The four white spaces occur between the five black bars. In the first step, focus on the bars. The number of permutations of five black bars when two are B and three are b is

$$\frac{5!}{2!3!} = 10$$

In the second step, consider the white spaces. A code has three narrow spaces w and one wide space W so there are four possible locations for the wide space. Therefore, the number of possible codes is $10 \times 4 = 40$. If one code is held back as a start/stop delimiter, then 39 other characters can be coded by this system (and the name comes from this result).

Combinations

Another counting problem of interest is the number of subsets of r elements that can be selected from a set of n elements. Here, order is not important. These are called **combinations**. Every subset of r elements can be indicated by listing the elements in the set and marking each element with a "$*$" if it is to be included in the subset. Therefore, each permutation of r $*$'s and $n - r$ blanks indicates a different subset, and the numbers of these are obtained from Equation 2-3. For example, if the set is $S = \{a,b,c,d\}$, the subset $\{a,c\}$ can be indicated as

$$\begin{matrix} a & b & c & d \\ * & & * & \end{matrix}$$

Combinations

The number of combinations, subsets of r elements that can be selected from a set of n elements, is denoted as $\binom{n}{r}$ or C_r^n and

$$C_r^n = \binom{n}{r} = \frac{n!}{r!(n-r)!} \tag{2-4}$$

Example 2-13 **Printed Circuit Board Layout** A printed circuit board has eight different locations in which a component can be placed. If five identical components are to be placed on the board, how many different designs are possible?

Each design is a subset of size five from the eight locations that are to contain the components. From Equation 2-4, the number of possible designs is

$$\frac{8!}{5!3!} = 56$$

The following example uses the multiplication rule in combination with Equation 2-4 to answer a more difficult, but common, question. In random experiments in which items are selected from a batch, an item may or may not be replaced before the next one is selected. This is referred to as sampling **with** or **without replacement**, respectively.

Example 2-14 **Sampling without Replacement** A bin of 50 manufactured parts contains 3 defective parts and 47 nondefective parts. A sample of 6 parts is selected from the 50 parts without replacement. That is, each part can be selected only once, and the sample is a subset of the 50 parts. How many different samples are there of size 6 that contain exactly 2 defective parts?

A subset containing exactly 2 defective parts can be formed by first choosing the 2 defective parts from the three defective parts. Using Equation 2-4, this step can be completed in

$$\binom{3}{2} = \frac{3!}{2!1!} = 3 \text{ different ways}$$

Then, the second step is to select the remaining 4 parts from the 47 acceptable parts in the bin. The second step can be completed in

$$\binom{47}{4} = \frac{47!}{4!\,43!} = 178,365 \text{ different ways}$$

Therefore, from the multiplication rule, the number of subsets of size 6 that contain exactly 2 defective parts is

$$3 \times 178,365 = 535,095$$

As an additional computation, the total number of different subsets of size 6 is found to be

$$\binom{50}{6} = \frac{50!}{6!\,44!} = 15,890,700$$

Exercises FOR SECTION 2-1

⊕ Problem available in *WileyPLUS* at instructor's discretion.
⊕ Go Tutorial Tutoring problem available in *WileyPLUS* at instructor's discretion

Provide a reasonable description of the sample space for each of the random experiments in Exercises 2-1 to 2-17. There can be more than one acceptable interpretation of each experiment. Describe any assumptions you make.

2-1. Each of three machined parts is classified as either above or below the target specification for the part.

2-2. Each of four transmitted bits is classified as either in error or not in error.

2-3. In the final inspection of electronic power supplies, either units pass, or three types of nonconformities might occur: functional, minor, or cosmetic. Three units are inspected.

2-4. The number of hits (views) is recorded at a high-volume Web site in a day.

2-5. Each of 24 Web sites is classified as containing or not containing banner ads.

2-6. An ammeter that displays three digits is used to measure current in milliamperes.

2-7. A scale that displays two decimal places is used to measure material feeds in a chemical plant in tons.

2-8. ⊕ The following two questions appear on an employee survey questionnaire. Each answer is chosen from the five-point scale 1 (never), 2, 3, 4, 5 (always).

Is the corporation willing to listen to and fairly evaluate new ideas?

How often are my coworkers important in my overall job performance?

2-9. The concentration of ozone to the nearest part per billion.

2-10. The time until a service transaction is requested of a computer to the nearest millisecond.

2-11. The pH reading of a water sample to the nearest tenth of a unit.

2-12. The voids in a ferrite slab are classified as small, medium, or large. The number of voids in each category is measured by an optical inspection of a sample.

2-13. The time of a chemical reaction is recorded to the nearest millisecond.

2-14. An order for an automobile can specify either an automatic or a standard transmission, either with or without air conditioning, and with any one of the four colors red, blue, black, or white. Describe the set of possible orders for this experiment.

2-15. A sampled injection-molded part could have been produced in either one of two presses and in any one of the eight cavities in each press.

2-16. An order for a computer system can specify memory of 4, 8, or 12 gigabytes and disk storage of 200, 300, or 400 gigabytes. Describe the set of possible orders.

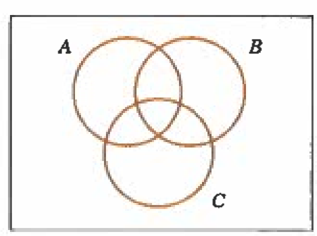
2-17. ⊕ Calls are repeatedly placed to a busy phone line until a connection is achieved.

2-18. Three attempts are made to read data in a magnetic storage device before an error recovery procedure that repositions the magnetic head is used. The error recovery procedure attempts three repositionings before an "abort" message is sent to the operator. Let

s denote the success of a read operation
f denote the failure of a read operation
S denote the success of an error recovery procedure
F denote the failure of an error recovery procedure
A denote an abort message sent to the operator

Describe the sample space of this experiment with a tree diagram.

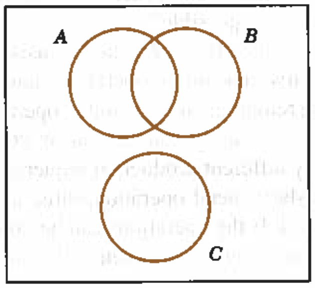
2-19. Three events are shown on the Venn diagram in the following figure:



Reproduce the figure and shade the region that corresponds to each of the following events.

(a) A' (b) $A \cap B$ (c) $(A \cap B) \cup C$
(d) $(B \cup C)'$ (e) $(A \cap B)' \cup C$

2-20. Three events are shown on the Venn diagram in the following figure:



Reproduce the figure and shade the region that corresponds to each of the following events.

(a) A' (b) $(A \cap B) \cup (A \cap B')$
(c) $(A \cap B) \cup C$ (d) $(B \cup C)'$
(e) $(A \cap B)' \cup C$

2-21. ⊕ A digital scale that provides weights to the nearest gram is used.

(a) What is the sample space for this experiment?

Let A denote the event that a weight exceeds 11 grams, let B denote the event that a weight is less than or equal to 15 grams, and let C denote the event that a weight is greater than or equal to 8 grams and less than 12 grams.

Describe the following events.

(b) $A \cup B$ (c) $A \cap B$
(d) A' (e) $A \cup B \cup C$
(f) $(A \cup C)'$ (g) $A \cap B \cap C$
(h) $B' \cap C$ (i) $A \cup (B \cap C)$

2-22. In an injection-molding operation, the length and width, denoted as X and Y, respectively, of each molded part are evaluated. Let

A denote the event of $48 < X < 52$ centimeters
B denote the event of $9 < Y < 11$ centimeters

Construct a Venn diagram that includes these events. Shade the areas that represent the following:

(a) A (b) $A \cap B$
(c) $A' \cup B$ (d) $A \cap B$
(e) If these events were mutually exclusive, how successful would this production operation be? Would the process produce parts with $X = 50$ centimeters and $Y = 10$ centimeters?

2-23. ⊕ **Go Tutorial** Four bits are transmitted over a digital communications channel. Each bit is either distorted or received without distortion. Let Ai denote the event that the ith bit is distorted, $i = 1, \ldots, 4$.

(a) Describe the sample space for this experiment.
(b) Are the A_i's mutually exclusive?

Describe the outcomes in each of the following events:

(c) A_1 (d) A_1'
(e) $A_1 \cap A_2 \cap A_3 \cap A_4$ (f) $(A_1 \cap A_2) \cup (A_3 \cap A_4)$

2-24. In light-dependent photosynthesis, light quality refers to the wavelengths of light that are important. The wavelength of a sample of photosynthetically active radiations (PAR) is measured to the nearest nanometer. The red range is 675–700 nm and the blue range is 450–500 nm. Let A denote the event that PAR occurs in the red range, and let B denote the event that PAR occurs in the blue range. Describe the sample space and indicate each of the following events:

(a) A (b) B (c) $A \cap B$ (d) $A \cup B$

2-25. In control replication, cells are replicated over a period of two days. Not until mitosis is completed can freshly synthesized DNA be replicated again. Two control mechanisms have been identified—one positive and one negative. Suppose that a replication is observed in three cells. Let A denote the event that all cells are identified as positive, and let B denote the event that all cells are negative. Describe the sample space graphically and display each of the following events:

(a) A (b) B
(c) $A \cap B$ (d) $A \cup B$

2-26. ⊕ Disks of polycarbonate plastic from a supplier are analyzed for scratch and shock resistance. The results from 100 disks are summarized here:

		Shock Resistance	
		High	Low
Scratch	High	70	9
Resistance	Low	16	5

Let A denote the event that a disk has high shock resistance, and let B denote the event that a disk has high scratch resistance. Determine the number of disks in $A \cap B$, A', and $A \cup B$.

2-27. Samples of a cast aluminum part are classified on the basis of surface finish (in microinches) and edge finish. The results of 100 parts are summarized as follows:

		Edge Finish	
		Excellent	Good
Surface	Excellent	80	2
Finish	Good	10	8

(a) Let A denote the event that a sample has excellent surface finish, and let B denote the event that a sample has excellent edge finish. Determine the number of samples in $A' \cap B$, B' and in $A \cup B$.
(b) Assume that each of two samples is to be classified on the basis of surface finish, either excellent or good, and on the basis of edge finish, either excellent or good. Use a tree diagram to represent the possible outcomes of this experiment.

2-28. ⊕ Samples of emissions from three suppliers are classified for conformance to air-quality specifications. The results from 100 samples are summarized as follows:

		Conforms	
		Yes	No
	1	22	8
Supplier	2	25	5
	3	30	10

Let A denote the event that a sample is from supplier 1, and let B denote the event that a sample conforms to specifications. Determine the number of samples in $A' \cap B$, B', and $A \cup B$.

2-29. The rise time of a reactor is measured in minutes (and fractions of minutes). Let the sample space be positive, real numbers. Define the events A and B as follows: $A = \{x \mid x < 72.5\}$ and $B = \{x \mid x > 52.5\}$.
Describe each of the following events:
(a) A' (b) B'
(c) $A \cap B$ (d) $A \cup B$

2-30. A sample of two items is selected without replacement from a batch. Describe the (ordered) sample space for each of the following batches:
(a) The batch contains the items $\{a, b, c, d\}$.
(b) The batch contains the items $\{a, b, c, d, e, f, g\}$.
(c) The batch contains 4 defective items and 20 good items.
(d) The batch contains 1 defective item and 20 good items.

2-31. A sample of two printed circuit boards is selected without replacement from a batch. Describe the (ordered) sample space for each of the following batches:
(a) The batch contains 90 boards that are not defective, 8 boards with minor defects, and 2 boards with major defects.
(b) The batch contains 90 boards that are not defective, 8 boards with minor defects, and 1 board with major defects.

2-32. Counts of the Web pages provided by each of two computer servers in a selected hour of the day are recorded. Let A denote the event that at least 10 pages are provided by server 1, and let B denote the event that at least 20 pages are provided by server 2. Describe the sample space for the numbers of pages for the two servers graphically in an $x-y$ plot. Show each of the following events on the sample space graph:
(a) A (b) B
(c) $A \cap B$ (d) $A \cup B$

2-33. A reactor's rise time is measured in minutes (and fractions of minutes). Let the sample space for the rise time of each batch be positive, real numbers. Consider the rise times of *two* batches. Let A denote the event that the rise time of batch 1 is less than 72.5 minutes, and let B denote the event that the rise time of batch 2 is greater than 52.5 minutes.

Describe the sample space for the rise time of two batches graphically and show each of the following events on a two-dimensional plot:
(a) A (b) B'
(c) $A \cap B$ (d) $A \cup B$

2-34. A wireless garage door opener has a code determined by the up or down setting of 12 switches. How many outcomes are in the sample space of possible codes?

2-35. An order for a computer can specify any one of five memory sizes, any one of three types of displays, and any one of four sizes of a hard disk, and can either include or not include a pen tablet. How many different systems can be ordered?

2-36. In a manufacturing operation, a part is produced by machining, polishing, and painting. If there are three machine tools, four polishing tools, and three painting tools, how many different routings (consisting of machining, followed by polishing, and followed by painting) for a part are possible?

2-37. New designs for a wastewater treatment tank have proposed three possible shapes, four possible sizes, three locations for input valves, and four locations for output valves. How many different product designs are possible?

2-38. A manufacturing process consists of 10 operations that can be completed in any order. How many different production sequences are possible?

2-39. A manufacturing operation consists of 10 operations. However, five machining operations must be completed before any of the remaining five assembly operations can begin. Within each set of five, operations can be completed in any order. How many different production sequences are possible?

2-40. In a sheet metal operation, three notches and four bends are required. If the operations can be done in any order, how many different ways of completing the manufacturing are possible?

2-41. A batch of 140 semiconductor chips is inspected by choosing a sample of 5 chips. Assume 10 of the chips do not conform to customer requirements.
(a) How many different samples are possible?
(b) How many samples of five contain exactly one nonconforming chip?
(c) How many samples of five contain at least one nonconforming chip?

2-42. In the layout of a printed circuit board for an electronic product, 12 different locations can accommodate chips.
(a) If five different types of chips are to be placed on the board, how many different layouts are possible?
(b) If the five chips that are placed on the board are of the same type, how many different layouts are possible?

2-43. In the laboratory analysis of samples from a chemical process, five samples from the process are analyzed daily. In addition, a control sample is analyzed twice each day to check the calibration of the laboratory instruments.
(a) How many different sequences of process and control samples are possible each day? Assume that the five process samples are considered identical and that the two control samples are considered identical.
(b) How many different sequences of process and control samples are possible if we consider the five process samples to be different and the two control samples to be identical?
(c) For the same situation as part (b), how many sequences are possible if the first test of each day must be a control sample?

2-44. In the design of an electromechanical product, 12 components are to be stacked into a cylindrical casing in a manner that minimizes the impact of shocks. One end of the casing is designated as the bottom and the other end is the top.
(a) If all components are different, how many different designs are possible?
(b) If seven components are identical to one another, but the others are different, how many different designs are possible?
(c) If three components are of one type and identical to one another, and four components are of another type and identical to one another, but the others are different, how many different designs are possible?

2-45. Consider the design of a communication system.
(a) How many three-digit phone prefixes that are used to represent a particular geographic area (such as an area code) can be created from the digits 0 through 9?

(b) As in part (a), how many three-digit phone prefixes are possible that do not start with 0 or 1, but contain 0 or 1 as the middle digit?
(c) How many three-digit phone prefixes are possible in which no digit appears more than once in each prefix?

2-46. A *byte* is a sequence of eight bits and each bit is either 0 or 1.
(a) How many different bytes are possible?
(b) If the first bit of a byte is a parity check, that is, the first byte is determined from the other seven bits, how many different bytes are possible?

2-47. In a chemical plant, 24 holding tanks are used for final product storage. Four tanks are selected at random and without replacement. Suppose that six of the tanks contain material in which the viscosity exceeds the customer requirements.
(a) What is the probability that exactly one tank in the sample contains high-viscosity material?
(b) What is the probability that at least one tank in the sample contains high-viscosity material?
(c) In addition to the six tanks with high viscosity levels, four different tanks contain material with high impurities. What is the probability that exactly one tank in the sample contains high-viscosity material and exactly one tank in the sample contains material with high impurities?

2-48. ⊕ Plastic parts produced by an injection-molding operation are checked for conformance to specifications. Each tool contains 12 cavities in which parts are produced, and these parts fall into a conveyor when the press opens. An inspector chooses 3 parts from among the 12 at random. Two cavities are affected by a temperature malfunction that results in parts that do not conform to specifications.
(a) How many samples contain exactly 1 nonconforming part?
(b) How many samples contain at least 1 nonconforming part?

2-49. A bin of 50 parts contains 5 that are defective. A sample of 10 parts is selected at random, without replacement. How many samples contain at least four defective parts?

2-50. The following table summarizes 204 endothermic reactions involving sodium bicarbonate.

Final Temperature Conditions	Heat Absorbed (cal)	
	Below Target	Above Target
266 K	12	40
271 K	44	16
274 K	56	36

Let A denote the event that a reaction's final temperature is 271 K or less. Let B denote the event that the heat absorbed is below target. Determine the number of reactions in each of the following events.

(a) $A \cap B$ (b) A' (c) $A \cup B$ (d) $A \cup B'$ (e) $A' \cap B'$

2-51. A Web ad can be designed from four different colors, three font types, five font sizes, three images, and five text phrases. How many different designs are possible?

2-52. Consider the hospital emergency department data in Example 2-8. Let A denote the event that a visit is to hospital 1, and let B denote the event that a visit results in admittance to any hospital. Determine the number of persons in each of the following events.

(a) $A \cap B$ (b) A' (c) $A \cup B$ (d) $A \cup B'$ (e) $A' \cap B'$

2-53. An article in *The Journal of Data Science* ["A Statistical Analysis of Well Failures in Baltimore County" (2009, Vol. 7, pp. 111–127)] provided the following table of well failures for different geological formation groups in Baltimore County.

Geological Formation Group	Wells: Failed	Wells: Total
Gneiss	170	1685
Granite	2	28
Loch raven schist	443	3733
Mafic	14	363
Marble	29	309
Prettyboy schist	60	1403
Other schists	46	933
Serpentine	3	39

Let A denote the event that the geological formation has more than 1000 wells, and let B denote the event that a well failed. Determine the number of wells in each of the following events.
(a) $A \cap B$ (b) A' (c) $A \cup B$ (d) $A \cap B'$ (e) $A' \cap B'$

2-54. Similar to the hospital schedule in Example 2-11, suppose that an operating room needs to handle three knee, four hip, and five shoulder surgeries.
(a) How many different sequences are possible?
(b) How many different sequences have all hip, knee, and shoulder surgeries scheduled consecutively?
(c) How many different schedules begin and end with a knee surgery?

2-55. Consider the bar code in Example 2-12. One code is still held back as a delimiter. For each of the following cases, how many characters can be encoded?
(a) The constraint of exactly two wide bars is replaced with one that requires exactly one wide bar.
(b) The constraint of exactly two wide bars is replaced with one that allows either one or two wide bars.
(c) The constraint of exactly two wide bars is dropped.
(d) The constraints of exactly two wide bars and one wide space are dropped.

2-56. A computer system uses passwords that contain exactly eight characters, and each character is 1 of the 26 lowercase letters (*a–z*) or 26 uppercase letters (*A–Z*) or 10 integers (0–9). Let Ω denote the set of all possible passwords, and let A and B denote the events that consist of passwords with only letters or only integers, respectively. Determine the number of passwords in each of the following events.
(a) Ω (b) A (c) $A' \cap B'$
(d) Passwords that contain at least 1 integer
(e) Passwords that contain exactly 1 integer

2-57. The article "Term Efficacy of Ribavirin Plus Interferon Alfa in the Treatment of Chronic Hepatitis C," [*Gastroenterology* (1996, Vol. 111, no. 5, pp. 1307–1312)], considered the effect of two treatments and a control for treatment of hepatitis C. The following table provides the total patients in each group and the number that showed a complete (positive) response after 24 weeks of treatment.

	Complete Response	Total
Ribavirin plus interferon alfa	16	21
Interferon alfa	6	19
Untreated controls	0	20

Let A denote the event that the patient was treated with ribavirin plus interferon alfa, and let B denote the event that the response was complete. Determine the number of patients in each of the following events.

(a) A (b) $A \cap B$ (c) $A \cup B$ (d) $A' \cap B'$

2-2 Interpretations and Axioms of Probability

In this chapter, we introduce probability for **discrete sample spaces**—those with only a finite (or countably infinite) set of outcomes. The restriction to these sample spaces enables us to simplify the concepts and the presentation without excessive mathematics.

Probability is used to quantify the likelihood, or chance, that an outcome of a random experiment will occur. "The chance of rain today is 30%" is a statement that quantifies our feeling about the possibility of rain. The likelihood of an outcome is quantified by assigning a number from the interval [0, 1] to the outcome (or a percentage from 0 to 100%). Higher numbers indicate that the outcome is more likely than lower numbers. A 0 indicates an outcome will not occur. A probability of 1 indicates that an outcome will occur with certainty.

The probability of an outcome can be interpreted as our subjective probability, or **degree of belief**, that the outcome will occur. Different individuals will no doubt assign different probabilities to the same outcomes. Another interpretation of probability is based on the conceptual model of repeated replications of the random experiment. The probability of an outcome is interpreted as the limiting value of the proportion of times the outcome occurs in n repetitions of the random experiment as n increases beyond all bounds. For example, if we assign probability 0.2 to the outcome that there is a corrupted pulse in a digital signal, we might interpret this assignment as implying that, if we analyze many pulses, approximately 20% of them will be corrupted. This example provides a **relative frequency** interpretation of probability. The proportion, or relative frequency, of replications of the experiment that result in the outcome is 0.2. Probabilities are chosen so that the sum of the probabilities of all outcomes in an experiment adds up to 1. This convention facilitates the relative frequency interpretation of probability. Fig. 2-10 illustrates the concept of relative frequency.

Probabilities for a random experiment are often assigned on the basis of a reasonable model of the system under study. One approach is to base probability assignments on the simple concept of equally likely outcomes. For example, suppose that we select 1 laser diode **randomly** from a batch of 100. *Randomly* implies that it is reasonable to assume that each diode in the batch has an equal chance of being selected. Because the sum of the probabilities must equal 1, the probability model for this experiment assigns probability of 0.01 to each of the 100 outcomes. We can interpret the probability by imagining many replications of the experiment. Each time we start with all 100 diodes and select 1 at random. The probability 0.01 assigned to a particular diode represents the proportion of replicates in which a particular diode is selected. When the model of **equally likely outcomes** is assumed, the probabilities are chosen to be equal.

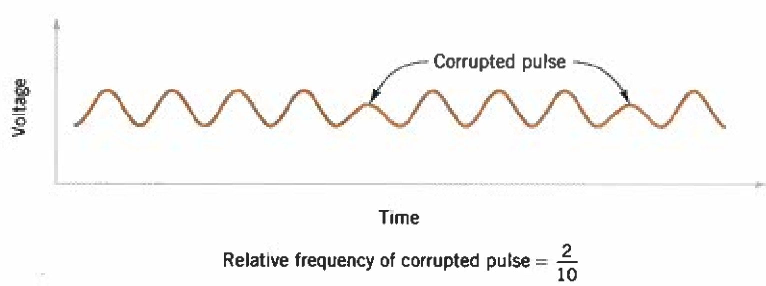


FIGURE 2-10 Relative frequency of corrupted pulses sent over a communication channel.

Axioms of Probability

Probability is a number that is assigned to each member of a collection of events from a random experiment that satisfies the following properties:

If S is the sample space and E is any event in a random experiment,

(1) $P(S)=1$

(2) $0\le P(E)\le 1$

(3) For two events E_1 and E_2 with $E_1\cap E_2=\varnothing$

$$P(E_1\cup E_2)=P(E_1)+P(E_2)$$

The property that $0\le P(E)\le 1$ is equivalent to the requirement that a relative frequency must be between 0 and 1. The property that $P(S)=1$ is a consequence of the fact that an outcome from the sample space occurs on every trial of an experiment. Consequently, the relative frequency of S is 1. Property 3 implies that if the events E_1 and E_2 have no outcomes in common, the relative frequency of outcomes in $E_1\cup E_2$ is the sum of the relative frequencies of the outcomes in E_1 and E_2.

These axioms imply the following results. The derivations are left as exercises at the end of this section. Now,

$$P(\varnothing)=0$$

and for any event E,

$$P(E')=1-P(E)$$

For example, if the probability of the event E is 0.4, our interpretation of relative frequency implies that the probability of E' is 0.6. Furthermore, if the event E_1 is contained in the event E_2,

$$P(E_1)\le P(E_2)$$

Exercises FOR SECTION 2-2

⊕ Problem available in *WileyPLUS* at instructor's discretion.

⊕ **Go Tutorial** Tutoring problem available in *WileyPLUS* at instructor's discretion

2-58. Each of the possible five outcomes of a random experiment is equally likely. The sample space is $\{a,b,c,d,e\}$. Let A denote the event $\{a,b\}$, and let B denote the event $\{c,d,e\}$. Determine the following:

(a) $P(A)$ (b) $P(B)$ (c) $P(A')$
(d) $P(A\cup B)$ (e) $P(A\cap B)$

2-59. ⊕ The sample space of a random experiment is $\{a,b,c,d,e\}$ with probabilities 0.1, 0.1, 0.2, 0.4, and 0.2, respectively. Let A denote the event $\{a,b,c\}$, and let B denote the event $\{c,d,e\}$. Determine the following:

(a) $P(A)$ (b) $P(B)$ (c) $P(A')$
(d) $P(A\cup B)$ (e) $P(A\cap B)$

2-60. ⊕ Orders for a computer are summarized by the optional features that are requested as follows:

	Proportion of Orders
No optional features	0.3
One optional feature	0.5
More than one optional feature	0.2

(a) What is the probability that an order requests at least one optional feature?

(b) What is the probability that an order does not request more than one optional feature?

2-61. ⊕ If the last digit of a weight measurement is equally likely to be any of the digits 0 through 9,

(a) What is the probability that the last digit is 0?

(b) What is the probability that the last digit is greater than or equal to 5?

2-62. A part selected for testing is equally likely to have been produced on any one of six cutting tools.

(a) What is the sample space?

(b) What is the probability that the part is from tool 1?

(c) What is the probability that the part is from tool 3 or tool 5?

(d) What is the probability that the part is not from tool 4?

2-63. ⊕ An injection-molded part is equally likely to be obtained from any one of the eight cavities on a mold.

(a) What is the sample space?

(b) What is the probability that a part is from cavity 1 or 2?

(c) What is the probability that a part is from neither cavity 3 nor 4?

2-64. ⊕ In an acid-base titration, a base or acid is gradually added to the other until they have completely neutralized each other. Because acids and bases are usually colorless (as are the water and salt produced in the neutralization reaction), pH is measured to monitor the reaction. Suppose that the equivalence point is reached after approximately 100 mL of an NaOH solution has been added (enough to react with all the acetic acid present) but that replicates are equally likely to indicate from 95 to 104 mL to the nearest mL. Assume that volumes are measured to the nearest mL and describe the sample space.

(a) What is the probability that equivalence is indicated at 100 mL?

(b) What is the probability that equivalence is indicated at less than 100 mL?

(c) What is the probability that equivalence is indicated between 98 and 102 mL (inclusive)?

2-65. ⊕ In a NiCd battery, a fully charged cell is composed of nickelic hydroxide. Nickel is an element that has multiple oxidation states and that is usually found in the following states:

Nickel Charge	Proportions Found
0	0.17
+2	0.35
+3	0.33
+4	0.15

(a) What is the probability that a cell has at least one of the positive nickel-charged options?

(b) What is the probability that a cell is *not* composed of a positive nickel charge greater than +3?

2-66. ⊕ A credit card contains 16 digits between 0 and 9. However, only 100 million numbers are valid. If a number is entered randomly, what is the probability that it is a valid number?

2-67. ⊕ Suppose your vehicle is licensed in a state that issues license plates that consist of three digits (between 0 and 9) followed by three letters (between *A* and *Z*). If a license number is selected randomly, what is the probability that yours is the one selected?

2-68. A message can follow different paths through servers on a network. The sender's message can go to one of five servers for the first step; each of them can send to five servers at the second step; each of those can send to four servers at the third step; and then the message goes to the recipient's server.

(a) How many paths are possible?

(b) If all paths are equally likely, what is the probability that a message passes through the first of four servers at the third step?

2-69. Magnesium alkyls are used as homogenous catalysts in the production of linear low-density polyethylene (LLDPE), which requires a finer magnesium powder to sustain a reaction. Redox reaction experiments using four different amounts of magnesium powder are performed. Each result may or may not be further reduced in a second step using three different magnesium powder amounts. Each of these results may or may not be further reduced in a third step using three different amounts of magnesium powder.

(a) How many experiments are possible?

(b) If all outcomes are equally likely, what is the probability that the best result is obtained from an experiment that uses all three steps?

(c) Does the result in part (b) change if five or six or seven different amounts are used in the first step? Explain.

2-70. ⊕ Disks of polycarbonate plastic from a supplier are analyzed for scratch and shock resistance. The results from 100 disks are summarized as follows:

		Shock Resistance	
		High	Low
Scratch	High	70	9
Resistance	Low	16	5

Let *A* denote the event that a disk has high shock resistance, and let *B* denote the event that a disk has high scratch resistance. If a disk is selected at random, determine the following probabilities:

(a) $P(A)$ (b) $P(B)$ (c) $P(A')$

(d) $P(A \cap B)$ (e) $P(A \cup B)$ (f) $P(A' \cup B)$

2-71. Samples of emissions from three suppliers are classified for conformance to air-quality specifications. The results from 100 samples are summarized as follows:

		Conforms	
		Yes	No
Supplier	1	22	8
	2	25	5
	3	30	10

Let *A* denote the event that a sample is from supplier 1, and let *B* denote the event that a sample conforms to specifications. If a sample is selected at random, determine the following probabilities:

(a) $P(A)$ (b) $P(B)$ (c) $P(A')$

(d) $P(A \cap B)$ (e) $P(A \cup B)$ (f) $P(A' \cap B)$

2-72. ⊕ An article in the *Journal of Database Management* ["Experimental Study of a Self-Tuning Algorithm for DBMS Buffer Pools" (2005, Vol. 16, pp. 1–20)] provided the workload used in the TPC-C OLTP (Transaction Processing Performance Council's Version C On-Line Transaction Processing) benchmark, which simulates a typical order entry application. See Table 2E-1.

The frequency of each type of transaction (in the second column) can be used as the percentage of each type of transaction. The average number of *selects* operations required for each type of transaction is shown. Let *A* denote the event of transactions with an average number of *selects* operations of 12 or fewer. Let *B* denote the event of transactions with an average number of *updates* operations of 12 or fewer. Calculate the following probabilities.

(a) $P(A)$ (b) $P(B)$ (c) $P(A \cap B)$

(d) $P(A \cap B')$ (e) $P(A \cup B)$

2-73. Use the axioms of probability to show the following:

(a) For any event E, $P(E') = 1 - P(E)$.

(b) $P(\varnothing) = 0$

(c) If *A* is contained in *B*, then $P(A) \le P(B)$.

2-74. Consider the endothermic reaction's in Exercise 2-50. Let *A* denote the event that a reaction's final temperature is 271 K or less. Let *B* denote the event that the heat absorbed is above target. Determine the following probabilities.

TABLE • 2E-1 Average Frequencies and Operations in TPC-C

Transaction	Frequency	Selects	Updates	Inserts	Deletes	Nonunique Selects	Joins
New order	43	23	11	12	0	0	0
Payment	44	4.2	3	1	0	0.6	0
Order status	4	11.4	0	0	0	0.6	0
Delivery	5	130	120	0	10	0	0
Stock level	4	0	0	0	0	0	1

(a) $P(A \cap B)$ (b) $P(A')$ (c) $P(A \cup B)$
(d) $P(A \cup B')$ (e) $P(A' \cap B')$

2-75. A Web ad can be designed from four different colors, three font types, five font sizes, three images, and five text phrases. A specific design is randomly generated by the Web server when you visit the site. If you visit the site five times, what is the probability that you will not see the same design?

2-76. Consider the hospital emergency room data in Example 2-8. Let A denote the event that a visit is to hospital 4, and let B denote the event that a visit results in LWBS (at any hospital). Determine the following probabilities.
(a) $P(A \cap B)$ (b) $P(A')$ (c) $P(A \cup B)$
(d) $P(A \cup B')$ (e) $P(A' \cap B')$

2-77. Consider the well failure data in Exercise 2-53. Let A denote the event that the geological formation has more than 1000 wells, and let B denote the event that a well failed. Determine the following probabilities.
(a) $P(A \cap B)$ (b) $P(A')$ (c) $P(A \cup B)$
(d) $P(A \cup B')$ (e) $P(A' \cap B')$

2-78. Consider the bar code in Example 2-12. Suppose that all 40 codes are equally likely (none is held back as a delimiter). Determine the probability for each of the following:
(a) A wide space occurs before a narrow space.
(b) Two wide bars occur consecutively.
(c) Two consecutive wide bars are at the start or end.
(d) The middle bar is wide.

2-79. Similar to the hospital schedule in Example 2-11, suppose that an operating room needs to schedule three knee, four hip, and five shoulder surgeries. Assume that all schedules are equally likely. Determine the probability for each of the following:
(a) All hip surgeries are completed before another type of surgery.
(b) The schedule begins with a hip surgery.
(c) The first and last surgeries are hip surgeries.
(d) The first two surgeries are hip surgeries.

2-80. Suppose that a patient is selected randomly from the those described in Exercise 2-57. Let A denote the event that the patient is in the group treated with interferon alfa, and let B denote the event that the patient has a complete response. Determine the following probabilities.
(a) $P(A)$ (b) $P(B)$
(c) $P(A \cap B)$ (d) $P(A \cup B)$ (e) $P(A' \cup B)$

2-81. A computer system uses passwords that contain exactly eight characters, and each character is one of 26 lowercase letters (a–z) or 26 uppercase letters (A–Z) or 10 integers (0–9). Let Ω denote the set of all possible passwords, and let A and B denote the events that consist of passwords with only letters or only integers, respectively. Suppose that all passwords in Ω are equally likely. Determine the probability of each of the following:
(a) A (b) B
(c) A password contains at least 1 integer.
(d) A password contains exactly 2 integers.

2-3 Addition Rules

Joint events are generated by applying basic set operations to individual events. Unions of events, such as $A \cup B$; intersections of events, such as $A \cap B$; and complements of events, such as A'—are commonly of interest. The probability of a joint event can often be determined from the probabilities of the individual events that it comprises. Basic set operations are also sometimes helpful in determining the probability of a joint event. In this section, the focus is on unions of events.

Example 2-19 **Semiconductor Wafers** Table 2-1 lists the history of 940 wafers in a semiconductor manufacturing process. Suppose that 1 wafer is selected at random. Let H denote the event that the wafer contains high levels of contamination. Then, $P(H) = 358/940$.

Let C denote the event that the wafer is in the center of a sputtering tool. Then, $P(C) = 626/940$. Also, $P(H \cap C)$ is the probability that the wafer is from the center of the sputtering tool and contains high levels of contamination. Therefore,

$$P(H \cap C) = 112/940$$

The event $H \cup C$ is the event that a wafer is from the center of the sputtering tool or contains high levels of contamination (or both). From the table, $P(H \cup C) = 872/940$. An alternative calculation of $P(H \cup C)$ can be obtained as follows. The 112 wafers in the event $H \cap C$ are included once in the calculation of $P(H)$ and again in the calculation of $P(C)$. Therefore, $P(H \cup C)$ can be determined to be

$$P(H \cup C) = P(H) + P(C) - P(H \cap C)$$
$$= 358/940 + 626/940 - 112/940 = 872/940$$

Practical Interpretation: To better understand the sources of contamination, yield from defferent locations on wafers are routinely aggregated.

TABLE • 2-1 Wafers in Semiconductor Manufacturing Classified by Contamination and Location

	Location in Sputtering Tool		
Contamination	**Center**	**Edge**	**Total**
Low	514	68	582
High	112	246	358
Total	626	314	

The preceding example illustrates that the probability of A or B is interpreted as $P(A \cup B)$ and that the following general **addition rule** applies.

Probability of a Union

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{2-5}$$

Example 2-20 **Semiconductor Wafers and Location** The wafers in Example 2-19 were further classified by the degree of contamination. Table 2-2 shows the proportion of wafers in each category. What is the probability that a wafer was either at the edge or that it contains four or more particles? Let E_1 denote the event that a wafer contains four or more particles, and let E_2 denote the event that a wafer was at the edge.

The requested probability is $P(E_1 \cup E_2)$. Now, $P(E_1) = 0.15$ and $P(E_2) = 0.28$. Also, from the table, $P(E_1 \cap E_2) = 0.04$. Therefore, using Equation 2-1, we find that

$$P(E_1 \cup E_2) = 0.15 + 0.28 - 0.04 = 0.39$$

TABLE • 2-2 Wafers Classified by Contamination and Location

Number of Contamination Particles	Center	Edge	Totals
0	0.30	0.10	0.40
1	0.15	0.05	0.20
2	0.10	0.05	0.15
3	0.06	0.04	0.10
4	0.04	0.01	0.05
5 or more	0.07	0.03	0.10
Totals	0.72	0.28	1.00

What is the probability that a wafer contains less than two particles or that it is both at the edge and contains more than four particles? Let E_1 denote the event that a wafer contains less than two particles, and let E_2 denote the event that a wafer is both at the edge and contains more than four particles. The requested probability is $P(E_1 \cup E_2)$. Now, $P(E_1) = 0.60$ and $P(E_2) = 0.03$. Also, E_1 and E_2 are mutually exclusive. Consequently, there are no wafers in the intersection and $P(E_1 \cap E_2) = 0$. Therefore,

$$P(E_1 \cup E_2) = 0.60 + 0.03 = 0.63$$

Recall that two events A and B are said to be mutually exclusive if $A \cap B = \emptyset$. Then, $P(A \cap B) = 0$, and the general result for the probability of $A \cup B$ simplifies to the third axiom of probability.

If A and B are mutually exclusive events,

$$P(A \cup B) = P(A) + P(B) \tag{2-6}$$

Three or More Events

More complicated probabilities, such as $P(A \cup B \cup C)$, can be determined by repeated use of Equation 2-5 and by using some basic set operations. For example,

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

Upon expanding $P(A \cup B)$ by Equation 2-5 and using the distributed rule for set operations to simplify $P[(A \cup B) \cap C]$, we obtain

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) - P(A \cap B) + P(C) - P[(A \cap C) \cup (B \cap C)] \\ &= P(A) + P(B) - P(A \cap B) + P(C) - [P(A \cap C) + P(B \cap C) - P(A \cap B \cap C)] \\ &= P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C) \end{aligned}$$

We have developed a formula for the probability of the union of three events. Formulas can be developed for the probability of the union of any number of events, although the formulas become very complex. As a summary, for the case of three events,

$$\begin{aligned} P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) \\ - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C) \end{aligned} \tag{2-7}$$

Results for three or more events simplify considerably if the events are mutually exclusive. In general, a collection of events, $E_1, E_2, \ldots, E_k$, is said to be mutually exclusive if there is no overlap among any of them. The Venn diagram for several mutually exclusive events is shown in Fig. 2-12. By generalizing the reasoning for the union of two events, the following result can be obtained:

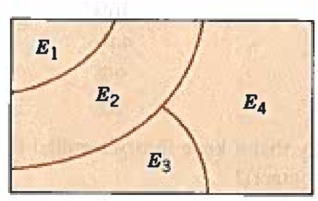


FIGURE 2-12 Venn diagram of four mutually exclusive events.

Mutually Exclusive Events

A collection of events, $E_1, E_2, \ldots, E_k$, is said to be **mutually exclusive** if for all pairs,

$$E_i \cap E_j = \varnothing$$

For a collection of mutually exclusive events,

$$P(E_1 \cup E_2 \cup \ldots \cup E_k) = P(E_1) + P(E_2) + \ldots P(E_k) \qquad \text{(2-8)}$$

Example 2-21 pH Here is a simple example of mutually exclusive events, which will be used quite frequently. Let X denote the pH of a sample. Consider the event that X is greater than 6.5 but less than or equal to 7.8. This probability is the sum of any collection of mutually exclusive events with union equal to the same range for X. One example is

$$P(6.5 < X \le 7.8) = P(6.5 \le X \le 7.0) + P(7.0 < X \le 7.5) + P(7.5 < X \le 7.8)$$

Another example is

$$P(6.5 < X \le 7.8) = P(6.5 < X \le 6.6) + P(6.6 < X \le 7.1) + P(7.1 < X \le 7.4) + P(7.4 < X \le 7.8)$$

The best choice depends on the particular probabilities available.

Practical Interpretation: The partition of an event into mutually exclusive subsets is widely used in later chapters to calculate probabilities.

Exercises FOR SECTION 2-3

⊕ Problem available in *WileyPLUS* at instructor's discretion.

⊕ Go Tutorial Tutoring problem available in *WileyPLUS* at instructor's discretion

2-82. ⊕ If $P(A) = 0.3$, $P(B) = 0.2$, and $P(A \cap B) = 0.1$, determine the following probabilities:

(a) $P(A')$ (b) $P(A \cup B)$ (c) $P(A' \cap B)$
(d) $P(A \cap B')$ (e) $P[(A \cup B)']$ (f) $P(A' \cup B)$

2-83. ⊕ If A, B, and C are mutually exclusive events with $P(A) = 0.2$, $P(B) = 0.3$, and $P(C) = 0.4$, determine the following probabilities:

(a) $P(A \cup B \cup C)$ (b) $P(A \cap B \cap C)$
(c) $P(A \cap B)$ (d) $P[(A \cup B) \cap C]$
(e) $P(A' \cap B' \cap C')$

2-84. ⊕ In the article "ACL Reconstruction Using Bone-Patellar Tendon-Bone Press-Fit Fixation: 10-Year Clinical Results" in *Knee Surgery, Sports Traumatology, Arthroscopy* (2005, Vol. 13, pp. 248–255), the following causes for knee injuries were considered:

Activity	Percentage of Knee Injuries
Contact sport	46%
Noncontact sport	44%
Activity of daily living	9%
Riding motorcycle	1%

(a) What is the probability that a knee injury resulted from a sport (contact or noncontact)?

(b) What is the probability that a knee injury resulted from an activity other than a sport?

2-85. ⊕ Disks of polycarbonate plastic from a supplier are analyzed for scratch and shock resistance. The results from 100 disks are summarized as follows:

		Shock Resistance	
		High	Low
Scratch	High	70	9
Resistance	Low	16	5

(a) If a disk is selected at random, what is the probability that its scratch resistance is high and its shock resistance is high?

(b) If a disk is selected at random, what is the probability that its scratch resistance is high or its shock resistance is high?

(c) Consider the event that a disk has high scratch resistance and the event that a disk has high shock resistance. Are these two events mutually exclusive?

2-86. ⊕ Strands of copper wire from a manufacturer are analyzed for strength and conductivity. The results from 100 strands are as follows:

	Strength	
	High	Low
High conductivity	74	8
Low conductivity	15	3

(a) If a strand is randomly selected, what is the probability that its conductivity is high and its strength is high?
(b) If a strand is randomly selected, what is the probability that its conductivity is low or its strength is low?
(c) Consider the event that a strand has low conductivity and the event that the strand has low strength. Are these two events mutually exclusive?

2-87. ⊕ The analysis of shafts for a compressor is summarized by conformance to specifications.

		Roundness Conforms	
		Yes	No
Surface Finish	Yes	345	5
Conforms	No	12	8

(a) If a shaft is selected at random, what is the probability that it conforms to surface finish requirements?
(b) What is the probability that the selected shaft conforms to surface finish requirements or to roundness requirements?
(c) What is the probability that the selected shaft either conforms to surface finish requirements or does not conform to roundness requirements?
(d) What is the probability that the selected shaft conforms to both surface finish and roundness requirements?

2-88. ⊕ Cooking oil is produced in two main varieties: mono- and polyunsaturated. Two common sources of cooking oil are corn and canola. The following table shows the number of bottles of these oils at a supermarket:

		Type of oil	
		Canola	Corn
Type of Unsaturation	Mono	7	13
	Poly	93	77

(a) If a bottle of oil is selected at random, what is the probability that it belongs to the polyunsaturated category?
(b) What is the probability that the chosen bottle is monounsaturated canola oil?

2-89. ⊕ A manufacturer of front lights for automobiles tests lamps under a high-humidity, high-temperature environment using intensity and useful life as the responses of interest. The following table shows the performance of 130 lamps:

		Useful life	
		Satisfactory	Unsatisfactory
Intensity	Satisfactory	117	3
	Unsatisfactory	8	2

(a) Find the probability that a randomly selected lamp will yield unsatisfactory results under any criteria.
(b) The customers for these lamps demand 95% satisfactory results. Can the lamp manufacturer meet this demand?

2-90. ⊕ A computer system uses passwords that are six characters, and each character is one of the 26 letters (a–z) or 10 integers (0–9). Uppercase letters are not used. Let A denote the event that a password begins with a vowel (either a, e, i, o, or u), and let B denote the event that a password ends with an even number (either 0, 2, 4, 6, or 8). Suppose a hacker selects a password at random. Determine the following probabilities:

(a) $P(A)$ (b) $P(B)$
(c) $P(A \cap B)$ (d) $P(A \cup B)$

2-91. Consider the endothermic reactions in Exercise 2-50. Let A denote the event that a reaction's final temperature is 271 K or less. Let B denote the event that the heat absorbed is above target. Use the addition rules to calculate the following probabilities.

(a) $P(A \cup B)$ (b) $P(A \cap B')$ (c) $P(A' \cup B')$

2-92. A Web ad can be designed from four different colors, three font types, five font sizes, three images, and five text phrases. A specific design is randomly generated by the Web server when you visit the site. Let A denote the event that the design color is red, and let B denote the event that the font size is not the smallest one. Use the addition rules to calculate the following probabilities.

(a) $P(A \cup B)$ (b) $P(A \cup B')$ (c) $P(A' \cup B')$

2-93. Consider the hospital emergency room data in Example 2-8. Let A denote the event that a visit is to hospital 4, and let B denote the event that a visit results in LWBS (at any hospital). Use the addition rules to calculate the following probabilities.

(a) $P(A \cup B)$ (b) $P(A \cup B')$ (c) $P(A' \cup B')$

2-94. Consider the well failure data in Exercise 2-53. Let A denote the event that the geological formation has more than 1000 wells, and let B denote the event that a well failed. Use the addition rules to calculate the following probabilities.

(a) $P(A \cup B)$ (b) $P(A \cup B')$ (c) $P(A' \cup B')$

2-95. Consider the bar code in Example 2-12. Suppose that all 40 codes are equally likely (none is held back as a delimiter). Determine the probability for each of the following:
(a) The first bar is wide or the second bar is wide.
(b) Neither the first nor the second bar is wide.
(c) The first bar is wide or the second bar is not wide.
(d) The first bar is wide or the first space is wide.

2-96. Consider the three patient groups in Exercise 2-57. Let A denote the event that the patient was treated with ribavirin plus interferon alfa, and let B denote the event that the response was complete. Determine the following probabilities:

(a) $P(A \cup B)$ (b) $P(A' \cup B)$ (c) $P(A \cup B')$

2-97. A computer system uses passwords that contain exactly eight characters, and each character is one of the 26 lowercase letters (a–z) or 26 uppercase letters (A–Z) or 10 integers (0–9). Assume all passwords are equally likely. Let A and B denote the events that consist of passwords with only letters or only integers, respectively. Determine the following probabilities:

(a) $P(A \cup B)$ (b) $P(A' \cup B)$
(c) P (Password contains exactly 1 or 2 integers)

2-98. The article ["Clinical and Radiographic Outcomes of Four Different Treatment Strategies in Patients with Early Rheumatoid Arthritis," *Arthritis & Rheumatism* (2005, Vol. 52, pp. 3381–3390)] considered four treatment groups. The groups consisted of patients with different drug therapies (such as prednisone and infliximab): sequential monotherapy (group 1), step-up combination therapy (group 2), initial combination therapy (group 3), or initial combination therapy with infliximab (group 4). Radiographs of hands and feet were used to evaluate disease progression. The number of patients without progression of joint damage was 76 of

114 patients (67%), 82 of 112 patients (73%), 104 of 120 patients (87%), and 113 of 121 patients (93%) in groups 1–4, respectively. Suppose that a patient is selected randomly. Let A denote the event that the patient is in group 1, and let B denote the event that there is no progression. Determine the following probabilities:
(a) $P(A \cup B)$ (b) $P(A' \cup B')$ (c) $P(A \cup B')$

2-4 Conditional Probability

Sometimes probabilities need to be reevaluated as additional information becomes available. A useful way to incorporate additional information into a probability model is to assume that the outcome that will be generated is a member of a given event. This event, say A, defines the conditions that the outcome is known to satisfy. Then probabilities can be revised to include this knowledge. The probability of an event B under the knowledge that the outcome will be in event A is denoted as

$$P(B|A)$$

and this is called the **conditional probability of B given A.**

A digital communication channel has an error rate of 1 bit per every 1000 transmitted. Errors are rare, but when they occur, they tend to occur in bursts that affect many consecutive bits. If a single bit is transmitted, we might model the probability of an error as 1/1000. However, if the previous bit was in error because of the bursts, we might believe that the probability that the next bit will be in error is greater than 1/1000.

In a thin film manufacturing process, the proportion of parts that are not acceptable is 2%. However, the process is sensitive to contamination problems that can increase the rate of parts that are not acceptable. If we knew that during a particular shift there were problems with the filters used to control contamination, we would assess the probability of a part being unacceptable as higher than 2%.

In a manufacturing process, 10% of the parts contain visible surface flaws and 25% of the parts with surface flaws are (functionally) defective parts. However, only 5% of parts without surface flaws are defective parts. The probability of a defective part depends on our knowledge of the presence or absence of a surface flaw. Let D denote the event that a part is defective, and let F denote the event that a part has a surface flaw. Then we denote the probability of D given or assuming that a part has a surface flaw, as $P(D \mid F)$. Because 25% of the parts with surface flaws are defective, our conclusion can be stated as $P(D \mid F) = 0.25$. Furthermore, because F' denotes the event that a part does not have a surface flaw and because 5% of the parts without surface flaws are defective, we have $P(D \mid F') = 0.05$. These results are shown graphically in Fig. 2-13.

Example 2-22 **Surface Flaws and Defectives** Table 2-3 provides an example of 400 parts classified by surface flaws and as (functionally) defective. For this table, the conditional probabilities match those discussed previously in this section. For example, of the parts with surface flaws (40 parts), the number of defective ones is 10. Therefore,

$$P(D \mid F) = 10/40 = 0.25$$

and of the parts without surface flaws (360 parts), the number of defective ones is 18. Therefore,

$$P(D \mid F') = 18/360 = 0.05$$

Practical Interpretation: The probability of being defective is five times greater for parts with surface flaws. This calculation illustrates how probabilities are adjusted for additional information. The result also suggests that there may be a link between surface flaws and functionally defective parts, which should be investigated.

TABLE • 2-3 Parts Classified

		Surface Flaws		
		Yes (event F)	No	Total
Defective	Yes (event D)	10	18	28
	No	30	342	372
	Total	40	360	400

Let A and B denote the events that the first and second part selected are defective, respectively. The probability requested can be expressed as $P(B|A)$. If the first part is defective, prior to selecting the second part the batch contains 49 parts, of which 2 are defective. Therefore,

$$P(B \mid A) = \frac{2}{49}$$

Example 2-25 Continuing Example 2-24, what is the probability that the first two parts selected are defective and the third is not defective?

This probability can be described in shorthand notation as $P(d_1 d_2 n_3)$, where d and n denote parts that are defective and not defective, respectively. Here

$$P(d_1 d_2 n_3) = P(n_3 \mid d_1 d_2) P(d_1 d_2) = P(n_3 \mid d_1 d_2) P(d_2 \mid d_1) P(d_1) = \frac{47}{48} \cdot \frac{2}{49} \cdot \frac{3}{50} = 0.0024$$

The probabilities for the first and second selections are similar to those in the previous example. The $P(n_3|d_1 d_2)$ is based on the fact that after the first 2 parts are selected, 1 defective and 47 nondefective parts remain.

When the probability is written to account for the order of the selections, it is easy to solve this question from the definition of conditional probability. There are other ways to express the probability, such as $P(d_1 d_2 n_3) = P(d_2|d_1 n_3) P(d_1 n_3)$. However, such alternatives do not lead to conditional probabilities that can be easily calculated.

Exercises FOR SECTION 2-4

⊕ Problem available in *WileyPLUS* at instructor's discretion.

⊕ **Go Tutorial** Tutoring problem available in *WileyPLUS* at instructor's discretion

2-99. ⊕ Disks of polycarbonate plastic from a supplier are analyzed for scratch and shock resistance. The results from 100 disks are summarized as follows:

		Shock Resistance	
		High	Low
Scratch	High	70	9
Resistance	Low	16	5

Let A denote the event that a disk has high shock resistance, and let B denote the event that a disk has high scratch resistance. Determine the following probabilities:

(a) $P(A)$ (b) $P(B)$

(c) $P(A \mid B)$ (d) $P(B \mid A)$

2-100. ⊕ Samples of skin experiencing desquamation are analyzed for both moisture and melanin content. The results from 100 skin samples are as follows:

		Melanin Content	
		High	Low
Moisture	High	13	7
Content	Low	48	32

Let A denote the event that a sample has low melanin content, and let B denote the event that a sample has high moisture content. Determine the following probabilities:

(a) $P(A)$ (b) $P(B)$

(c) $P(A \mid B)$ (d) $P(B \mid A)$

2-101. ⊕ The analysis of results from a leaf transmutation experiment (turning a leaf into a petal) is summarized by type of transformation completed:

		Total Textural Transformation	
		Yes	No
Total Color	Yes	243	26
Transformation	No	13	18

(a) If a leaf completes the color transformation, what is the probability that it will complete the textural transformation?

(b) If a leaf does not complete the textural transformation, what is the probability it will complete the color transformation?

2-102. ⊕ Samples of a cast aluminum part are classified on the basis of surface finish (in microinches) and length measurements. The results of 100 parts are summarized as follows:

		Length	
		Excellent	Good
Surface	Excellent	80	2
Finish	Good	10	8

Let A denote the event that a sample has excellent surface finish, and let B denote the event that a sample has excellent length. Determine:

(a) $P(A)$ (b) $P(B)$

(c) $P(A \mid B)$ (d) $P(B \mid A)$

(e) If the selected part has excellent surface finish, what is the probability that the length is excellent?
(f) If the selected part has good length, what is the probability that the surface finish is excellent?

2-103. The following table summarizes the analysis of samples of galvanized steel for coating weight and surface roughness:

		Coating Weight	
		High	Low
Surface	High	12	16
Roughness	Low	88	34

(a) If the coating weight of a sample is high, what is the probability that the surface roughness is high?
(b) If the surface roughness of a sample is high, what is the probability that the coating weight is high?
(c) If the surface roughness of a sample is low, what is the probability that the coating weight is low?

2-104. Consider the data on wafer contamination and location in the sputtering tool shown in Table 2-2. Assume that one wafer is selected at random from this set. Let A denote the event that a wafer contains four or more particles, and let B denote the event that a wafer is from the center of the sputtering tool. Determine:

(a) $P(A)$ (b) $P(A \mid B)$
(c) $P(B)$ (d) $P(B \mid A)$
(e) $P(A \cap B)$ (f) $P(A \cup B)$

2-105. The following table summarizes the number of deceased beetles under autolysis (the destruction of a cell after its death by the action of its own enzymes) and putrefaction (decomposition of organic matter, especially protein, by microorganisms, resulting in production of foul-smelling matter):

		Autolysis	
		High	Low
Putrefaction	High	14	59
	Low	18	9

(a) If the autolysis of a sample is high, what is the probability that the putrefaction is low?
(b) If the putrefaction of a sample is high, what is the probability that the autolysis is high?
(c) If the putrefaction of a sample is low, what is the probability that the autolysis is low?

2-106. A maintenance firm has gathered the following information regarding the failure mechanisms for air conditioning systems:

		Evidence of Gas Leaks	
		Yes	No
Evidence of electrical failure	Yes	55	17
	No	32	3

The units without evidence of gas leaks or electrical failure showed other types of failure. If this is a representative sample of AC failure, find the probability

(a) That failure involves a gas leak
(b) That there is evidence of electrical failure given that there was a gas leak
(c) That there is evidence of a gas leak given that there is evidence of electrical failure

2-107. A lot of 100 semiconductor chips contains 20 that are defective. Two are selected randomly, without replacement, from the lot.

(a) What is the probability that the first one selected is defective?
(b) What is the probability that the second one selected is defective given that the first one was defective?
(c) What is the probability that both are defective?
(d) How does the answer to part (b) change if chips selected were replaced prior to the next selection?

2-108. A batch of 500 containers for frozen orange juice contains 5 that are defective. Two are selected, at random, without replacement from the batch.

(a) What is the probability that the second one selected is defective given that the first one was defective?
(b) What is the probability that both are defective?
(c) What is the probability that both are acceptable?
Three containers are selected, at random, without replacement, from the batch.
(d) What is the probability that the third one selected is defective given that the first and second ones selected were defective?
(e) What is the probability that the third one selected is defective given that the first one selected was defective and the second one selected was okay?
(f) What is the probability that all three are defective?

2-109. A batch of 350 samples of rejuvenated mitochondria contains 8 that are mutated (or defective). Two are selected from the batch, at random, without replacement.

(a) What is the probability that the second one selected is defective given that the first one was defective?
(b) What is the probability that both are defective?
(c) What is the probability that both are acceptable?

2-110. A computer system uses passwords that are exactly seven characters and each character is one of the 26 letters (a–z) or 10 integers (0–9). You maintain a password for this computer system. Let A denote the subset of passwords that begin with a vowel (either a, e, i, o, or u) and let B denote the subset of passwords that end with an even number (either 0, 2, 4, 6, or 8).

(a) Suppose a hacker selects a password at random. What is the probability that your password is selected?
(b) Suppose a hacker knows that your password is in event A and selects a password at random from this subset. What is the probability that your password is selected?
(c) Suppose a hacker knows that your password is in A and B and selects a password at random from this subset. What is the probability that your password is selected?

2-111. If $P(A \mid B) = 1$, must $A = B$? Draw a Venn diagram to explain your answer.

2-112. Suppose A and B are mutually exclusive events. Construct a Venn diagram that contains the three events A, B, and C such that $P(A \mid C) = 1$ and $P(B \mid C) = 0$.

2-113. Consider the endothermic reactions in Exercise 2-50. Let A denote the event that a reaction's final temperature is 271 K or less. Let B denote the event that the heat absorbed is above target. Determine the following probabilities.

(a) $P(A \mid B)$ (b) $P(A' \mid B)$
(c) $P(A \mid B')$ (d) $P(B \mid A)$

2-114. Consider the hospital emergency room data in Example 2-8. Let A denote the event that a visit is to hospital 4, and let B denote the event that a visit results in LWBS (at any hospital). Determine the following probabilities.

(a) $P(A \mid B)$ (b) $P(A' \mid B)$
(c) $P(A \mid B')$ (d) $P(B \mid A)$

2-115. Consider the well failure data in Exercise 2-53.

(a) What is the probability of a failure given there are more than 1,000 wells in a geological formation?
(b) What is the probability of a failure given there are fewer than 500 wells in a geological formation?

2-116. An article in the *The Canadian Entomologist* (Harcourt et al., 1977, Vol. 109, pp. 1521–1534) reported on the life of the alfalfa weevil from eggs to adulthood. The following table shows the number of larvae that survived at each stage of development from eggs to adults.

Eggs	Early Larvae	Late Larvae	Pre-pupae	Late Pupae	Adults
421	412	306	45	35	31

(a) What is the probability an egg survives to adulthood?
(b) What is the probability of survival to adulthood given survival to the late larvae stage?
(c) What stage has the lowest probability of survival to the next stage?

2-117. Consider the bar code in Example 2-12. Suppose that all 40 codes are equally likely (none is held back as a delimiter). Determine the probability for each of the following:

(a) The second bar is wide given that the first bar is wide.
(b) The third bar is wide given that the first two bars are not wide.
(c) The first bar is wide given that the last bar is wide.

2-118. Suppose that a patient is selected randomly from those described in Exercise 2-57. Let A denote the event that the patient is treated with ribavirin plus interferon alfa, and let B denote the event that the response is complete. Determine the following probabilities:

(a) $P(B \mid A)$ (b) $P(A \mid B)$
(c) $P(A \mid B')$ (d) $P(A' \mid B)$

2-119. Suppose that a patient is selected randomly from those described in Exercise 2-98. Let A denote the event that the patient is in group 1, and let B denote the event that there is no progression. Determine the following probabilities:

(a) $P(B \mid A)$ (b) $P(A \mid B)$
(c) $P(A \mid B')$ (d) $P(A' \mid B)$

2-120. A computer system uses passwords that contain exactly eight characters, and each character is one of the 26 lowercase letters (a–z) or 26 uppercase letters (A–Z) or 10 integers (0–9). Let Ω denote the set of all possible passwords. Suppose that all passwords in Ω are equally likely. Determine the probability for each of the following:

(a) Password contains all lowercase letters given that it contains only letters
(b) Password contains at least 1 uppercase letter given that it contains only letters
(c) Password contains only even numbers given that is contains all numbers

2-5 Multiplication and Total Probability Rules

The probability of the intersection of two events is often needed. The conditional probability definition in Equation 2-9 can be rewritten to provide a formula known as the **multiplication rule** for probabilities.

Multiplication Rule

$$P(A \cap B) = P(B \mid A)P(A) = P(A \mid B)P(B) \qquad (2\text{-}10)$$

The last expression in Equation 2-10 is obtained by interchanging A and B.

Example 2-26 **Machining Stages** The probability that the first stage of a numerically controlled machining operation for high-rpm pistons meets specifications is 0.90. Failures are due to metal variations, fixture alignment, cutting blade condition, vibration, and ambient environmental conditions. Given that the first stage meets specifications, the probability that a second stage of machining meets specifications is 0.95. What is the probability that both stages meet specifications?

Let A and B denote the events that the first and second stages meet specifications, respectively. The probability requested is

$$P(A \cap B) = P(B \mid A)P(A) = 0.95(0.90) = 0.855$$

Although it is also true that $P(A \cap B) = P(A \mid B)P(B)$, the information provided in the problem does not match this second formulation.

Practical Interpretation: The probability that both stages meet specifications is approximately 0.85, and if additional stages were needed to complete a piston, the probability would decrease further. Consequently, the probability that each stage is completed successfully needs to be large in order for a piston to meet all specifications.

Sometimes the probability of an event is given under each of several conditions. With enough of these conditional probabilities, the probability of the event can be recovered. For example, suppose that in semiconductor manufacturing, the probability is 0.10 that a chip subjected to high levels of contamination during manufacturing causes a product failure. The probability is 0.005 that a chip not subjected to high contamination levels during manufacturing causes a product failure. In a particular production run, 20% of the chips are subject to high levels of contamination. What is the probability that a product using one of these chips fails?

Clearly, the requested probability depends on whether or not the chip was exposed to high levels of contamination. For any event B, we can write B as the union of the part of B in A and the part of B in A'. That is,

$$B = (A \cap B) \cup (A' \cap B)$$

This result is shown in the Venn diagram in Fig. 2-15. Because A and A' are mutually exclusive, $A \cap B$ and $A' \cap B$ are mutually exclusive. Therefore, from the probability of the union of mutually exclusive events in Equation 2-6 and the multiplication rule in Equation 2-10, the following **total probability rule** is obtained.

Total Probability Rule (Two Events)

For any events A and B,

$$P(B) = P(B \cap A) + P(B \cap A') = P(B \mid A)P(A) + P(B \mid A')P(A') \qquad (2\text{-}11)$$

Example 2-27 **Semiconductor Contamination** Consider the contamination discussion at the start of this section. The information is summarized here.

Probability of Failure	Level of Contamination	Probability of Level
0.1	High	0.2
0.005	Not high	0.8

Let F denote the event that the product fails, and let H denote the event that the chip is exposed to high levels of contamination. The requested probability is $P(F)$, and the information provided can be represented as

$$P(F \mid H) = 0.10 \quad \text{and} \quad P(F \mid H') = 0.005$$

$$P(H) = 0.20 \quad \text{and} \quad P(H') = 0.80$$

From Equation 2-11,

$$P(F) = 0.10(0.20) + 0.005(0.80) = 0.024$$

which can be interpreted as just the weighted average of the two probabilities of failure.

The reasoning used to develop Equation 2-11 can be applied more generally. Because $A \cup A' = S$, we know $(A \cap B) \cup (A' \cap B)$ equals B, and because $A \cap A' = \phi$, we know $A \cap B$ and $A' \cap B$ are mutually exclusive. In general, a collection of sets $E_1, E_2, \ldots, E_k$ such that $E_1 \cup E_2 \cup \ldots \cup E_k = S$ is said to be **exhaustive**. A graphical display of partitioning an event B among a collection of mutually exclusive and exhaustive events is shown in Fig. 2-16.

Total Probability Rule (Multiple Events)

Assume $E_1, E_2, \ldots, E_k$ are k mutually exclusive and exhaustive sets. Then

$$\begin{aligned} P(B) &= P(B \cap E_1) + P(B \cap E_2) + \cdots + P(B \cap E_k) \\ &= P(B \mid E_1)P(E_1) + P(B \mid E_2)P(E_2) + \cdots + P(B \mid E_k)P(E_k) \end{aligned} \qquad (2\text{-}12)$$

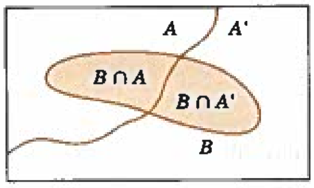
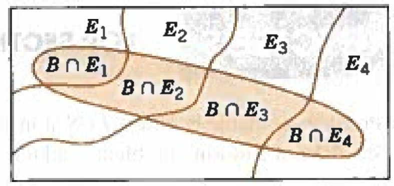


FIGURE 2-15 Partitioning an event into two mutually exclusive subsets.



$B = (B \cap E_1) \cup (B \cap E_2) \cup (B \cap E_3) \cup (B \cap E_4)$

FIGURE 2-16 Partitioning an event into several mutually exclusive subsets.

Example 2-28 **Semiconductor Failures** Continuing with semiconductor manufacturing, assume the following probabilities for product failure subject to levels of contamination in manufacturing:

Probability of Failure	Level of Contamination
0.10	High
0.01	Medium
0.001	Low

In a particular production run, 20% of the chips are subjected to high levels of contamination, 30% to medium levels of contamination, and 50% to low levels of contamination. What is the probability that a product using one of these chips fails? Let

- H denote the event that a chip is exposed to high levels of contamination
- M denote the event that a chip is exposed to medium levels of contamination
- L denote the event that a chip is exposed to low levels of contamination

Then,

$$P(F) = P(F \mid H)P(H) + P(F \mid M)P(M) + P(F \mid L)P(L)$$
$$= 0.10(0.20) + 0.01(0.30) + 0.001(0.50) = 0.0235$$

The calculations are conveniently organized with the tree diagram in Fig. 2-17.

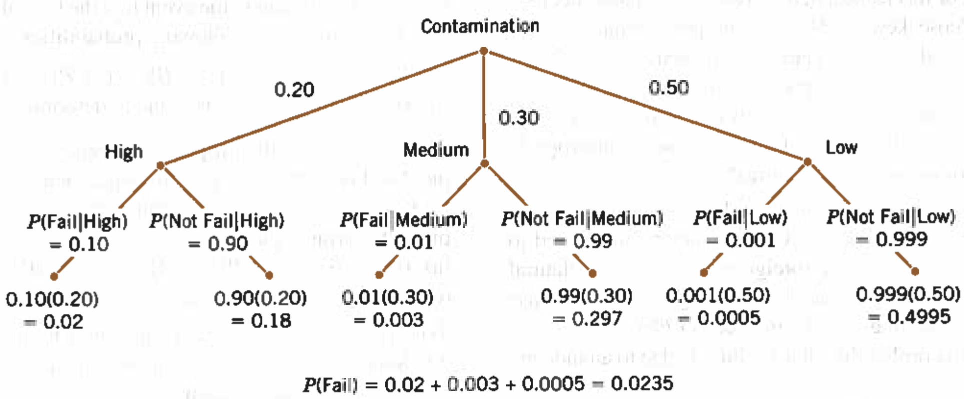


FIGURE 2-17 Tree diagram for Example 2-28.

Exercises FOR SECTION 2-5

⊕ Problem available in *WileyPLUS* at instructor's discretion.
⊕ **Go Tutorial** Tutoring problem available in *WileyPLUS* at instructor's discretion

2-121. ⊕ Suppose that $P(A \mid B) = 0.4$ and $P(B) = 0.5$. Determine the following:
(a) $P(A \cap B)$ (b) $P(A' \cap B)$

2-122. ⊕ Suppose that $P(A \mid B) = 0.2$, $P(A \mid B') = 0.3$, and $P(B) = 0.8$. What is $P(A)$?

2-123. ⊕ The probability is 1% that an electrical connector that is kept dry fails during the warranty period of a portable computer. If the connector is ever wet, the probability of a failure during the warranty period is 5%. If 90% of the connectors are kept dry and 10% are wet, what proportion of connectors fail during the warranty period?

2-124. ⊕ Suppose 2% of cotton fabric rolls and 3% of nylon fabric rolls contain flaws. Of the rolls used by a manufacturer, 70% are cotton and 30% are nylon. What is the probability that a randomly selected roll used by the manufacturer contains flaws?

2-125. ⊕ The edge roughness of slit paper products increases as knife blades wear. Only 1% of products slit with new blades have rough edges, 3% of products slit with blades of average sharpness exhibit roughness, and 5% of products slit with worn blades exhibit roughness. If 25% of the blades in manufacturing are new, 60% are of average sharpness, and 15% are worn, what is the proportion of products that exhibit edge roughness?

2-126. ⊕ In the 2012 presidential election, exit polls from the critical state of Ohio provided the following results:

Total	Obama	Romney
No college degree (60%)	52%	45%
College graduate (40%)	47%	51%

What is the probability a randomly selected respondent voted for Obama?

2-127. ⊕ Computer keyboard failures are due to faulty electrical connects (12%) or mechanical defects (88%). Mechanical defects are related to loose keys (27%) or improper assembly (73%). Electrical connect defects are caused by defective wires (35%), improper connections (13%), or poorly welded wires (52%).
(a) Find the probability that a failure is due to loose keys.
(b) Find the probability that a failure is due to improperly connected or poorly welded wires.

2-128. ⊕ Heart failures are due to either natural occurrences (87%) or outside factors (13%). Outside factors are related to induced substances (73%) or foreign objects (27%). Natural occurrences are caused by arterial blockage (56%), disease (27%), and infection (e.g., staph infection) (17%).
(a) Determine the probability that a failure is due to an induced substance.
(b) Determine the probability that a failure is due to disease or infection.

2-129. ⊕ A batch of 25 injection-molded parts contains 5 parts that have suffered excessive shrinkage.
(a) If two parts are selected at random, and without replacement, what is the probability that the second part selected is one with excessive shrinkage?
(b) If three parts are selected at random, and without replacement, what is the probability that the third part selected is one with excessive shrinkage?

2-130. ⊕ A lot of 100 semiconductor chips contains 20 that are defective.
(a) Two are selected, at random, without replacement, from the lot. Determine the probability that the second chip selected is defective.
(b) Three are selected, at random, without replacement, from the lot. Determine the probability that all are defective.

2-131. An article in the *British Medical Journal* ["Comparison of treatment of renal calculi by operative surgery, percutaneous nephrolithotomy, and extracorporeal shock wave lithotripsy" (1986, Vol. 82, pp. 879–892)] provided the following discussion of success rates in kidney stone removals. Open surgery had a success rate of 78% (273/350) and a newer method, percutaneous nephrolithotomy (PN), had a success rate of 83% (289/350). This newer method looked better, but the results changed when stone diameter was considered. For stones with diameters less than 2 centimeters, 93% (81/87) of cases of open surgery were successful compared with only 83% (234/270) of cases of PN. For stones greater than or equal to 2 centimeters, the success rates were 73% (192/263) and 69% (55/80) for open surgery and PN, respectively. Open surgery is better for both stone sizes, but less successful in total. In 1951, E. H. Simpson pointed out this apparent contradiction (known as **Simpson's paradox**), and the hazard still persists today. Explain how open surgery can be better for both stone sizes but worse in total.

2-132. Consider the endothermic reactions in Exercise 2-50. Let A denote the event that a reaction's final temperature is 271 K or less. Let B denote the event that the heat absorbed is above target. Determine the following probabilities.
(a) $P(A \cap B)$ (b) $P(A \cup B)$ (c) $P(A' \cup B')$
(d) Use the total probability rule to determine $P(A)$

2-133. Consider the hospital emergency room data in Example 2-8. Let A denote the event that a visit is to hospital 4 and let B denote the event that a visit results in LWBS (at any hospital). Determine the following probabilities.
(a) $P(A \cap B)$ (b) $P(A \cup B)$ (c) $P'(A' \cup B')$
(d) Use the total probability rule to determine $P(A)$

2-134. Consider the hospital emergency room data in Example 2-8. Suppose that three visits that resulted in LWBS are selected randomly (without replacement) for a follow-up interview.
(a) What is the probability that all three are selected from hospital 2?
(b) What is the probability that all three are from the same hospital?

2-135. Consider the well failure data in Exercise 2-53. Let A denote the event that the geological formation has more than 1000 wells, and let B denote the event that a well failed. Determine the following probabilities:

(a) $P(A \cap B)$ (b) $P(A \cup B)$ (c) $P(A' \cup B')$

(d) Use the total probability rule to determine $P(A)$

2-136. Consider the well failure data in Exercise 2-53. Suppose that two failed wells are selected randomly (without replacement) for a follow-up review.

(a) What is the probability that both are from the gneiss geological formation group?

(b) What is the probability that both are from the same geological formation group?

2-137. A Web ad can be designed from four different colors, three font types, five font sizes, three images, and five text phrases. A specific design is randomly generated by the Web server when you visit the site. Determine the probability that the ad color is red and the font size is not the smallest one.

2-138. Consider the code in Example 2-12. Suppose that all 40 codes are equally likely (none is held back as a delimiter). Determine the probability for each of the following:

(a) The code starts and ends with a wide bar.

(b) Two wide bars occur consecutively.

(c) Two consecutive wide bars occur at the start or end.

(d) The middle bar is wide.

2-139. Similar to the hospital schedule in Example 2-11, suppose that an operating room needs to schedule three knee, four hip, and five shoulder surgeries. Assume that all schedules are equally likely. Determine the following probabilities:

(a) All hip surgeries are completed first given that all knee surgeries are last.

(b) The schedule begins with a hip surgery given that all knee surgeries are last.

(c) The first and last surgeries are hip surgeries given that knee surgeries are scheduled in time periods 2 through 4.

(d) The first two surgeries are hip surgeries given that all knee surgeries are last.

2-140. Suppose that a patient is selected randomly from those described in Exercise 2-98. Let A denote the event that the patient is in group 1, and let B denote the event for which there is no progression. Determine the following probabilities:

(a) $P(A \cap B)$ (b) $P(B)$

(c) $P(A' \cap B)$ (d) $P(A \cup B)$ (e) $P(A' \cup B)$

2-141. A computer system uses passwords that contain exactly eight characters, and each character is one of the 26 lowercase letters (a–z) or 26 uppercase letters (A–Z) or 10 integers (0–9). Let Ω denote the set of all possible password, and let A and B denote the events that consist of passwords with only letters or only integers, respectively. Suppose that all passwords in Ω are equally likely. Determine the following robabilities:

(a) $P(A|B')$

(b) $P(A' \cap B)$

(c) P (password contains exactly 2 integers given that it contains at least 1 integer)

2-6 Independence

In some cases, the conditional probability of $P(B \mid A)$ might equal $P(B)$. In this special case, knowledge that the outcome of the experiment is in event A does not affect the probability that the outcome is in event B.

Example 2-29 **Sampling with Replacement** Consider the inspection described in Example 2-14. Six parts are selected randomly from a bin of 50 parts, but assume that the selected part is replaced before the next one is selected. The bin contains 3 defective parts and 47 nondefective parts. What is the probability that the second part is defective given that the first part is defective?

In shorthand notation, the requested probability is $P(B|A)$, where A and B denote the events that the first and second parts are defective, respectively. Because the first part is replaced prior to selecting the second part, the bin still contains 50 parts, of which 3 are defective. Therefore, the probability of B does not depend on whether or not the first part is defective. That is,

$$P(B|A) = \frac{3}{50}$$

Also, the probability that both parts are defective is

$$P(A \cap B) = P(B \mid A)P(A) = \frac{3}{50} \cdot \frac{3}{50} = \frac{9}{2500}$$

Example 2-30 **Flaws and Functions** The information in Table 2-3 related surface flaws to functionally defective parts. In that case, we determined that $P(D \mid F) = 10/40 = 0.25$ and $P(D) = 28/400 = 0.07$. Suppose that the situation is different and follows Table 2-4. Then,

$$P(D \mid F) = 2/40 = 0.05 \quad \text{and} \quad P(D) = 20/400 = 0.05$$

That is, the probability that the part is defective does not depend on whether it has surface flaws. Also,

$$P(F \mid D) = 2/20 = 0.10 \quad \text{and} \quad P(F) = 40/400 = 0.10$$

so the probability of a surface flaw does not depend on whether the part is defective. Furthermore, the definition of conditional probability implies that

$$P(F \cap D) = P(D \mid F)P(F)$$

but in the special case of this problem,

$$P(F \cap D) = P(D)P(F) = \frac{2}{40} \cdot \frac{2}{20} = \frac{1}{200}$$

TABLE • 2-4 Parts Classified

		Surface Flaws		
		Yes (event F)	No	Total
Defective	Yes (event D)	2	18	20
	No	38	342	380
	Total	40	360	400

The preceding example illustrates the following conclusions. In the special case that $P(B \mid A) = P(B)$, we obtain

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

and

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

These conclusions lead to an important definition.

Independence (two events)

Two events are **independent** if any one of the following equivalent statements is true:

(1) $P(A \mid B) = P(A)$

(2) $P(B \mid A) = P(B)$

(3) $P(A \cap B) = P(A)P(B)$ (2-13)

It is left as a mind-expanding exercise to show that independence implies related results such as

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

The concept of independence is an important relationship between events and is used throughout this text. A mutually exclusive relationship between two events is based only on the outcomes that compose the events. However, an independence relationship depends on the probability model used for the random experiment. Often, independence is assumed to be part of the random experiment that describes the physical system under study.

Example 2-31 Consider the inspection described in Example 2-14. Six parts are selected randomly without replacement from a bin of 50 parts. The bin contains 3 defective parts and 47 nondefective parts. Let A and B denote the events that the first and second parts are defective, respectively.

We suspect that these two events are not independent because the knowledge that the first part is defective suggests that it is less likely that the second part selected is defective. Indeed, $P(B \mid A) = 2/49$. Now, what is $P(B)$? Finding the unconditional $P(B)$ takes some work because the possible values of the first selection need to be considered:

$$P(B) = P(B \mid A)P(A) + P(B \mid A')P(A')$$

$$= \frac{2}{49} \cdot \frac{3}{50} + \frac{3}{49} \cdot \frac{47}{50} = \frac{3}{50}$$

Interestingly, $P(B)$, the unconditional probability that the second part selected is defective, without any knowledge of the first part, is the same as the probability that the first part selected is defective. Yet our goal is to assess independence. Because $P(B|A)$ does not equal $P(B)$, the two events are not independent, as we expected.

When considering three or more events, we can extend the definition of independence with the following general result.

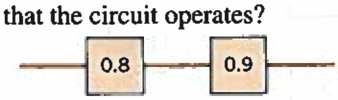
Independence (multiple events)

The events $E_1, E_2, \ldots, E_n$ are independent if and only if for any subset of these events

$$P(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_k}) = P(E_{i_1}) \times P(E_{i_2}) \times \cdots \times P(E_{i_k}) \quad (2\text{-}14)$$

This definition is typically used to calculate the probability that several events occur, assuming that they are independent and the individual event probabilities are known. The knowledge that the events are independent usually comes from a fundamental understanding of the random experiment.

Example 2-32 **Series Circuit** The following circuit operates only if there is a path of functional devices from left to right. The probability that each device functions is shown on the graph. Assume that devices fail independently. What is the probability that the circuit operates?



Let L and R denote the events that the left and right devices operate, respectively. There is a path only if both operate. The probability that the circuit operates is

$$P(L \text{ and } R) = P(L \cap R) = P(L)P(R) = 0.80(0.90) = 0.72$$

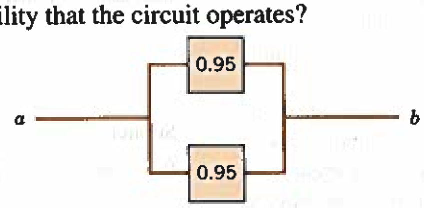
Practical Interpretation: Notice that the probability that the circuit operates degrades to approximately 0.5 when all devices are required to be functional. The probability that each device is functional needs to be large for a circuit to operate when many devices are connected in series.

Example 2-33 **Semiconductor Wafers** Assume that the probability that a wafer contains a large particle of contamination is 0.01 and that the wafers are independent; that is, the probability that a wafer contains a large particle does not depend on the characteristics of any of the other wafers. If 15 wafers are analyzed, what is the probability that no large particles are found?

Let E_i denote the event that the ith wafer contains no large particles, $i = 1, 2, \ldots, 15$. Then, $P(E_i) = 0.99$. The probability requested can be represented as $P(E_1 \cap E_2 \cap \cdots \cap E_{15})$. From the independence assumption and Equation 2-14,

$$P(E_1 \cap E_2 \cap \cdots \cap E_{15}) = P(E_1) \times P(E_2) \times \cdots \times P(E_{15}) = 0.99^{15} = 0.86$$

Example 2-34 **Parallel Circuit** The following circuit operates only if there is a path of functional devices from left to right. The probability that each device functions is shown on the graph. Assume that devices fail independently. What is the probability that the circuit operates?



Let T and B denote the events that the top and bottom devices operate, respectively. There is a path if at least one device operates. The probability that the circuit operates is

$$P(T \text{ or } B) = 1 - P[(T \text{ or } B)'] = 1 - P(T' \text{ and } B')$$

A simple formula for the solution can be derived from the complements T' and B'. From the independence assumption,

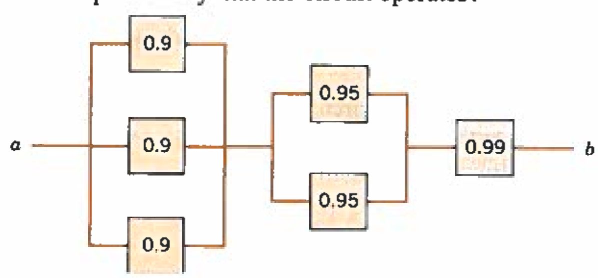
$$P(T' \text{ and } B') = P(T')P(B') = (1-0.95)^2 = 0.05^2$$

so

$$P(T \text{ or } B) = 1 - 0.05^2 = 0.9975$$

Practical Interpretation: Notice that the probability that the circuit operates is larger than the probability that either device is functional. This is an advantage of a parallel architecture. A disadvantage is that multiple devices are needed.

Example 2-35 Advanced Circuit The following circuit operates only if there is a path of functional devices from left to right. The probability that each device functions is shown on the graph. Assume that devices fail independently. What is the probability that the circuit operates?



The solution can be obtained from a partition of the graph into three columns. Let L denote the event that there is a path of functional devices only through the three units on the left. From the independence and based upon the previous example,

$$P(L) = 1 - 0.1^3$$

Similarly, let M denote the event that there is a path of functional devices only through the two units in the middle. Then,

$$P(M) = 1 - 0.05^2$$

The probability that there is a path of functional devices only through the one unit on the right is simply the probability that the device functions, namely, 0.99. Therefore, with the independence assumption used again, the solution is

$$(1-0.1^3)(1-0.05^2)(0.99) = 0.987$$

Exercises FOR SECTION 2-6

⊕ Problem available in *WileyPLUS* at instructor's discretion.
⊕ Go Tutorial Tutoring problem available in *WileyPLUS* at instructor's discretion

2-142. ⊕ If $P(A|B) = 0.4$, $P(B) = 0.8$, and $P(A) = 0.5$, are the events A and B independent?

2-143. ⊕ If $P(A|B) = 0.3$, $P(B) = 0.8$, and $P(A) = 0.3$, are the events B and the complement of A independent?

2-144. ⊕ If $P(A) = 0.2$, $P(B) = 0.2$, and A and B are mutually exclusive, are they independent?

2-145. ⊕ A batch of 500 containers of frozen orange juice contains 5 that are defective. Two are selected, at random, without replacement, from the batch. Let A and B denote the events that the first and second containers selected are defective, respectively.
(a) Are A and B independent events?
(b) If the sampling were done with replacement, would A and B be independent?

2-146. ⊕ Disks of polycarbonate plastic from a supplier are analyzed for scratch and shock resistance. The results from 100 disks are summarized as follows:

		Shock Resistance	
		High	Low
Scratch	High	70	9
Resistance	Low	16	5

Let A denote the event that a disk has high shock resistance, and let B denote the event that a disk has high scratch resistance. Are events A and B independent?

2-147. Samples of emissions from three suppliers are classified for conformance to air-quality specifications. The results from 100 samples are summarized as follows:

		Conforms	
		Yes	No
	1	22	8
Supplier	2	25	5
	3	30	10

Let A denote the event that a sample is from supplier 1, and let B denote the event that a sample conforms to specifications.

(a) Are events A and B independent?

(b) Determine $P(B \mid A)$.

2-148. ⊕ Redundant array of inexpensive disks (RAID) is a technology that uses multiple hard drives to increase the speed of data transfer and provide instant data backup. Suppose that the probability of any hard drive failing in a day is 0.001 and the drive failures are independent.

(a) A RAID 0 scheme uses two hard drives, each containing a mirror image of the other. What is the probability of data loss? Assume that data loss occurs if both drives fail within the same day.

(b) A RAID 1 scheme splits the data over two hard drives. What is the probability of data loss? Assume that data loss occurs if at least one drive fails within the same day.

2-149. ⊕ The probability that a lab specimen contains high levels of contamination is 0.10. Five samples are checked, and the samples are independent.

(a) What is the probability that none contain high levels of contamination?

(b) What is the probability that exactly one contains high levels of contamination?

(c) What is the probability that at least one contains high levels of contamination?

2-150. ⊕ In a test of a printed circuit board using a random test pattern, an array of 10 bits is equally likely to be 0 or 1. Assume the bits are independent.

(a) What is the probability that all bits are 1s?

(b) What is the probability that all bits are 0s?

(c) What is the probability that exactly 5 bits are 1s and 5 bits are 0s?

2-151. Six tissues are extracted from an ivy plant infested by spider mites. The plant in infested in 20% of its area. Each tissue is chosen from a randomly selected area on the ivy plant.

(a) What is the probability that four successive samples show the signs of infestation?

(b) What is the probability that three out of four successive samples show the signs of infestation?

2-152. ⊕ A player of a video game is confronted with a series of four opponents and an 80% probability of defeating each opponent. Assume that the results from opponents are independent (and that when the player is defeated by an opponent the game ends).

(a) What is the probability that a player defeats all four opponents in a game?

(b) What is the probability that a player defeats at least two opponents in a game?

(c) If the game is played three times, what is the probability that the player defeats all four opponents at least once?

2-153. In an acid-base titration, a base or acid is gradually added to the other until they have completely neutralized each other. Because acids and bases are usually colorless (as are the water and salt produced in the neutralization reaction), pH is measured to monitor the reaction. Suppose that the equivalence point is reached after approximately 100 mL of an NaOH solution has been added (enough to react with all the acetic acid present) but that replicates are equally likely to indicate from 95 to 104 mL, measured to the nearest mL. Assume that two technicians each conduct titrations independently.

(a) What is the probability that both technicians obtain equivalence at 100 mL?

(b) What is the probability that both technicians obtain equivalence between 98 and 104 mL (inclusive)?

(c) What is the probability that the average volume at equivalence from the technicians is 100 mL?

2-154. A credit card contains 16 digits. It also contains the month and year of expiration. Suppose there are 1 million credit card holders with unique card numbers. A hacker randomly selects a 16-digit credit card number.

(a) What is the probability that it belongs to a user?

(b) Suppose a hacker has a 25% chance of correctly guessing the year your card expires and randomly selects 1 of the 12 months. What is the probability that the hacker correctly selects the month and year of expiration?

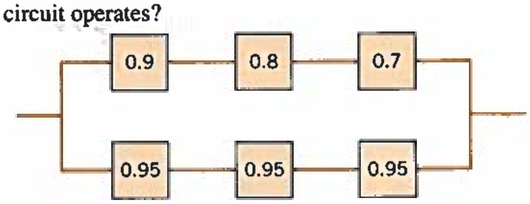
2-155. ⊕ Eight cavities in an injection-molding tool produce plastic connectors that fall into a common stream. A sample is chosen every several minutes. Assume that the samples are independent.

(a) What is the probability that five successive samples were all produced in cavity 1 of the mold?

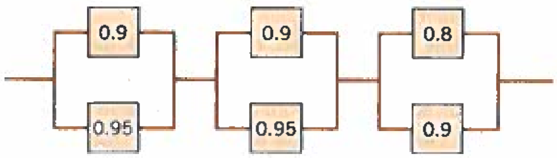
(b) What is the probability that five successive samples were all produced in the same cavity of the mold?

(c) What is the probability that four out of five successive samples were produced in cavity 1 of the mold?

2-156. The following circuit operates if and only if there is a path of functional devices from left to right. The probability that each device functions is as shown. Assume that the probability that a device is functional does not depend on whether or not other devices are functional. What is the probability that the circuit operates?



2-157. The following circuit operates if and only if there is a path of functional devices from left to right. The probability that each device functions is as shown. Assume that the probability that a device is functional does not depend on whether or not other devices are functional. What is the probability that the circuit operates?



2-158. Consider the endothermic reactions in Exercise 2-50. Let A denote the event that a reaction's final temperature is 271 K or less. Let B denote the event that the heat absorbed is above target. Are these events independent?

2-159. Consider the hospital emergency room data in Example 2-8. Let A denote the event that a visit is to hospital 4, and let B denote the event that a visit results in LWBS (at any hospital). Are these events independent?

2-160. Consider the well failure data in Exercise 2-53. Let A denote the event that the geological formation has more than 1000 wells, and let B denote the event that a well failed. Are these events independent?

2-161. A Web ad can be designed from four different colors, three font types, five font sizes, three images, and five text phrases. A specific design is randomly generated by the Web server when you visit the site. Let A denote the event that the design color is red, and let B denote the event that the font size is not the smallest one. Are A and B independent events? Explain why or why not.

2-162. Consider the code in Example 2-12. Suppose that all 40 codes are equally likely (none is held back as a delimiter). Let A and B denote the event that the first bar is wide and B denote the event that the second bar is wide. Determine the following:

(a) $P(A)$ (b) $P(B)$ (c) $P(A \cap B)$
(d) Are A and B independent events?

2-163. An integrated circuit contains 10 million logic gates (each can be a logical AND or OR circuit). Assume the probability of a gate failure is p and that the failures are independent. The integrated circuit fails to function if any gate fails. Determine the value for p so that the probability that the integrated circuit functions is 0.95.

2-164. Table 2-1 provides data on wafers categorized by location and contamination levels. Let A denote the event that contamination is *low*, and let B denote the event that the location is *center.* Are A and B independent? Why or why not?

2-165. Table 2-1 provides data on wafers categorized by location and contamination levels. More generally, let the number of wafers with *low* contamination from the *center* and *edge* locations be denoted as n_{lc} and n_{le}, respectively. Similarly, let n_{hc} and n_{he} denote the number of wafers with *high* contamination from the *center* and *edge* locations, respectively. Suppose that $n_{lc} = 10n_{hc}$ and $n_{le} = 10n_{he}$. That is, there are 10 times as many *low* contamination wafers as *high* ones from each location. Let A denote the event that contamination is low, and let B denote the event that the location is *center.* Are A and B independent? Does your conclusion change if the multiplier of 10 (between *low* and *high* contamination wafers) is changed from 10 to another positive integer?

2-7 Bayes' Theorem

The examples in this chapter indicate that information is often presented in terms of conditional probabilities. These conditional probabilities commonly provide the probability of an event (such as failure) given a condition (such as high or low contamination). But after a random experiment generates an outcome, we are naturally interested in the probability that a condition was present (high contamination) given an outcome (a semiconductor failure). Thomas Bayes addressed this essential question in the 1700s and developed the fundamental result known as **Bayes' theorem.** Do not let the simplicity of the mathematics conceal the importance. There is extensive interest in such probabilities in modern statistical analysis.

From the definition of conditional probability,

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

Now, considering the second and last terms in the preceding expression, we can write

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)} \quad \text{for} \quad P(B) > 0 \qquad (2\text{-}15)$$

This is a useful result that enables us to solve for $P(A \mid B)$ in terms of $P(B \mid A)$.

and chlorinated compounds—instead of having to use a single test for each pollutant. The makers of the test claim that it can detect high levels of organic pollutants with 99.7% accuracy, volatile solvents with 99.95% accuracy, and chlorinated compounds with 89.7% accuracy. If a pollutant is not present, the test does not signal. Samples are prepared for the calibration of the test and 60% of them are contaminated with organic pollutants, 27% with volatile solvents, and 13% with traces of chlorinated compounds. A test sample is selected randomly.

(a) What is the probability that the test will signal?

(b) If the test signals, what is the probability that chlorinated compounds are present?

2-174. Consider the endothermic reactions in Exercise 2-50. Use Bayes' theorem to calculate the probability that a reaction's final temperature is 271 K or less given that the heat absorbed is above target.

2-175. Consider the hospital emergency room data in Example 2-8. Use Bayes' theorem to calculate the probability that a person visits hospital 4 given they are LWBS.

2-176. Consider the well failure data in Exercise 2-53. Use Bayes' theorem to calculate the probability that a randomly selected well is in the gneiss group given that the well has failed.

2-177. Two Web colors are used for a site advertisement. If a site visitor arrives from an affiliate, the probabilities of the blue or green colors being used in the advertisement are 0.8 and 0.2, respectively. If the site visitor arrives from a search site, the probabilities of blue and green colors in the advertisement are 0.4 and 0.6, respectively. The proportions of visitors from affiliates and search sites are 0.3 and 0.7, respectively. What is the probability that a visitor is from a search site given that the blue ad was viewed?

2-178. Suppose that a patient is selected randomly from those described in Exercise 2-98. Let A denote the event that the patient is in group 1, and let B denote the event that there is no progression. Determine the following probabilities:

(a) $P(B)$ (b) $P(B|A)$ (c) $P(A|B)$

2-179. An e-mail filter is planned to separate valid e-mails from spam. The word *free* occurs in 60% of the spam messages and only 4% of the valid messages. Also, 20% of the messages are spam. Determine the following probabilities:

(a) The message contains *free*.

(b) The message is spam given that it contains *free*.

(c) The message is valid given that it does not contain *free*.

2-180. A recreational equipment supplier finds that among orders that include tents, 40% also include sleeping mats. Only 5% of orders that do not include tents do include sleeping mats. Also, 20% of orders include tents. Determine the following probabilities:

(a) The order includes sleeping mats.

(b) The order includes a tent given it includes sleeping mats.

2-181. The probabilities of poor print quality given no printer problem, misaligned paper, high ink viscosity, or printer-head debris are 0, 0.3, 0.4, and 0.6, respectively. The probabilities of no printer problem, misaligned paper, high ink viscosity, or printer-head debris are 0.8, 0.02, 0.08, and 0.1, respectively.

(a) Determine the probability of high ink viscosity given poor print quality.

(b) Given poor print quality, what problem is most likely?

2-8 Random Variables

We often summarize the outcome from a random experiment by a simple number. In many of the examples of random experiments that we have considered, the sample space has been a description of possible outcomes. In some cases, descriptions of outcomes are sufficient, but in other cases, it is useful to associate a number with each outcome in the sample space. Because the particular outcome of the experiment is not known in advance, the resulting value of our variable is not known in advance. For this reason, the variable that associates a number with the outcome of a random experiment is referred to as a **random variable**.

Random Variable

> A **random variable** is a function that assigns a real number to each outcome in the sample space of a random experiment.

Notation is used to distinguish between a random variable and the real number.

Notation

> A random variable is denoted by an uppercase letter such as X. After an experiment is conducted, the measured value of the random variable is denoted by a lowercase letter such as $x = 70$ milliamperes.

Sometimes a measurement (such as current in a copper wire or length of a machined part) can assume any value in an interval of real numbers (at least theoretically). Then arbitrary precision in the measurement is possible. Of course, in practice, we might round off to the nearest tenth or hundredth of a unit. The random variable that represents this measurement is said to be a **continuous random variable.** The range of the random variable includes all values in an interval of real numbers; that is, the range can be thought of as a continuum.

In other experiments, we might record a count such as the number of transmitted bits that are received in error. Then, the measurement is limited to integers. Or we might record that a proportion such as 0.0042 of the 10,000 transmitted bits were received in error. Then, the measurement is fractional, but it is still limited to discrete points on the real line. Whenever the measurement is limited to discrete points on the real line, the random variable is said to be a **discrete random variable.**

Discrete and Continuous Random Variables

A **discrete random variable** is a random variable with a finite (or countably infinite) range.

A **continuous random variable** is a random variable with an interval (either finite or infinite) of real numbers for its range.

In some cases, the random variable X is actually discrete but, because the range of possible values is so large, it might be more convenient to analyze X as a continuous random variable. For example, suppose that current measurements are read from a digital instrument that displays the current to the nearest 100th of a milliampere. Because the possible measurements are limited, the random variable is discrete. However, it might be a more convenient, simple approximation to assume that the current measurements are values of a continuous random variable.

Examples of Random Variables

Examples of **continuous** random variables:
electrical current, length, pressure, temperature, time, voltage, weight

Examples of **discrete** random variables:
number of scratches on a surface, proportion of defective parts among 1000 tested, number of transmitted bits received in error

Exercises FOR SECTION 2-8

⊕ Problem available in *WileyPLUS* at instructor's discretion.
⊕ Go Tutorial Tutoring problem available in *WileyPLUS* at instructor's discretion

2-182. Decide whether a discrete or continuous random variable is the best model for each of the following variables:
(a) The time until a projectile returns to earth.
(b) The number of times a transistor in a computer memory changes state in one operation.
(c) The volume of gasoline that is lost to evaporation during the filling of a gas tank.
(d) The outside diameter of a machined shaft.

2-183. Decide whether a discrete or continuous random variable is the best model for each of the following variables:
(a) The number of cracks exceeding one-half inch in 10 miles of an interstate highway.
(b) The weight of an injection-molded plastic part.
(c) The number of molecules in a sample of gas.
(d) The concentration of output from a reactor.
(e) The current in an electronic circuit.

2-184. Decide whether a discrete or continuous random variable is the best model for each of the following variables:
(a) The time for a computer algorithm to assign an image to a category.
(b) The number of bytes used to store a file in a computer.
(c) The ozone concentration in micrograms per cubic meter.
(d) The ejection fraction (volumetric fraction of blood pumped from a heart ventricle with each beat).
(e) The fluid flow rate in liters per minute.

Supplemental Exercises

⊕ Problem available in *WileyPLUS* at instructor's discretion.

⊕ **Go Tutorial** Tutoring problem available in *WileyPLUS* at instructor's discretion

2-185. ⊕ Samples of laboratory glass are in small, light packaging or heavy, large packaging. Suppose that 2% and 1%, respectively, of the sample shipped in small and large packages, respectively, break during transit. If 60% of the samples are shipped in large packages and 40% are shipped in small packages, what proportion of samples break during shipment?

2-186. A sample of three calculators is selected from a manufacturing line, and each calculator is classified as either defective or acceptable. Let A, B, and C denote the events that the first, second, and third calculators, respectively, are defective.

(a) Describe the sample space for this experiment with a tree diagram.

Use the tree diagram to describe each of the following events:

(b) A (c) B (d) $A \cap B$ (e) $B \cup C$

2-187. ⊕ Samples of a cast aluminum part are classified on the basis of surface finish (in microinches) and edge finish. The results of 100 parts are summarized as follows:

		Edge Finish	
		Excellent	Good
Surface	Excellent	80	2
Finish	Good	10	8

Let A denote the event that a sample has excellent surface finish, and let B denote the event that a sample has excellent edge finish. If a part is selected at random, determine the following probabilities:

(a) $P(A)$ (b) $P(B)$ (c) $P(A')$

(d) $P(A \cap B)$ (e) $P(A \cup B)$ (f) $P(A' \cup B)$

2-188. Shafts are classified in terms of the machine tool that was used for manufacturing the shaft and conformance to surface finish and roundness.

Tool 1		Roundness Conforms	
		Yes	No
Surface Finish	Yes	200	1
Conforms	No	4	2

Tool 2		Roundness Conforms	
		Yes	No
Surface Finish	Yes	145	4
Conforms	No	8	6

(a) If a shaft is selected at random, what is the probability that the shaft conforms to surface finish requirements or to roundness requirements or is from tool 1?

(b) If a shaft is selected at random, what is the probability that the shaft conforms to surface finish requirements or does not conform to roundness requirements or is from tool 2?

(c) If a shaft is selected at random, what is the probability that the shaft conforms to both surface finish and roundness requirements or the shaft is from tool 2?

(d) If a shaft is selected at random, what is the probability that the shaft conforms to surface finish requirements or the shaft is from tool 2?

2-189. If A, B, and C are mutually exclusive events, is it possible for $P(A) = 0.3$, $P(B) = 0.4$, and $P(C) = 0.5$? Why or why not?

2-190. ⊕ The analysis of shafts for a compressor is summarized by conformance to specifications:

		Roundness Conforms	
		Yes	No
Surface finish	Yes	345	5
Conforms	No	12	8

(a) If we know that a shaft conforms to roundness requirements, what is the probability that it conforms to surface finish requirements?

(b) If we know that a shaft does not conform to roundness requirements, what is the probability that it conforms to surface finish requirements?

2-191. ⊕ A researcher receives 100 containers of oxygen. Of those containers, 20 have oxygen that is not ionized, and the rest are ionized. Two samples are randomly selected, without replacement, from the lot.

(a) What is the probability that the first one selected is not ionized?

(b) What is the probability that the second one selected is not ionized given that the first one was ionized?

(c) What is the probability that both are ionized?

(d) How does the answer in part (b) change if samples selected were replaced prior to the next selection?

2-192. ⊕ A lot contains 15 castings from a local supplier and 25 castings from a supplier in the next state. Two castings are selected randomly, without replacement, from the lot of 40. Let A be the event that the first casting selected is from the local supplier, and let B denote the event that the second casting is selected from the local supplier. Determine:

(a) $P(A)$ (b) $P(B \mid A)$ (c) $P(A \cap B)$ (d) $P(A \cup B)$

Suppose that 3 castings are selected at random, without replacement, from the lot of 40. In addition to the definitions of events A and B, let C denote the event that the third casting selected is from the local supplier. Determine:

(e) $P(A \cap B \cap C)$ (f) $P(A \cap B \cap C')$

2-193. ⊕ In the manufacturing of a chemical adhesive, 3% of all batches have raw materials from two different lots. This occurs when holding tanks are replenished and the remaining portion of a lot is insufficient to fill the tanks.

Only 5% of batches with material from a single lot require reprocessing. However, the viscosity of batches consisting of two or more lots of material is more difficult to control, and 40% of such batches require additional processing to achieve the required viscosity.

Let A denote the event that a batch is formed from two different lots, and let B denote the event that a lot requires additional processing. Determine the following probabilities:
(a) $P(A)$ (b) $P(A')$ (c) $P(B|A)$ (d) $P(B|A')$
(e) $P(A \cap B)$ (f) $P(A \cap B')$ (g) $P(B)$

2-194. Incoming calls to a customer service center are classified as complaints (75% of calls) or requests for information (25% of calls). Of the complaints, 40% deal with computer equipment that does not respond and 57% deal with incomplete software installation; in the remaining 3% of complaints, the user has improperly followed the installation instructions. The requests for information are evenly divided on technical questions (50%) and requests to purchase more products (50%).

(a) What is the probability that an incoming call to the customer service center will be from a customer who has not followed installation instructions properly?
(b) Find the probability that an incoming call is a request for purchasing more products.

2-195. A congested computer network has a 0.002 probability of losing a data packet, and packet losses are independent events. A lost packet must be resent.

(a) What is the probability that an e-mail message with 100 packets will need to be resent?
(b) What is the probability that an e-mail message with 3 packets will need exactly 1 to be resent?
(c) If 10 e-mail messages are sent, each with 100 packets, what is the probability that at least 1 message will need some packets to be resent?

2-196. Samples of a cast aluminum part are classified on the basis of surface finish (in microinches) and length measurements. The results of 100 parts are summarized as follows:

		Length	
		Excellent	Good
Surface	Excellent	80	2
Finish	Good	10	8

Let A denote the event that a sample has excellent surface finish, and let B denote the event that a sample has excellent length. Are events A and B independent?

2-197. Go Tutorial An optical storage device uses an error recovery procedure that requires an immediate satisfactory readback of any written data. If the readback is not successful after three writing operations, that sector of the disk is eliminated as unacceptable for data storage. On an acceptable portion of the disk, the probability of a satisfactory readback is 0.98. Assume the readbacks are independent. What is the probability that an acceptable portion of the disk is eliminated as unacceptable for data storage?

2-198. Semiconductor lasers used in optical storage products require higher power levels for write operations than for read operations. High-power-level operations lower the useful life of the laser.

Lasers in products used for backup of higher-speed magnetic disks primarily write, and the probability that the useful life exceeds five years is 0.95. Lasers that are in products that are used for main storage spend approximately an equal amount of time reading and writing, and the probability that the useful life exceeds five years is 0.995. Now, 25% of the products from a manufacturer are used for backup and 75% of the products are used for main storage.

Let A denote the event that a laser's useful life exceeds five years, and let B denote the event that a laser is in a product that is used for backup.

Use a tree diagram to determine the following:
(a) $P(B)$ (b) $P(A|B)$ (c) $P(A|B')$
(d) $P(A \cap B)$ (e) $P(A \cap B')$ (f) $P(A)$
(g) What is the probability that the useful life of a laser exceeds five years?
(h) What is the probability that a laser that failed before five years came from a product used for backup?

2-199. Energy released from cells breaks the molecular bond and converts ATP (adenosine triphosphate) into ADP (adenosine diphosphate). Storage of ATP in muscle cells (even for an athlete) can sustain maximal muscle power only for less than five seconds (a short dash). Three systems are used to replenish ATP—phosphagen system, glycogen-lactic acid system (anaerobic), and aerobic respiration—but the first is useful only for less than 10 seconds, and even the second system provides less than two minutes of ATP. An endurance athlete needs to perform below the anaerobic threshold to sustain energy for extended periods. A sample of 100 individuals is described by the energy system used in exercise at different intensity levels.

	Primarily Aerobic	
Period	Yes	No
1	50	7
2	13	30

Let A denote the event that an individual is in period 2, and let B denote the event that the energy is primarily aerobic. Determine the number of individuals in
(a) $A' \cap B$ (b) B' (c) $A \cup B$

2-200. A sample preparation for a chemical measurement is completed correctly by 25% of the lab technicians, completed with a minor error by 70%, and completed with a major error by 5%.

(a) If a technician is selected randomly to complete the preparation, what is the probability that it is completed without error?
(b) What is the probability that it is completed with either a minor or a major error?

2-201. In circuit testing of printed circuit boards, each board either fails or does not fail the test. A board that fails the test is then checked further to determine which one of five defect types is the primary failure mode. Represent the sample space for this experiment.

2-202. The data from 200 machined parts are summarized as follows:

	Depth of Bore	
Edge Condition	Above Target	Below Target
Coarse	15	10
Moderate	25	20
Smooth	50	80

(a) What is the probability that a part selected has a moderate edge condition and a below-target bore depth?
(b) What is the probability that a part selected has a moderate edge condition or a below-target bore depth?
(c) What is the probability that a part selected does not have a moderate edge condition or does not have a below-target bore depth?

2-203. ⊕ Computers in a shipment of 100 units contain a portable hard drive, solid-state memory, or both, according to the following table:

	Portable Hard Drive	
Solid-state memory	Yes	No
Yes	15	80
No	4	1

Let A denote the event that a computer has a portable hard drive, and let B denote the event that a computer has a solid-state memory. If one computer is selected randomly, compute

(a) $P(A)$ (b) $P(A \cap B)$ (c) $P(A \cup B)$
(d) $P(A' \cap B)$ (e) $P(A \mid B)$

2-204. ⊕ The probability that a customer's order is not shipped on time is 0.05. A particular customer places three orders, and the orders are placed far enough apart in time that they can be considered to be independent events.
(a) What is the probability that all are shipped on time?
(b) What is the probability that exactly one is not shipped on time?
(c) What is the probability that two or more orders are not shipped on time?

2-205. Let E_1, E_2, and E_3 denote the samples that conform to a percentage of solids specification, a molecular weight specification, and a color specification, respectively. A total of 240 samples are classified by the E_1, E_2, and E_3 specifications, where *yes* indicates that the sample conforms.

E_3 yes

		E_2		
		Yes	No	Total
E_1	Yes	200	1	201
	No	5	4	9
Total		205	5	210

E_3 no

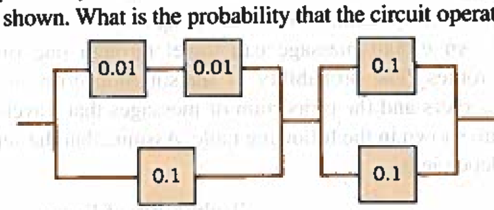
		E_2		
		Yes	No	Total
E_1	Yes	20	4	24
	No	6	0	6
Total		26	4	30

(a) Are E_1, E_2, and E_3 mutually exclusive events?
(b) Are E_1', E_2', and E_3' mutually exclusive events?
(c) What is $P(E_1'$ or E_2' or $E_3')$?
(d) What is the probability that a sample conforms to all three specifications?
(e) What is the probability that a sample conforms to the E_1 or E_3 specification?
(f) What is the probability that a sample conforms to the E_1 or E_2 or E_3 specification?

2-206. ⊕ Transactions to a computer database are either new items or changes to previous items. The addition of an item can be completed in less than 100 milliseconds 90% of the time, but only 20% of changes to a previous item can be completed in less than this time. If 30% of transactions are changes, what is the probability that a transaction can be completed in less than 100 milliseconds?

2-207. ⊕ A steel plate contains 20 bolts. Assume that 5 bolts are not torqued to the proper limit. 4 bolts are selected at random, without replacement, to be checked for torque.
(a) What is the probability that all 4 of the selected bolts are torqued to the proper limit?
(b) What is the probability that at least 1 of the selected bolts is *not* torqued to the proper limit?

2-208. The following circuit operates if and only if there is a path of functional devices from left to right. Assume devices fail independently and that the probability of *failure* of each device is as shown. What is the probability that the circuit operates?



2-209. ⊕ The probability that concert tickets are available by telephone is 0.92. For the same event, the probability that tickets are available through a Web site is 0.95. Assume that these two ways to buy tickets are independent. What is the probability that someone who tries to buy tickets through the Web and by phone will obtain tickets?

2-210. The British government has stepped up its information campaign regarding foot-and-mouth disease by mailing brochures to farmers around the country. It is estimated that 99% of Scottish farmers who receive the brochure possess enough information to deal with an outbreak of the disease, but only 90% of those without the brochure can deal with an outbreak. After the first three months of mailing, 95% of the farmers in Scotland had received the informative brochure. Compute the probability that a randomly selected farmer will have enough information to deal effectively with an outbreak of the disease.

2-211. ⊕ In an automated filling operation, the probability of an incorrect fill when the process is operated at a low speed is 0.001. When the process is operated at a high speed, the probability of an incorrect fill is 0.01. Assume that 30% of the containers are filled when the process is operated at a high speed and the remainder are filled when the process is operated at a low speed.
(a) What is the probability of an incorrectly filled container?
(b) If an incorrectly filled container is found, what is the probability that it was filled during the high-speed operation?

2-212. ⊕ An encryption-decryption system consists of three elements: encode, transmit, and decode. A faulty encode occurs in 0.5% of the messages processed, transmission errors occur

in 1% of the messages, and a decode error occurs in 0.1% of the messages. Assume the errors are independent.
(a) What is the probability of a completely defect-free message?
(b) What is the probability of a message that has either an encode or a decode error?

2-213. ⊕ It is known that two defective copies of a commercial software program were erroneously sent to a shipping lot that now has a total of 75 copies of the program. A sample of copies will be selected from the lot without replacement.
(a) If three copies of the software are inspected, determine the probability that exactly one of the defective copies will be found.
(b) If three copies of the software are inspected, determine the probability that both defective copies will be found.
(c) If 73 copies are inspected, determine the probability that both copies will be found. (*Hint*: Work with the copies that remain in the lot.)

2-214. ⊕ A robotic insertion tool contains 10 primary components. The probability that any component fails during the warranty period is 0.01. Assume that the components fail independently and that the tool fails if any component fails. What is the probability that the tool fails during the warranty period?

2-215. An e-mail message can travel through one of two server routes. The probability of transmission error in each of the servers and the proportion of messages that travel each route are shown in the following table. Assume that the servers are independent.

		Probability of Error			
	Percentage of Messages	Server 1	Server 2	Server 3	Server 4
Route 1	30	0.01	0.015	—	—
Route 2	70	—	—	0.02	0.003

(a) What is the probability that a message will arrive without error?
(b) If a message arrives in error, what is the probability it was sent through route 1?

2-216. A machine tool is idle 15% of the time. You request immediate use of the tool on five different occasions during the year. Assume that your requests represent independent events.
(a) What is the probability that the tool is idle at the time of all of your requests?
(b) What is the probability that the machine is idle at the time of exactly four of your requests?
(c) What is the probability that the tool is idle at the time of at least three of your requests?

2-217. ⊕ A lot of 50 spacing washers contains 30 washers that are thicker than the target dimension. Suppose that 3 washers are selected at random, without replacement, from the lot.
(a) What is the probability that all 3 washers are thicker than the target?
(b) What is the probability that the third washer selected is thicker than the target if the first 2 washers selected are thinner than the target?
(c) What is the probability that the third washer selected is thicker than the target?

2-218. Continuing Exercise 2-217, washers are selected from the lot at random without replacement.
(a) What is the minimum number of washers that need to be selected so that the probability that all the washers are thinner than the target is less than 0.10?
(b) What is the minimum number of washers that need to be selected so that the probability that 1 or more washers are thicker than the target is at least 0.90?

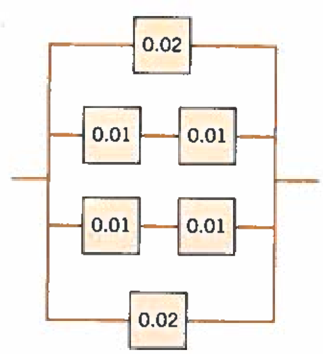
2-219. ⊕ The following table lists the history of 940 orders for features in an entry-level computer product.

		Extra Memory	
		No	Yes
Optional high-speed processor	No	514	68
	Yes	112	246

Let A be the event that an order requests the optional high-speed processor, and let B be the event that an order requests extra memory. Determine the following probabilities:
(a) $P(A \cup B)$ (b) $P(A \cap B)$
(c) $P(A' \cup B)$ (d) $P(A' \cap B')$
(e) What is the probability that an order requests an optional high-speed processor given that the order requests extra memory?
(f) What is the probability that an order requests extra memory given that the order requests an optional high-speed processor?

2-220. ⊕ The alignment between the magnetic media and head in a magnetic storage system affects the system's performance. Suppose that 10% of the read operations are degraded by skewed alignments, 5% of the read operations are degraded by off-center alignments, and the remaining read operations are properly aligned. The probability of a read error is 0.01 from a skewed alignment, 0.02 from an off-center alignment, and 0.001 from a proper alignment.
(a) What is the probability of a read error?
(b) If a read error occurs, what is the probability that it is due to a skewed alignment?

2-221. The following circuit operates if and only if there is a path of functional devices from left to right. Assume that devices fail independently and that the probability of *failure* of each device is as shown. What is the probability that the circuit does not operate?



2-222. A company that tracks the use of its Web site determined that the more pages a visitor views, the more likely the visitor is to provide contact information. Use the following tables to answer the questions:

Number of pages viewed:	1	2	3	4 or more
Percentage of visitors:	40	30	20	10
Percentage of visitors in each page-view category that provides contact information:	10	10	20	40

(a) What is the probability that a visitor to the Web site provides contact information?
(b) If a visitor provides contact information, what is the probability that the visitor viewed four or more pages?

2-223. An article in *Genome Research* ["An Assessment of Gene Prediction Accuracy in Large DNA Sequences" (2000, Vol. 10, pp. 1631–1642)], considered the accuracy of commercial software to predict nucleotides in gene sequences. The following table shows the number of sequences for which the programs produced predictions and the number of nucleotides correctly predicted (computed globally from the total number of prediction successes and failures on all sequences).

	Number of Sequences	Proportion
GenScan	177	0.93
Blastx default	175	0.91
Blastx topcomboN	174	0.97
Blastx 2 stages	175	0.90
GeneWise	177	0.98
Procrustes	177	0.93

Assume the prediction successes and failures are independent among the programs.
(a) What is the probability that all programs predict a nucleotide correctly?
(b) What is the probability that all programs predict a nucleotide incorrectly?
(c) What is the probability that at least one Blastx program predicts a nucleotide correctly?

2-224. ⊕ A batch contains 36 bacteria cells. Assume that 12 of the cells are not capable of cellular replication. Of the cells, 6 are selected at random, without replacement, to be checked for replication.
(a) What is the probability that all 6 of the selected cells are able to replicate?
(b) What is the probability that at least 1 of the selected cells is not capable of replication?

2-225. A computer system uses passwords that are exactly seven characters, and each character is one of the 26 letters (a–z) or 10 integers (0–9). Uppercase letters are not used.
(a) How many passwords are possible?
(b) If a password consists of exactly 6 letters and 1 number, how many passwords are possible?
(c) If a password consists of 5 letters followed by 2 numbers, how many passwords are possible?

2-226. Natural red hair consists of two genes. People with red hair have two dominant genes, two regressive genes, or one dominant and one regressive gene. A group of 1000 people was categorized as follows:

	Gene 2		
Gene 1	Dominant	Regressive	Other
Dominant	5	25	30
Regressive	7	63	35
Other	20	15	800

Let A denote the event that a person has a dominant red hair gene, and let B denote the event that a person has a regressive red hair gene. If a person is selected at random from this group, compute the following:
(a) $P(A)$ (b) $P(A \cap B)$ (c) $P(A \cup B)$
(d) $P(A' \cap B)$ (e) $P(A \mid B)$
(f) Probability that the selected person has red hair

2-227. ⊕ Two suppliers each supplied 2000 parts that were evaluated for conformance to specifications. One part type was more complex than the other. The proportion of nonconforming parts of each type are shown in the table.

Supplier		Simple Component	Complex Assembly	Total
1	Nonconforming	2	10	12
	Total	1000	1000	2000
2	Nonconforming	4	6	10
	Total	1600	400	2000

One part is selected at random from each supplier. For each supplier, separately calculate the following probabilities:
(a) What is the probability a part conforms to specifications?
(b) What is the probability a part conforms to specifications given it is a complex assembly?
(c) What is the probability a part conforms to specifications given it is a simple component?
(d) Compare your answers for each supplier in part (a) to those in parts (b) and (c) and explain any unusual results.

2-228. Consider the treatments in Exercise 2-57. Suppose a patient is selected randomly. Let A denote the event that the patient is treated with ribavirin plus interferon alfa or interferon alfa, and let B denote the event that the response is complete. Determine the following probabilities.
(a) $P(A \mid B)$ (b) $P(B \mid A)$ (c) $P(A \cap B)$ (d) $P(A \cup B)$

2-229. Consider the patient groups in Exercise 2-98. Suppose a patient is selected randomly. Let A denote the event that the patient is in group 1 or 2, and let B denote the event that there is no progression. Determine the following probabilities:

(a) $P(A \mid B)$ (b) $P(B \mid A)$ (c) $P(A \cap B)$ (d) $P(A \cup B)$

Mind-expanding exercises

2-230. Suppose documents in a lending organization are selected randomly (without replacement) for review. In a set of 50 documents, suppose that 2 actually contain errors.

(a) What is the minimum sample size such that the probability exceeds 0.90 that at least 1 document in error is selected?

(b) Comment on the effectiveness of sampling inspection to detect errors.

2-231. Suppose that a lot of washers is large enough that it can be assumed that the sampling is done with replacement. Assume that 60% of the washers exceed the target thickness.

(a) What is the minimum number of washers that need to be selected so that the probability that none is thicker than the target is less than 0.10?

(b) What is the minimum number of washers that need to be selected so that the probability that 1 or more washers are thicker than the target is at least 0.90?

2-232. A biotechnology manufacturing firm can produce diagnostic test kits at a cost of $20. Each kit for which there is a demand in the week of production can be sold for $100. However, the half-life of components in the kit requires the kit to be scrapped if it is not sold in the week of production. The cost of scrapping the kit is $5. The weekly demand is summarized as follows:

	Weekly Demand			
Number of units	0	50	100	200
Probability of demand	0.05	0.4	0.3	0.25

How many kits should be produced each week to maximize the firm's mean earnings?

2-233. Assume the following characteristics of the inspection process in Exercise 2-207. If an operator checks a bolt, the probability that an incorrectly torqued bolt is identified is 0.95. If a checked bolt is correctly torqued, the operator's conclusion is always correct. What is the probability that at least one bolt in the sample of four is identified as being incorrectly torqued?

2-234. If the events A and B are independent, show that A' and B' are independent.

2-235. Suppose that a table of part counts is generalized as follows:

		Conforms	
		Yes	No
Supplier	1	ka	kb
	2	a	b

where a, b, and k are positive integers. Let A denote the event that a part is from supplier 1, and let B denote the event that a part conforms to specifications. Show that A and B are independent events.

This exercise illustrates the result that whenever the rows of a table (with r rows and c columns) are proportional, an event defined by a row category and an event defined by a column category are independent.

Important Terms and Concepts

Addition rule
Axioms of probability
Bayes' theorem
Combination
Conditional probability
Counting techniques
Equally likely outcomes
Event
Independence
Multiplication rule
Mutually exclusive events
Outcome
Permutation
Probability
Random samples
Random variables—discrete and continuous
Sample spaces—discrete and continuous
Simpson's paradox
Total probability rule
Tree diagram
Venn diagram
With or without replacement

Example 3-3 Define the random variable X to be the number of contamination particles on a wafer in semiconductor manufacturing. Although wafers possess a number of characteristics, the random variable X summarizes the wafer only in terms of the number of particles.

The possible values of X are integers from zero up to some large value that represents the maximum number of particles that can be found on one of the wafers. If this maximum number is large, we might simply assume that the range of X is the set of integers from zero to infinity.

Note that more than one random variable can be defined on a sample space. In Example 3-3, we might also define the random variable Y to be the number of chips from a wafer that fails the final test.

Exercises FOR SECTION 3-1

⊕ Problem available in *WileyPLUS* at instructor's discretion.
⊕ **Go Tutorial** Tutoring problem available in *WileyPLUS* at instructor's discretion

For each of the following exercises, determine the range (possible values) of the random variable.

3-1. ⊕ The random variable is the number of nonconforming solder connections on a printed circuit board with 1000 connections.

3-2. ⊕ In a voice communication system with 50 lines, the random variable is the number of lines in use at a particular time.

3-3. ⊕ An electronic scale that displays weights to the nearest pound is used to weigh packages. The display shows only five digits. Any weight greater than the display can indicate is shown as 99999. The random variable is the displayed weight.

3-4. ⊕ A batch of 500 machined parts contains 10 that do not conform to customer requirements. The random variable is the number of parts in a sample of five parts that do not conform to customer requirements.

3-5. ⊕ A batch of 500 machined parts contains 10 that do not conform to customer requirements. Parts are selected successively, without replacement, until a nonconforming part is obtained. The random variable is the number of parts selected.

3-6. ⊕ The random variable is the moisture content of a lot of raw material, measured to the nearest percentage point.

3-7. The random variable is the number of surface flaws in a large coil of galvanized steel.

3-8. The random variable is the number of computer clock cycles required to complete a selected arithmetic calculation.

3-9. ⊕ An order for an automobile can select the base model or add any number of 15 options. The random variable is the number of options selected in an order.

3-10. ⊕ Wood paneling can be ordered in thicknesses of 1/8, 1/4, or 3/8 inch. The random variable is the total thickness of paneling in two orders.

3-11. ⊕ A group of 10,000 people are tested for a gene called Ifi202 that has been found to increase the risk for lupus. The random variable is the number of people who carry the gene.

3-12. ⊕ In an acid-base titration, the milliliters of base that are needed to reach equivalence are measured to the nearest milliliter between 0.1 and 0.15 liters (inclusive).

3-13. ⊕ The number of mutations in a nucleotide sequence of length 40,000 in a DNA strand after exposure to radiation is measured. Each nucleotide may be mutated.

3-14. A healthcare provider schedules 30 minutes for each patient's visit, but some visits require extra time. The random variable is the number of patients treated in an eight-hour day.

3-15. A Web site contains 100 interconnected pages. The random variable is the number of unique pages viewed by a visitor to the Web site.

3-2 Probability Distributions and Probability Mass Functions

Random variables are so important in random experiments that sometimes we essentially ignore the original sample space of the experiment and focus on the probability distribution of the random variable. For example, in Example 3-1, our analysis might focus exclusively on the integers $\{0, 1, \ldots, 48\}$ in the range of X. In Example 3-2, we might summarize the random experiment in terms of the three possible values of X, namely $\{0, 1, 2\}$. In this manner, a random variable can simplify the description and analysis of a random experiment.

The **probability distribution** of a random variable X is a description of the probabilities associated with the possible values of X. For a discrete random variable, the distribution is often specified by just a list of the possible values along with the probability of each. In some cases, it is convenient to express the probability in terms of a formula.

Example 3-4 **Digital Channel** There is a chance that a bit transmitted through a digital transmission channel is received in error. Let X equal the number of bits in error in the next four bits transmitted. The possible values for X are $\{0, 1, 2, 3, 4\}$. Based on a model for the errors that is presented in the following section, probabilities for these values will be determined. Suppose that the probabilities are

$$P(X = 0) = 0.6561 \qquad P(X = 1) = 0.2916$$
$$P(X = 2) = 0.0486 \qquad P(X = 3) = 0.0036$$
$$P(X = 4) = 0.0001$$

The probability distribution of X is specified by the possible values along with the probability of each. A graphical description of the probability distribution of X is shown in Fig. 3-1.

Practical Interpretation: A random experiment can often be summarized with a random variable and its distribution. The details of the sample space can often be omitted.

Suppose that a loading on a long, thin beam places mass only at discrete points. See Fig. 3-2. The loading can be described by a function that specifies the mass at each of the discrete points. Similarly, for a discrete random variable X, its distribution can be described by a function that specifies the probability at each of the possible discrete values for X.

Probability Mass Function

For a discrete random variable X with possible values $x_1, x_2, \ldots, x_n$, a **probability mass function** is a function such that

(1) $f(x_i) \geq 0$

(2) $\sum_{i=1}^{n} f(x_i) = 1$

(3) $f(x_i) = P(X = x_i)$ (3-1)

For the bits in error in Example 3-4, $f(0) = 0.6561$, $f(1) = 0.2916$, $f(2) = 0.0486$, $f(3) = 0.0036$, and $f(4) = 0.0001$. Check that the probabilities sum to 1.

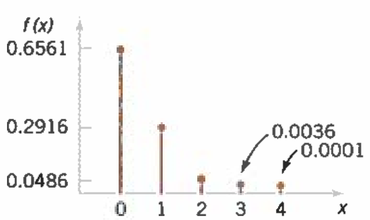


FIGURE 3-1 Probability distribution for bits in error.

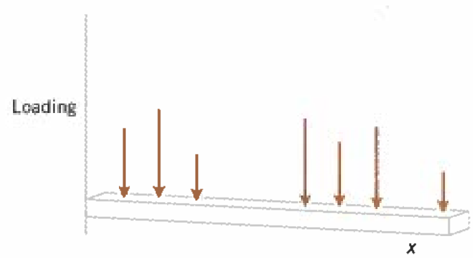


FIGURE 3-2 Loadings at discrete points on a long, thin beam.

Example 3-5 **Wafer Contamination** Let the random variable X denote the number of semiconductor wafers that need to be analyzed in order to detect a large particle of contamination. Assume that the probability that a wafer contains a large particle is 0.01 and that the wafers are independent. Determine the probability distribution of X.

Let p denote a wafer in which a large particle is present, and let a denote a wafer in which it is absent. The sample space of the experiment is infinite, and it can be represented as all possible sequences that start with a string of a's and end with p. That is,

$$s = \{p, ap, aap, aaap, aaaap, aaaaap, \text{and so forth}\}$$

Consider a few special cases. We have $P(X=1)=P(p)=0.01$. Also, using the independence assumption,

$$P(X=2)=P(ap)=0.99(0.01)=0.0099$$

A general formula is

$$P(X=x)=P(\underbrace{aa\ldots a}_{(x-1)a's}p)=0.99^{x-1}(0.01),\ \text{for } x=1,2,3,\ldots$$

Describing the probabilities associated with X in terms of this formula is a simple method to define the distribution of X in this example. Clearly $f(x)\geq 0$. The fact that the sum of the probabilities is 1 is left as an exercise. This is an example of a geometric random variable for which details are provided later in this chapter.

Practical Interpretation: The random experiment here has an unbounded number of outcomes, but it can still be conveniently modeled with a discrete random variable with a (countably) infinite range.

Exercises FOR SECTION 3-2

⊕ Problem available in *WileyPLUS* at instructor's discretion.

⊕ **Go Tutorial** Tutoring problem available in *WileyPLUS* at instructor's discretion

3-16. The sample space of a random experiment is $\{a, b, c, d, e, f\}$, and each outcome is equally likely. A random variable is defined as follows:

outcome	a	b	c	d	e	f
x	0	0	1.5	1.5	2	3

Determine the probability mass function of a. Use the probability mass function to determine the following probabilities:

(a) $P(X=1.5)$ (b) $P(0.5<X<2.7)$
(c) $P(X>3)$ (d) $P(0\leq X<2)$
(e) $P(X=0 \text{ or } X=2)$

For Exercises 3-17 to 3-21, verify that the following functions are probability mass functions, and determine the requested probabilities.

3-17. ⊕

x	−2	−1	0	1	2
$f(x)$	0.2	0.4	0.1	0.2	0.1

(a) $P(X\leq 2)$ (b) $P(X>-2)$
(c) $P(-1\leq X\leq 1)$ (d) $P(X\leq -1 \text{ or } X=2)$

3-18. ⊕ $f(x)=(8/7)(1/2)^x,\quad x=1,2,3$
(a) $P(X\leq 1)$ (b) $P(X>1)$
(c) $P(2<X<6)$ (d) $P(X\leq 1 \text{ or } X>1)$

3-19. ⊕ $f(x)=\dfrac{2x+1}{25},\quad x=0,1,2,3,4$
(a) $P(X=4)$ (b) $P(X\leq 1)$
(c) $P(2\leq X<4)$ (d) $P(X>-10)$

3-20. ⊕ $f(x)=(3/4)(1/4)^x,\quad x=0,1,2,\ldots$
(a) $P(X=2)$ (b) $P(X\leq 2)$
(c) $P(X>2)$ (d) $P(X\geq 1)$

3-21.

x	1.25	1.5	1.75	2	2.25
$f(x)$	0.2	0.4	0.1	0.2	0.1

(a) $P(X\geq 2)$ (b) $P(X<1.65)$
(c) $P(X=1.5)$ (d) $P(X<1.3 \text{ or } X>2.1)$

3-22. Consider the hospital patients in Example 2-8. Two patients are selected randomly, with replacement, from the total patients at Hospital 1. What is the probability mass function of the number of patients in the sample who are admitted?

3-23. ⊕ An article in *Knee Surgery, Sports Traumatology, Arthroscopy* ["Arthroscopic Meniscal Repair with an Absorbable Screw: Results and Surgical Technique" (2005, Vol. 13, pp. 273–279)] cites a success rate of more than 90% for meniscal tears with a rim width under 3 mm, but only a 67% success rate for tears of 3–6 mm. If you are unlucky enough to suffer a meniscal tear of under 3 mm on your left knee and one of width 3–6 mm on your right knee, what is the probability mass function of the number of successful surgeries? Assume that the surgeries are independent.

3-24. ⊕ An optical inspection system is used to distinguish among different part types. The probability of a correct classification of any part is 0.98. Suppose that three parts are inspected and that the classifications are independent. Let the random variable X denote the number of parts that are correctly classified. Determine the probability mass function of X.

3-25. ⊕ In a semiconductor manufacturing process, three wafers from a lot are tested. Each wafer is classified as *pass* or *fail*. Assume that the probability that a wafer passes the test is 0.8 and that wafers are independent. Determine the probability mass function of the number of wafers from a lot that pass the test.

3-26. ⊕ The space shuttle flight control system called Primary Avionics Software Set (PASS) uses four independent computers working in parallel. At each critical step, the computers "vote" to determine the appropriate step. The probability that a computer will ask for a roll to the left when a roll to the right is appropriate is 0.0001. Let X denote the number of computers that vote for a left roll when a right roll is appropriate. What is the probability mass function of X?

3-27. A disk drive manufacturer sells storage devices with capacities of one terabyte, 500 gigabytes, and 100 gigabytes with probabilities 0.5, 0.3, and 0.2, respectively. The revenues associated with the sales in that year are estimated to be \$50 million, \$25 million, and \$10 million, respectively. Let X denote the revenue of storage devices during that year. Determine the probability mass function of X.

3-28. ⊕ Marketing estimates that a new instrument for the analysis of soil samples will be very successful, moderately successful, or unsuccessful with probabilities 0.3, 0.6, and 0.1, respectively. The yearly revenue associated with a very successful, moderately successful, or unsuccessful product is \$10 million, \$5 million, and \$1 million, respectively. Let the random variable X denote the yearly revenue of the product. Determine the probability mass function of X.

3-29. ⊕ The distributor of a machine for cytogenics has developed a new model. The company estimates that when it is introduced into the market, it will be very successful with a probability 0.6, moderately successful with a probability 0.3, and not successful with probability 0.1. The estimated yearly profit associated with the model being very successful is \$15 million and with it being moderately successful is \$5 million; not successful would result in a loss of \$500,000. Let X be the yearly profit of the new model. Determine the probability mass function of X.

3-30. ⊕ An assembly consists of two mechanical components. Suppose that the probabilities that the first and second components meet specifications are 0.95 and 0.98, respectively. Assume that the components are independent. Determine the probability mass function of the number of components in the assembly that meet specifications.

3-31. ⊕ An assembly consists of three mechanical components. Suppose that the probabilities that the first, second, and third components meet specifications are 0.95, 0.98, and 0.99, respectively. Assume that the components are independent. Determine the probability mass function of the number of components in the assembly that meet specifications.

3-32. The data from 200 endothermic reactions involving sodium bicarbonate are summarized as follows:

Final Temperature Conditions	Number of Reactions
266 K	48
271 K	60
274 K	92

Calculate the probability mass function of final temperature.

3-33. ⊕ Actual lengths of stay at a hospital's emergency department in 2009 are shown in the following table (rounded to the nearest hour). Length of stay is the total of wait and service times. Some longer stays are also approximated as 15 hours in this table.

Hours	Count	Percent
1	19	3.80
2	51	10.20
3	86	17.20
4	102	20.40
5	87	17.40
6	62	12.40
7	40	8.00
8	18	3.60
9	14	2.80
10	11	2.20
15	10	2.00

Calculate the probability mass function of the wait time for service.

3-34. The distribution of the time until a Web site changes is important to Web crawlers that search engines use to maintain current information about Web sites. The distribution of the time until change (in days) of a Web site is approximated in the following table.

Days until Changes	Probability
1.5	0.05
3.0	0.25
4.5	0.35
5.0	0.20
7.0	0.15

Calculate the probability mass function of the days until change.

3-35 The following table shows the typical depth (rounded to the nearest foot) for nonfailed wells in geological formations in Baltimore County (*The Journal of Data Science,* 2009, Vol. 7, pp. 111–127).

Geological Formation Group	Number of Nonfailed Wells	Nonfailed Well Depth
Gneiss	1,515	255
Granite	26	218
Loch Raven Schist	3,290	317
Mafic	349	231
Marble	280	267
Prettyboy Schist	1,343	255
Other schists	887	267
Serpentine	36	217
Total	7,726	2,027

Calculate the probability mass function of depth for nonfailed wells from the table.

3-36. Consider the wafers with contamination particles in Example 2-17. Assume that wafers are independent with respect to contamination particles. Wafers are selected until one with five or more contamination particles occurs. What is the probability mass function of the number of wafers selected?

3-37. Consider the circuit in Example 2-32. Assume that devices fail independently. What is the probability mass function of the number of failed devices?

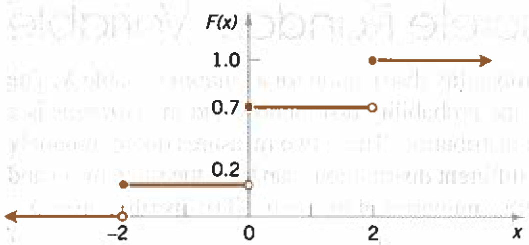


FIGURE 3-3 Cumulative distribution function for Example 3-7.

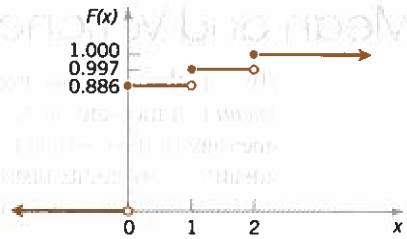


FIGURE 3-4 Cumulative distribution function for Example 3-8.

Exercises FOR SECTION 3-3

⊕ Problem available in *WileyPLUS* at instructor's discretion.
⊕ **Go Tutorial** Tutoring problem available in *WileyPLUS* at instructor's discretion

3-38. ⊕ Determine the cumulative distribution function of the random variable in Exercise 3-16.

3-39. ⊕ Determine the cumulative distribution function for the random variable in Exercise 3-17; also determine the following probabilities:

(a) $P(X \le 1.25)$ (b) $P(X \le 2.2)$
(c) $P(-1.1 < X \le 1)$ (d) $P(X > 0)$

3-40. ⊕ Determine the cumulative distribution function for the random variable in Exercise 3-18; also determine the following probabilities:

(a) $P(X < 1.5)$ (b) $P(X \le 3)$
(c) $P(X > 2)$ (d) $P(1 < X \le 2)$

3-41. Determine the cumulative distribution function for the random variable in Exercise 3-19.

3-42. Determine the cumulative distribution function for the random variable in Exercise 3-20.

3-43. Determine the cumulative distribution function for the random variable in Exercise 3-21.

3-44. Determine the cumulative distribution function for the random variable in Exercise 3-22.

3-45. ⊕ Determine the cumulative distribution function for the random variable in Exercise 3-25.

3-46. ⊕ Determine the cumulative distribution function for the random variable in Exercise 3-26.

3-47. ⊕ Determine the cumulative distribution function for the random variable in Exercise 3-27.

3-48. ⊕ Determine the cumulative distribution function for the variable in Exercise 3-28.

Verify that the following functions are cumulative distribution functions, and determine the probability mass function and the requested probabilities.

3-49. ⊕ $F(x) = \begin{cases} 0 & x < 1 \\ 0.5 & 1 \le x < 3 \\ 1 & 3 \le x \end{cases}$

(a) $P(X \le 3)$ (b) $P(X \le 2)$
(c) $P(1 \le X \le 2)$ (d) $P(X > 2)$

3-50. ⊕ Errors in an experimental transmission channel are found when the transmission is checked by a certifier that detects missing pulses. The number of errors found in an eight-bit byte is a random variable with the following distribution:

$$F(x) = \begin{cases} 0 & x < 1 \\ 0.7 & 1 \le x < 4 \\ 0.9 & 4 \le x < 7 \\ 1 & 7 \le x \end{cases}$$

Determine each of the following probabilities:

(a) $P(X \le 4)$ (b) $P(X > 7)$ (c) $P(X \le 5)$
(d) $P(X > 4)$ (e) $P(X \le 2)$

3-51. ⊕ $F(x) = \begin{cases} 0 & x < -10 \\ 0.25 & -10 \le x < 30 \\ 0.75 & 30 \le x < 50 \\ 1 & 50 \le x \end{cases}$

(a) $P(X \le 50)$ (b) $P(X \le 40)$
(c) $P(40 \le X \le 60)$ (d) $P(X < 0)$
(e) $P(0 \le X < 10)$ (f) $P(-10 < X < 10)$

3-52. ⊕ The thickness of wood paneling (in inches) that a customer orders is a random variable with the following cumulative distribution function:

$$F(x) = \begin{cases} 0 & x < 1/8 \\ 0.2 & 1/8 \le x < 1/4 \\ 0.9 & 1/4 \le x < 3/8 \\ 1 & 3/8 \le x \end{cases}$$

Determine the following probabilities:

(a) $P(X \le 1/18)$ (b) $P(X \le 1/4)$ (c) $P(X \le 5/16)$
(d) $P(X > 1/4)$ (e) $P(X \le 1/2)$

3-53. Determine the cumulative distribution function for the random variable in Exercise 3-32.

3-54. ⊕ Determine the cumulative distribution function for the random variable in Exercise 3-33.

3-55. Determine the cumulative distribution function for the random variable in Exercise 3-34.

3-56. Determine the cumulative distribution function for the random variable in Exercise 3-35.

3-4 Mean and Variance of a Discrete Random Variable

Two numbers are often used to summarize a probability distribution for a random variable X. The *mean* is a measure of the center or middle of the probability distribution, and the *variance* is a measure of the dispersion, or variability in the distribution. These two measures do not uniquely identify a probability distribution. That is, two different distributions can have the same mean and variance. Still, these measures are simple, useful summaries of the probability distribution of X.

Mean, Variance, and Standard Deviation

The **mean** or **expected value** of the discrete random variable X, denoted as μ or $E(X)$, is

$$\mu = E(X) = \sum_x x f(x) \tag{3-3}$$

The **variance** of X, denoted as σ^2 or $V(X)$, is

$$\sigma^2 = V(X) = E(X-\mu)^2 = \sum_x (x-\mu)^2 f(x) = \sum_x x^2 f(x) - \mu^2$$

The **standard deviation** of X is $\sigma = \sqrt{\sigma^2}$.

The mean of a discrete random variable X is a weighted average of the possible values of X with weights equal to the probabilities. If $f(x)$ is the probability mass function of a loading on a long, thin beam, $E(X)$ is the point at which the beam balances. Consequently, $E(X)$ describes the "center" of the distribution of X in a manner similar to the balance point of a loading. See Fig. 3-5.

The variance of a random variable X is a measure of dispersion or scatter in the possible values for X. The variance of X uses weight $f(x)$ as the multiplier of each possible squared deviation $(x-\mu)^2$. Figure 3-5 illustrates probability distributions with equal means but different variances. Properties of summations and the definition of μ can be used to show the equality of the formulas for variance.

$$V(X) = \sum_x (x-\mu)^2 f(x) = \sum_x x^2 f(x) - 2\mu \sum_x x f(x) + \mu^2 \sum_x f(x)$$

$$= \sum_x x^2 f(x) - 2\mu^2 + \mu^2 = \sum_x x^2 f(x) - \mu^2$$

Either formula for $V(x)$ can be used. Figure 3-6 illustrates that two probability distributions can differ even though they have identical means and variances.

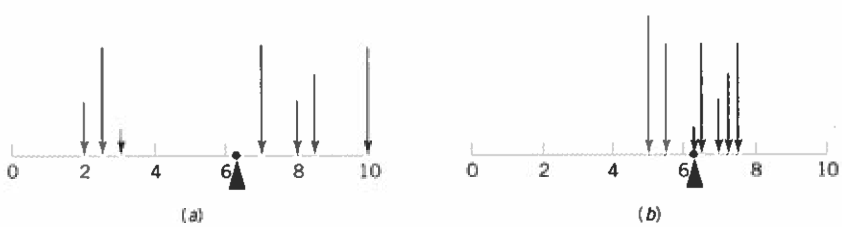


FIGURE 3-5 A probability distribution can be viewed as a loading with the mean equal to the balance point. Parts (a) and (b) illustrate equal means, but part (a) illustrates a larger variance.

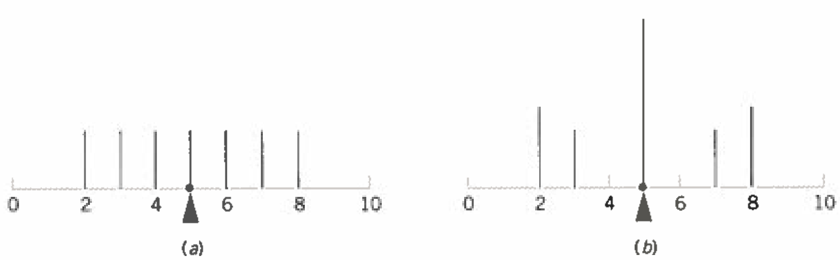


FIGURE 3-6 The probability distributions illustrated in parts (a) and (b) differ even though they have equal means and equal variances.

In Example 3-11, suppose that each e-mail message header reserves 15 kilobytes of memory space for storage. Let the random variable Y denote the memory space reserved for all message headers per hour (in kilobytes). Then $Y = h(X) = 15X$. Also, because $h(X)$ is a linear function,

$$E(Y) = 15E(X) = 15(12.5) = 187.5 \text{ kilobytes}$$

and

$$V(Y) = 15^2 V(X) = 15^2(1.85) = 416.25 \text{ square kilobytes}$$

Exercises FOR SECTION 3-4

⊕ Problem available in *WileyPLUS* at instructor's discretion.

⊕ **Go Tutorial** Tutoring problem available in *WileyPLUS* at instructor's discretion

3-57. ⊕ If the range of X is the set $\{0,1,2,3,4\}$ and $P(X = x) = 0.2$, determine the mean and variance of the random variable.

3-58. ⊕ Determine the mean and variance of the random variable in Exercise 3-16.

3-59. ⊕ Determine the mean and variance of the random variable in Exercise 3-17.

3-60. ⊕ Determine the mean and variance of the random variable in Exercise 3-18.

3-61. ⊕ Determine the mean and variance of the random variable in Exercise 3-19.

3-62. ⊕ Determine the mean and variance of the random variable in Exercise 3-20.

3-63. ⊕ Determine the mean and variance of the random variable in Exercise 3-23.

3-64. ⊕ Determine the mean and variance of the random variable in Exercise 3-24.

3-65. ⊕ The range of the random variable X is $[0, 1, 2, 3, x]$ where x is unknown. If each value is equally likely and the mean of X is 6, determine x.

3-66. In a NiCd battery, a fully charged cell is composed of nickelic hydroxide. Nickel is an element that has multiple oxidation states. Assume the following proportions of the states:

Nickel Charge	Proportions Found
0	0.17
+2	0.35
+3	0.33
+4	0.15

(a) Determine the cumulative distribution function of nickel charge.
(b) Determine the mean and variance of the nickel charge.

3-67. ⊕ The space shuttle flight control system called Primary Avionics Software Set (PASS) uses four independent computers working in parallel. At each critical step, the computers "vote" to determine the appropriate step. The probability that a computer will ask for a roll to the left when a roll to the right is appropriate is 0.0001. Let X denote the number of computers that vote for a left roll when a right roll is appropriate. What are the mean and variance of X?

3-68. ⊕ Trees are subjected to different levels of carbon dioxide atmosphere with 6% of them in a minimal growth condition at 350 parts per million (ppm), 10% at 450 ppm (slow growth), 47% at 550 ppm (moderate growth), and 37% at 650 ppm (rapid growth). What are the mean and standard deviation of the carbon dioxide atmosphere (in ppm) for these trees in ppm?

3-69. ⊕ An article in the *Journal of Database Management* ["Experimental Study of a Self-Tuning Algorithm for DBMS Buffer Pools" (2005, Vol. 16, pp. 1–20)] provided the workload used in the Transaction Processing Performance Council's Version C On-Line Transaction Processing (TPC-C OLTP) benchmark, which simulates a typical order entry application.

The frequency of each type of transaction (in the second column) can be used as the percentage of each type of transaction. The average number of *selects* operations required for each type of transaction is shown.

(a) Determine the mean and standard deviation of the number of *selects* operations for a transaction from the distribution of types shown in the table.
(b) Determine the mean and standard deviation of the total number of operations (*selects, updates*,..., and *joins*) for a transaction from the distribution of types shown in the table.

3-70. Calculate the mean and variance for the random variable in Exercise 3-32.

3-71. ⊕ Calculate the mean and variance for the random variable in Exercise 3-33.

3-72. Calculate the mean and variance for the random variable in Exercise 3-34.

3-73. Calculate the mean and variance for the random variable in Exercise 3-35.

3-74. Calculate the mean and variance for the random variable in Exercise 3-36.

3-75. Calculate the mean for the random variable in Exercise 3-37.

Average Frequencies and Operations in TPC-C

Transaction	Frequency	Selects	Updates	Inserts	Deletes	Nonunique Selects	Joins
New order	43	23	11	12	0	0	0
Payment	44	4.2	3	1	0	0.6	0
Order status	4	11.4	0	0	0	0.6	0
Delivery	5	130	120	0	10	0	0
Stock level	4	0	0	0	0	0	1

3-5 Discrete Uniform Distribution

The simplest discrete random variable is one that assumes only a finite number of possible values, each with equal probability. A random variable X that assumes each of the values $x_1, x_2, \ldots, x_n$ with equal probability $1/n$ is frequently of interest.

Discrete Uniform Distribution

A random variable X has a **discrete uniform distribution** if each of the n values in its range, $x_1, x_2, \ldots, x_n$, has equal probability. Then

$$f(x_i) = 1/n \tag{3-5}$$

Example 3-13 **Serial Number** The first digit of a part's serial number is equally likely to be any one of the digits 0 through 9. If one part is selected from a large batch and X is the first digit of the serial number, X has a discrete uniform distribution with probability 0.1 for each value in $R = \{0, 1, 2, \ldots, 9\}$. That is,

$$f(x) = 0.1$$

for each value in R. The probability mass function of X is shown in Fig. 3-7.

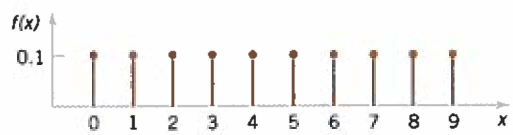


FIGURE 3-7 Probability mass function for a discrete uniform random variable.

Suppose that the range of the discrete random variable X equals the consecutive integers $a, a+1, a+2, \ldots, b$, for $a \le b$. The range of X contains $b-a+1$ values each with probability $1/(b-a+1)$. Now

$$\mu = \sum_{k=a}^{b} k\left(\frac{1}{b-a+1}\right)$$

The algebraic identity $\sum_{k=a}^{b} k = \frac{b(b+1)-(a-1)a}{2}$ can be used to simplify the result to $\mu = (b+a)/2$. The derivation of the variance is left as an exercise.

Mean and Variance

Suppose that X is a discrete uniform random variable on the consecutive integers $a, a+1, a+2, \ldots, b$, for $a \le b$. The mean of X is

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

The variance of X is

$$\sigma^2 = \frac{(b-a+1)^2 - 1}{12} \tag{3-6}$$

Example 3-14 **Number of Voice Lines** As in Example 3-1, let the random variable X denote the number of 48 voice lines that are used at a particular time. Assume that X is a discrete uniform random variable with a range of 0 to 48. Then,

$$E(X) = (48+0)/2 = 24$$

and

$$\sigma = \sqrt{\frac{(48-0+1)^2-1}{12}} = 14.14$$

Practical Interpretation: The average number of lines in use is 24, but the dispersion (as measured by σ) is large. Therefore, at many times far more or fewer than 24 lines are used.

Equation 3-6 is more useful than it might first appear. For example, suppose that the discrete uniform random variable Y has range $5, 10, \ldots, 30$. Then $Y = 5X$ where X has range $1, 2, \ldots, 6$. The mean and variance of Y are obtained from the formulas for a linear function of X in Section 3-4 to be

$$E(Y) = 5E(X) = 5\left(\frac{1+6}{2}\right) = 17.5$$

$$V(Y) = 5^2V(X) = 25\left[\frac{(6-1+1)^2-1}{12}\right] = 72.92$$

Example 3-15 **Proportion of Voice Lines** Let the random variable Y denote the proportion of the 48 voice lines used at a particular time, and X denote the number of lines used at a particular time. Then $Y = X/48$. Therefore,

$$E(Y) = E(X)/48 = 0.5$$

and

$$V(Y) = V(X)/48^2 = 0.087$$

Exercises FOR SECTION 3-5

⊕ Problem available in *WileyPLUS* at instructor's discretion.

⊕ **Go Tutorial** Tutoring problem available in *WileyPLUS* at instructor's discretion

3-76. Let the random variable X have a discrete uniform distribution on the integers $0 \le x \le 99$. Determine the mean and variance of X.

3-77. ⊕ Let the random variable X have a discrete uniform distribution on the integers $1 \le x \le 3$. Determine the mean and variance of X.

3-78. ⊕ Thickness measurements of a coating process are made to the nearest hundredth of a millimeter. The thickness measurements are uniformly distributed with values 0.15, 0.16, 0.17, 0.18, and 0.19. Determine the mean and variance of the coating thickness for this process.

3-79. ⊕ Product codes of two, three, four, or five letters are equally likely. What are the mean and standard deviation of the number of letters in the codes?

3-80. ⊕ The lengths of plate glass parts are measured to the nearest tenth of a millimeter. The lengths are uniformly distributed with values at every tenth of a millimeter starting at 590.0 and continuing through 590.9. Determine the mean and variance of the lengths.

3-81. ⊕ Assume that the wavelengths of photosynthetically active radiations (PAR) are uniformly distributed at integer nanometers in the red spectrum from 675 to 700 nm.

(a) What are the mean and variance of the wavelength distribution for this radiation?

(b) If the wavelengths are uniformly distributed at integer nanometers from 75 to 100 nanometers, how do the mean and variance of the wavelength distribution compare to the previous part? Explain.

3-82. ⊕ The probability of an operator entering alphanumeric data incorrectly into a field in a database is equally likely. The random variable X is the number of fields on a data entry form with 28 fields that have an error. Is X a discrete uniform random variable? Why or why not?

3-83. ⊕ Suppose that X has a discrete uniform distribution on the integers 0 through 9. Determine the mean, variance, and standard deviation of the random variable $Y = 5X$ and compare to the corresponding results for X.

3-84. Show that for a discrete uniform random variable X, if each of the values in the range of X is multiplied by the constant c, the effect is to multiply the mean of X by c and the variance of X by c^2. That is, show that $E(cX) = cE(X)$ and $V(cX) = c^2V(X)$.

3-85. The number of pages in a PDF document you create has a discrete uniform distribution from five to nine pages (including the end points). What are the mean and standard deviation of the number of pages in the document?

3-86. Suppose that nine-digit Social Security numbers are assigned at random. If you randomly select a number, what is the probability that it belongs to one of the 300 million people in the United States?

3-87. Suppose that 1000 seven-digit telephone numbers within your area code are dialed randomly. What is the probability that your number is called?

3-88. The probability that data are entered incorrectly into a field in a database is 0.005. A data entry form has 28 fields, and errors occur independently for each field. The random variable X is the number of fields on the form with an error. Does X have a discrete uniform distribution? Why or why not?

3-89. Each multiple-choice question on an exam has four choices. Suppose that there are 10 questions and the choice is selected randomly and independently for each question. Let X denote the number of questions answered correctly. Does X have a discrete uniform distribution? Why or why not?

3-90. Consider the hospital data in Example 2-8. Suppose a patient is selected randomly from the collection in the table. Let X denote the hospital number of the selected patient (either 1, 2, 3, or 4). Does X have a discrete uniform distribution? Why or why not?

3-6 Binomial Distribution

Consider the following random experiments and random variables:

1. Flip a coin 10 times. Let X = number of heads obtained.
2. A worn machine tool produces 1% defective parts. Let X = number of defective parts in the next 25 parts produced.
3. Each sample of air has a 10% chance of containing a particular rare molecule. Let X = the number of air samples that contain the rare molecule in the next 18 samples analyzed.
4. Of all bits transmitted through a digital transmission channel, 10% are received in error. Let X = the number of bits in error in the next five bits transmitted.
5. A multiple-choice test contains 10 questions, each with four choices, and you guess at each question. Let X = the number of questions answered correctly.
6. In the next 20 births at a hospital, let X = the number of female births.
7. Of all patients suffering a particular illness, 35% experience improvement from a particular medication. In the next 100 patients administered the medication, let X = the number of patients who experience improvement.

These examples illustrate that a general probability model that includes these experiments as particular cases would be very useful.

Each of these random experiments can be thought of as consisting of a series of repeated, random trials: 10 flips of the coin in experiment 1, the production of 25 parts in experiment 2, and so forth. The random variable in each case is a count of the number of trials that meet a specified criterion. The outcome from each trial either meets the criterion that X counts or it does not; consequently, each trial can be summarized as resulting in either a success or a failure. For example, in the multiple-choice experiment, for each question, only the choice that is correct is considered a success. Choosing any one of the three incorrect choices results in the trial being summarized as a failure.

The terms *success* and *failure* are just labels. We can just as well use A and B or 0 or 1. Unfortunately, the usual labels can sometimes be misleading. In experiment 2, because X counts defective parts, the production of a defective part is called a success.

A trial with only two possible outcomes is used so frequently as a building block of a random experiment that it is called a **Bernoulli trial**. It is usually assumed that the trials that constitute the random experiment are **independent**. This implies that the outcome from one trial has no effect on the outcome to be obtained from any other trial. Furthermore, it is often reasonable to assume that the **probability of a success in each trial is constant**. In the multiple-choice experiment, if the test taker has no knowledge of the material and just guesses at each question, we might assume that the probability of a correct answer is 1/4 *for each question.*

EXAMPLE 3-16 **Digital Channel** The chance that a bit transmitted through a digital transmission channel is received in error is 0.1. Also, assume that the transmission trials are independent. Let X = the number of bits in error in the next four bits transmitted. Determine $P(X = 2)$.

Let the letter E denote a bit in error, and let the letter O denote that the bit is okay, that is, received without error. We can represent the outcomes of this experiment as a list of four letters that indicate the bits that are in error and

Example 3-18 **Organic Pollution** Each sample of water has a 10% chance of containing a particular organic pollutant. Assume that the samples are independent with regard to the presence of the pollutant. Find the probability that in the next 18 samples, exactly 2 contain the pollutant.

Let X = the number of samples that contain the pollutant in the next 18 samples analyzed. Then X is a binomial random variable with $p = 0.1$ and $n = 18$. Therefore,

$$P(X = 2) = \binom{18}{2}(0.1)^2(0.9)^{16}$$

Now $\binom{18}{2} = 18!/[2!16!] = 18(17)/2 = 153$. Therefore,

$$P(X = 2) = 153(0.1)^2(0.9)^{16} = 0.284$$

Determine the probability that at least four samples contain the pollutant. The requested probability is

$$P(X \geq 4) = \sum_{x=4}^{18}\binom{18}{x}(0.1)^x(0.9)^{18-x}$$

However, it is easier to use the complementary event,

$$P(X \geq 4) = 1 - P(X < 4) = 1 - \sum_{x=0}^{3}\binom{18}{x}(0.1)^x(0.9)^{18-x}$$

$$= 1 - [0.150 + 0.300 + 0.284 + 0.168] = 0.098$$

Determine the probability that $3 \leq X < 7$. Now

$$P(3 \leq X < 7) = \sum_{x=3}^{6}\binom{18}{x}(0.1)^x(0.9)^{18-x}$$

$$= 0.168 + 0.070 + 0.022 + 0.005$$

$$= 0.265$$

Practical Interpretation: Binomial random variables are used to model many physical systems and probabilities for all such models can be obtained from the binomial probability mass function.

A table of cumulative binomial probabilities is provided in Appendix A, and it can simplify some calculations. For example, the binomial distribution in Example 3-16 has $p = 0.1$ and $n = 4$. A probability such as $P(X = 2)$ can be calculated from the table as

$$P(X = 2) = P(X \leq 2) - P(X \leq 1) = 0.9963 - 0.9477 = 0.0486$$

and this agrees with the result obtained previously.

The mean and variance of a binomial random variable can be obtained from an analysis of the independent trials that comprise the binomial experiment. Define new random variables

$$X_i = \begin{cases} 1 & \text{if } i\text{th trial is a success} \\ 0 & \text{otherwise} \end{cases}$$

for $i = 1, 2, \ldots, n$. Then

$$X = X_1 + X_2 + \cdots + X_n$$

Also, it is easy to derive the mean and variance of each X_i as

$$E(X_i) = 1p + 0(1-p) = p$$

and

$$V(X_i) = (1-p)^2 p + (0-p)^2 (1-p) = p(1-p)$$

Sums of random variables are discussed in Chapter 5, and there the intuitively reasonable result that

$$E(X) = E(X_1) + E(X_2) + \cdots + E(X_n)$$

is derived. Furthermore, for the independent trials of a binomial experiment, Chapter 5 also shows that

$$V(X) = V(X_1) + V(X_2) + \cdots + V(X_n)$$

Because $E(X_i) = p$ and $V(X_i) = p(1-p)$, we obtain the solution $E(X) = np$ and $V(X) = np(1-p)$.

Mean and Variance

> If X is a binomial random variable with parameters p and n,
>
> $$\mu = E(X) = np \quad \text{and} \quad \sigma^2 = V(X) = np(1-p) \qquad (3\text{-}8)$$

Example 3-19 **Mean and Variance** For the number of transmitted bits received in error in Example 3-16, $n = 4$ and $p = 0.1$, so

$$E(X) = 4(0.1) = 0.4 \quad \text{and} \quad V(X) = 4(0.1)(0.9) = 0.36$$

and these results match those obtained from a direct calculation in Example 3-9.

Exercises FOR SECTION 3-6

⊕ Problem available in *WileyPLUS* at instructor's discretion.

⊕ **Go Tutorial** Tutoring problem available in *WileyPLUS* at instructor's discretion

3-91. For each scenario (a)–(j), state whether or not the binomial distribution is a reasonable model for the random variable and why. State any assumptions you make.

(a) A production process produces thousands of temperature transducers. Let X denote the number of nonconforming transducers in a sample of size 30 selected at random from the process.

(b) From a batch of 50 temperature transducers, a sample of size 30 is selected without replacement. Let X denote the number of nonconforming transducers in the sample.

(c) Four identical electronic components are wired to a controller that can switch from a failed component to one of the remaining spares. Let X denote the number of components that have failed after a specified period of operation.

(d) Let X denote the number of accidents that occur along the federal highways in Arizona during a one-month period.

(e) Let X denote the number of correct answers by a student taking a multiple-choice exam in which a student can eliminate some of the choices as being incorrect in some questions and all of the incorrect choices in other questions.

(f) Defects occur randomly over the surface of a semiconductor chip. However, only 80% of defects can be found by testing. A sample of 40 chips with one defect each is tested. Let X denote the number of chips in which the test finds a defect.

(g) Reconsider the situation in part (f). Now suppose that the sample of 40 chips consists of chips with 1 and with 0 defects.

(h) A filling operation attempts to fill detergent packages to the advertised weight. Let X denote the number of detergent packages that are underfilled.

(i) Errors in a digital communication channel occur in bursts that affect several consecutive bits. Let X denote the number of bits in error in a transmission of 100,000 bits.

(j) Let X denote the number of surface flaws in a large coil of galvanized steel.

3-92. ⊕ Let X be a binomial random variable with $p = 0.2$ and $n = 20$. Use the binomial table in Appendix A to determine the following probabilities.

(a) $P(X \le 3)$ (b) $P(X > 10)$
(c) $P(X = 6)$ (d) $P(6 \le X \le 11)$

3-93. ⊕ Let X be a binomial random variable with $p = 0.1$ and $n = 10$. Calculate the following probabilities from the binomial probability mass function and from the binomial table in Appendix A and compare results.

(a) $P(X \le 2)$ (b) $P(X > 8)$
(c) $P(X = 4)$ (d) $P(5 \le X \le 7)$

3-94. ⊕ The random variable X has a binomial distribution with $n = 10$ and $p = 0.5$. Determine the following probabilities:
(a) $P(X = 5)$ (b) $P(X \le 2)$
(c) $P(X \ge 9)$ (d) $P(3 \le X < 5)$

3-95. ⊕ The random variable X has a binomial distribution with $n = 10$ and $p = 0.01$. Determine the following probabilities.
(a) $P(X = 5)$ (b) $P(X \le 2)$
(c) $P(X \ge 9)$ (d) $P(3 \le X < 5)$

3-96. The random variable X has a binomial distribution with $n = 10$ and $p = 0.5$. Sketch the probability mass function of X.
(a) What value of X is most likely?
(b) What value(s) of X is(are) least likely?

3-97. Sketch the probability mass function of a binomial distribution with $n = 10$ and $p = 0.01$ and comment on the shape of the distribution.
(a) What value of X is most likely?
(b) What value of X is least likely?

3-98. ⊕ Determine the cumulative distribution function of a binomial random variable with $n = 3$ and $p = 1/2$.

3-99. Determine the cumulative distribution function of a binomial random variable with $n = 3$ and $p = 1/4$.

3-100. ⊕ An electronic product contains 40 integrated circuits. The probability that any integrated circuit is defective is 0.01, and the integrated circuits are independent. The product operates only if there are no defective integrated circuits. What is the probability that the product operates?

3-101. The phone lines to an airline reservation system are occupied 40% of the time. Assume that the events that the lines are occupied on successive calls are independent. Assume that 10 calls are placed to the airline.
(a) What is the probability that for exactly three calls, the lines are occupied?
(b) What is the probability that for at least one call, the lines are not occupied?
(c) What is the expected number of calls in which the lines are all occupied?

3-102. A multiple-choice test contains 25 questions, each with four answers. Assume that a student just guesses on each question.
(a) What is the probability that the student answers more than 20 questions correctly?
(b) What is the probability that the student answers fewer than 5 questions correctly?

3-103. A particularly long traffic light on your morning commute is green 20% of the time that you approach it. Assume that each morning represents an independent trial.
(a) Over 5 mornings, what is the probability that the light is green on exactly one day?
(b) Over 20 mornings, what is the probability that the light is green on exactly four days?
(c) Over 20 mornings, what is the probability that the light is green on more than four days?

3-104. ⊕ Samples of rejuvenated mitochondria are mutated (defective) in 1% of cases. Suppose that 15 samples are studied and can be considered to be independent for mutation. Determine the following probabilities. The binomial table in Appendix A can help.
(a) No samples are mutated.
(b) At most one sample is mutated.
(c) More than half the samples are mutated.

3-105. ⊕ An article in *Information Security Technical Report* ["Malicious Software—Past, Present and Future" (2004, Vol. 9, pp. 6–18)] provided the following data on the top 10 malicious software instances for 2002. The clear leader in the number of registered incidences for the year 2002 was the Internet worm "Klez," and it is still one of the most widespread threats. This virus was first detected on 26 October 2001, and it has held the top spot among malicious software for the longest period in the history of virology.

The 10 most widespread malicious programs for 2002

Place	Name	% Instances
1	I-Worm.Klez	61.22%
2	I-Worm.Lentin	20.52%
3	I-Worm.Tanatos	2.09%
4	I-Worm.BadtransII	1.31%
5	Macro.Word97.Thus	1.19%
6	I-Worm.Hybris	0.60%
7	I-Worm.Bridex	0.32%
8	I-Worm.Magistr	0.30%
9	Win95.CIH	0.27%
10	I-Worm.Sircam	0.24%

(Source: Kaspersky Labs).

Suppose that 20 malicious software instances are reported. Assume that the malicious sources can be assumed to be independent.
(a) What is the probability that at least one instance is "Klez?"
(b) What is the probability that three or more instances are "Klez?"
(c) What are the mean and standard deviation of the number of "Klez" instances among the 20 reported?

3-106. ⊕ Heart failure is due to either natural occurrences (87%) or outside factors (13%). Outside factors are related to induced substances or foreign objects. Natural occurrences are caused by arterial blockage, disease, and infection. Suppose that 20 patients will visit an emergency room with heart failure. Assume that causes of heart failure for the individuals are independent.
(a) What is the probability that three individuals have conditions caused by outside factors?
(b) What is the probability that three or more individuals have conditions caused by outside factors?
(c) What are the mean and standard deviation of the number of individuals with conditions caused by outside factors?

3-107. A computer system uses passwords that are exactly six characters and each character is one of the 26 letters (a–z) or 10 integers (0–9). Suppose that 10,000 users of the system have unique passwords. A hacker randomly selects (with replacement) one billion passwords from the potential set, and a match to a user's password is called a *hit*.
(a) What is the distribution of the number of hits?
(b) What is the probability of no hits?
(c) What are the mean and variance of the number of hits?

3-108. Samples of 20 parts from a metal punching process are selected every hour. Typically, 1% of the parts require rework. Let X denote the number of parts in the sample of 20 that require rework. A process problem is suspected if X exceeds its mean by more than 3 standard deviations.

(a) If the percentage of parts that require rework remains at 1%, what is the probability that X exceeds its mean by more than 3 standard deviations?

(b) If the rework percentage increases to 4%, what is the probability that X exceeds 1?

(c) If the rework percentage increases to 4%, what is the probability that X exceeds 1 in at least one of the next five hours of samples?

3-109. Because all airline passengers do not show up for their reserved seat, an airline sells 125 tickets for a flight that holds only 120 passengers. The probability that a passenger does not show up is 0.10, and the passengers behave independently.

(a) What is the probability that every passenger who shows up can take the flight?

(b) What is the probability that the flight departs with empty seats?

3-110. This exercise illustrates that poor quality can affect schedules and costs. A manufacturing process has 100 customer orders to fill. Each order requires one component part that is purchased from a supplier. However, typically, 2% of the components are identified as defective, and the components can be assumed to be independent.

(a) If the manufacturer stocks 100 components, what is the probability that the 100 orders can be filled without reordering components?

(b) If the manufacturer stocks 102 components, what is the probability that the 100 orders can be filled without reordering components?

(c) If the manufacturer stocks 105 components, what is the probability that the 100 orders can be filled without reordering components?

3-111. Consider the lengths of stay at a hospital's emergency department in Exercise 3-33. Assume that five persons independently arrive for service.

(a) What is the probability that the length of stay of exactly one person is less than or equal to 4 hours?

(b) What is the probability that exactly two people wait more than 4 hours?

(c) What is the probability that at least one person waits more than 4 hours?

3-112. Consider the patient data in Example 2-8. Suppose that five patients are randomly selected with replacement from the total for hospital 4. Determine the following probabilities:

(a) Exactly one is LWBS. (b) Two or more are LWBS.

(c) At least one is LWBS.

3-113. Assume that a Web site changes its content according to the distribution in Exercise 3-34. Assume that 10 changes are made independently.

(a) What is the probability that the change is made in less than 4 days in 7 of the 10 updates?

(b) What is the probability that the change is made in less than 4 days in 2 or fewer of the 10 updates?

(c) What is the probability that at least one change is made in less than 4 days?

(d) What is the expected number of the 10 updates that occur in less than 4 days?

3-114. Consider the endothermic reactions in Exercise 3-32. A total of 20 independent reactions are to be conducted.

(a) What is the probability that exactly 12 reactions result in a final temperature less than 272 K?

(b) What is the probability that at least 19 reactions result in a final temperature less than 272 K?

(c) What is the probability that at least 18 reactions result in a final temperature less than 272 K?

(d) What is the expected number of reactions that result in a final temperature of less than 272 K?

3-115. The probability that a visitor to a Web site provides contact data for additional information is 0.01. Assume that 1000 visitors to the site behave independently. Determine the following probabilities:

(a) No visitor provides contact data.

(b) Exactly 10 visitors provide contact data.

(c) More than 3 visitors provide contact data.

3-116. Consider the circuit in Example 2-34. Assume that devices fail independently. What is the probability mass function of the number of device failures? Explain why a binomial distribution does not apply to the number of device failures in Example 2-32.

3-117. Consider the time to recharge the flash in cell-phone cameras as in Example 3-2. Assume that the probability that a camera passes the test is 0.8 and the cameras perform independently. What is the smallest sample size needed so that the probability of at least one camera failing is at least 95%?

3-118. Consider the patient data in Example 2-8. Suppose that patients are randomly selected with replacement from the total for hospital 4. What is the smallest sample size needed so that the probability is at least 90% that at least one patient is LWBS?

3-7 Geometric and Negative Binomial Distributions

3-7.1 GEOMETRIC DISTRIBUTION

Consider a random experiment that is closely related to the one used in the definition of a binomial distribution. Again, assume a series of Bernoulli trials (independent trials with constant probability p of a success on each trial). However, instead of a fixed number of trials, trials are conducted until a success is obtained. Let the random variable X denote the

Example 3-25 **Camera Flashes** Consider the time to recharge the flash in Example 3-25. The probability that a camera passes the test is 0.8, and the cameras perform independently. What is the probability that the third failure is obtained in five or fewer tests?

Let X denote the number of cameras tested until three failures have been obtained. The requested probability is $P(X \le 5)$. Here X has a negative binomial distribution with $p = 0.2$ and $r = 3$. Therefore,

$$P(X \le 5) = \sum_{x=3}^{5} \binom{x-1}{2} 0.2^3(0.8)^{x-3} = 0.2^3 + \binom{3}{2} 0.2^3(0.8) + \binom{4}{2} 0.2^3(0.8)^2 = 0.056$$

Exercises FOR SECTION 3-7

⊕ Problem available in *WileyPLUS* at instructor's discretion.
⊕ **Go Tutorial** Tutoring problem available in *WileyPLUS* at instructor's discretion

3-119. ⊕ Suppose that the random variable X has a geometric distribution with $p = 0.5$. Determine the following probabilities:
(a) $P(X = 1)$ (b) $P(X = 4)$ (c) $P(X = 8)$
(d) $P(X \le 2)$ (e) $P(X > 2)$

3-120. ⊕ Suppose that the random variable X has a geometric distribution with a mean of 2.5. Determine the following probabilities:
(a) $P(X = 1)$ (b) $P(X = 4)$ (c) $P(X = 5)$
(d) $P(X \le 3)$ (e) $P(X > 3)$

3-121. ⊕ Consider a sequence of independent Bernoulli trials with $p = 0.2$.
(a) What is the expected number of trials to obtain the first success?
(b) After the eighth success occurs, what is the expected number of trials to obtain the ninth success?

3-122. ⊕ Suppose that X is a negative binomial random variable with $p = 0.2$ and $r = 4$. Determine the following:
(a) $E(X)$ (b) $P(X = 20)$
(c) $P(X = 19)$ (d) $P(X = 21)$
(e) The most likely value for X

3-123. ⊕ The probability of a successful optical alignment in the assembly of an optical data storage product is 0.8. Assume that the trials are independent.
(a) What is the probability that the first successful alignment requires exactly four trials?
(b) What is the probability that the first successful alignment requires at most four trials?
(c) What is the probability that the first successful alignment requires at least four trials?

3-124. ⊕ In a clinical study, volunteers are tested for a gene that has been found to increase the risk for a disease. The probability that a person carries the gene is 0.1.
(a) What is the probability that four or more people need to be tested to detect two with the gene?
(b) What is the expected number of people to test to detect two with the gene?

3-125. ⊕ Assume that each of your calls to a popular radio station has a probability of 0.02 of connecting, that is, of not obtaining a busy signal. Assume that your calls are independent.
(a) What is the probability that your first call that connects is your 10th call?
(b) What is the probability that it requires more than five calls for you to connect?
(c) What is the mean number of calls needed to connect?

3-126. ⊕ A player of a video game is confronted with a series of opponents and has an 80% probability of defeating each one. Success with any opponent is independent of previous encounters. Until defeated, the player continues to contest opponents.
(a) What is the probability mass function of the number of opponents contested in a game?
(b) What is the probability that a player defeats at least two opponents in a game?
(c) What is the expected number of opponents contested in a game?
(d) What is the probability that a player contests four or more opponents in a game?
(e) What is the expected number of game plays until a player contests four or more opponents?

3-127. ⊕ Heart failure is due to either natural occurrences (87%) or outside factors (13%). Outside factors are related to induced substances or foreign objects. Natural occurrences are caused by arterial blockage, disease, and infection. Assume that causes of heart failure for the individuals are independent.
(a) What is the probability that the first patient with heart failure who enters the emergency room has the condition due to outside factors?
(b) What is the probability that the third patient with heart failure who enters the emergency room is the first one due to outside factors?
(c) What is the mean number of heart failure patients with the condition due to natural causes who enter the emergency room before the first patient with heart failure from outside factors?

3-128. ⊕ A computer system uses passwords constructed from the 26 letters (a–z) or 10 integers (0–9). Suppose that 10,000 users of the system have unique passwords. A hacker randomly selects (with replacement) passwords from the potential set.
(a) Suppose that 9900 users have unique six-character passwords and the hacker randomly selects six-character passwords. What are the mean and standard deviation of the number of attempts before the hacker selects a user password?
(b) Suppose that 100 users have unique three-character passwords and the hacker randomly selects three-character passwords.

What are the mean and standard deviation of the number of attempts before the hacker selects a user password?

(c) Comment on the security differences between six- and three-character passwords.

3-129. A trading company uses eight computers to trade on the New York Stock Exchange (NYSE). The probability of a computer failing in a day is 0.005, and the computers fail independently. Computers are repaired in the evening, and each day is an independent trial.

(a) What is the probability that all eight computers fail in a day?

(b) What is the mean number of days until a specific computer fails?

(c) What is the mean number of days until all eight computers fail on the same day?

3-130. Assume that 20 parts are checked each hour and that X denotes the number of parts in the sample of 20 that require rework. Parts are assumed to be independent with respect to rework.

(a) If the percentage of parts that require rework remains at 1%, what is the probability that hour 10 is the first sample at which X exceeds 1?

(b) If the rework percentage increases to 4%, what is the probability that hour 10 is the first sample at which X exceeds 1?

(c) If the rework percentage increases to 4%, what is the expected number of hours until X exceeds 1?

3-131. A fault-tolerant system that processes transactions for a financial services firm uses three separate computers. If the operating computer fails, one of the two spares can be immediately switched online. After the second computer fails, the last computer can be immediately switched online. Assume that the probability of a failure during any transaction is 10^{-8} and that the transactions can be considered to be independent events.

(a) What is the mean number of transactions before all computers have failed?

(b) What is the variance of the number of transactions before all computers have failed?

3-132. ⊕ In the process of meiosis, a single parent diploid cell goes through eight different phases. However, only 60% of the processes pass the first six phases and only 40% pass all eight. Assume that the results from each phase are independent.

(a) If the probability of a successful pass of each one of the first six phases is constant, what is the probability of a successful pass of a single one of these phases?

(b) If the probability of a successful pass of each one of the last two phases is constant, what is the probability of a successful pass of a single one of these phases?

3-133. Show that the probability density function of a negative binomial random variable equals the probability density function of a geometric random variable when $r = 1$. Show that the formulas for the mean and variance of a negative binomial random variable equal the corresponding results for a geometric random variable when $r = 1$.

3-134. Consider the endothermic reactions in Exercise 3-32. Assume that independent reactions are conducted.

(a) What is the probability that the first reaction to result in a final temperature less than 272 K is the tenth reaction?

(b) What is the mean number of reactions until the first final temperature is less than 272 K?

(c) What is the probability that the first reaction to result in a final temperature less than 272 K occurs within three or fewer reactions?

(d) What is the mean number of reactions until two reactions result in final temperatures less than 272 K?

3-135. A Web site randomly selects among 10 products to discount each day. The color printer of interest to you is discounted today.

(a) What is the expected number of days until this product is again discounted?

(b) What is the probability that this product is first discounted again exactly 10 days from now?

(c) If the product is not discounted for the next five days, what is the probability that it is first discounted again 15 days from now?

(d) What is the probability that this product is first discounted again within three or fewer days?

3-136. Consider the visits that result in leave without being seen (LWBS) at an emergency department in Example 2-8. Assume that people independently arrive for service at hospital 1.

(a) What is the probability that the fifth visit is the first one to LWBS?

(b) What is the probability that either the fifth or sixth visit is the first one to LWBS?

(c) What is the probability that the first visit to LWBS is among the first four visits?

(d) What is the expected number of visits until the third LWBS occurs?

3-137. Consider the time to recharge the flash in cell-phone cameras as in Example 3-2. Assume that the probability that a camera passes the test is 0.8 and the cameras perform independently. Determine the following:

(a) Probability that the second failure occurs on the tenth camera tested.

(b) Probability that the second failure occurs in tests of four or fewer cameras.

(b) Expected number of cameras tested to obtain the third failure.

3-138. An array of 30 LED bulbs is used in an automotive light. The probability that a bulb is defective is 0.001 and defective bulbs occur independently. Determine the following:

(a) Probability that an automotive light has two or more defective bulbs.

(b) Expected number of automotive lights to check to obtain one with two or more defective bulbs.

3-139. Consider the patient data in Example 2-8. Suppose that patients are randomly selected with replacement, from the total for hospital 4. Determine the following:

(a) Probability that the first patient admitted is the first one selected.

(b) Probability that four or fewer patients are selected to admit two.

(c) Expected number of patients selected to admit 10.

3-140. Customers visit a Web site, and the probability of an order if a customer views five or fewer pages is 0.01. However, if a customer views more than five pages, the probability of an order is 0.1. The probability a customer views five or more pages is 0.25. The customers behave independently.

(a) Is the number of customers who visit the site until an order is obtained a geometric random variable? Why or why not?

(b) What is the probability that the first order is obtained from the tenth customer to visit the site?

What is the probability that at least one part in the sample is from the local supplier?

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \frac{\binom{100}{0}\binom{200}{4}}{\binom{300}{4}} = 0.804$$

Practical Interpretation: Sampling without replacement is frequently used for inspection and the hypergeometric distribution simplifies the calculations.

The mean and variance of a hypergeometric random variable can be determined from the trials that compose the experiment. However, the trials are not independent, so the calculations are more difficult than for a binomial distribution. The results are stated as follows.

Mean and Variance

If X is a hypergeometric random variable with parameters N, K, and n, then

$$\mu = E(X) = np \quad \text{and} \quad \sigma^2 = V(X) = np(1-p)\left(\frac{N-n}{N-1}\right) \tag{3-14}$$

where $p = K/N$.

Here p is the proportion of successes in the set of N objects.

Example 3-28 **Mean and Variance** In Example 3-27, the sample size is four. The random variable X is the number of parts in the sample from the local supplier. Then, $p = 100/300 = 1/3$. Therefore,

$$E(X) = 4\left(\frac{100}{300}\right) = 1.33$$

and

$$V(X) = 4\left(\frac{1}{3}\right)\left(\frac{2}{3}\right)\left(\frac{300-4}{299}\right) = 0.88$$

For a hypergeometric random variable, $E(X)$ is similar to the mean of a binomial random variable. Also, $V(X)$ differs from the result for a binomial random variable only by the following term.

Finite Population Correction Factor

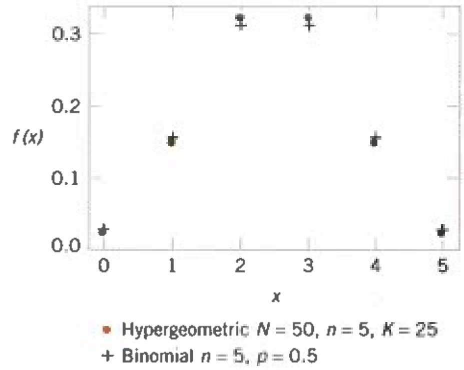
The term in the variance of a hypergeometric random variable

$$\frac{N-n}{N-1} \tag{3-15}$$

is called the **finite population correction factor**.

Sampling with replacement is equivalent to sampling from an infinite set because the proportion of success remains constant for every trial in the experiment. As mentioned previously, if sampling were done with replacement, X would be a binomial random variable and its variance would be $np(1-p)$. Consequently, the finite population correction represents the correction to the binomial variance that results because the sampling is without replacement from the finite set of size N.

If n is small relative to N, the correction is small and the hypergeometric distribution is similar to the binomial distribution. In this case, a binomial distribution can effectively approximate the hypergeometric distribution. A case is illustrated in Fig. 3-13.



	0	1	2	3	4	5
Hypergeometric probability	0.025	0.149	0.326	0.326	0.149	0.025
Binomial probability	0.031	0.156	0.312	0.312	0.156	0.031

FIGURE 3-13 Comparison of hypergeometric and binomial distributions.

Example 3-29 **Customer Sample** A list of customer accounts at a large corporation contains 1000 customers. Of these, 700 have purchased at least one of the corporation's products in the last three months. To evaluate a new product design, 50 customers are sampled at random from the corporate list. What is the probability that more than 45 of the sampled customers have purchased from the corporation in the last three months?

The sampling is without replacement. However, because the sample size of 50 is small relative to the number of customer accounts, 1000, the probability of selecting a customer who has purchased from the corporation in the last three months remains approximately constant as the customers are chosen.

For example, let A denote the event that the first customer selected has purchased from the corporation in the last three months, and let B denote the event that the second customer selected has purchased from the corporation in the last three months. Then, $P(A) = 700/1000 = 0.7$ and $P(B \mid A) = 699/999 = 0.6997$. That is, the trials are approximately independent.

Let X denote the number of customers in the sample who have purchased from the corporation in the last three months. Then, X is a hypergeometric random variable with $N = 1{,}000$, $n = 50$, and $K = 700$. Consequently, $p = K/N = 0.7$. The requested probability is $P(X > 45)$. Because the sample size is small relative to the batch size, the distribution of X can be approximated as binomial with $n = 50$ and $p = 0.7$. Using the binomial approximation to the distribution of X results in

$$P(X > 45) = \sum_{x=46}^{50} \binom{50}{x} 0.7^x (1-0.7)^{50-x} = 0.00017$$

The probability from the hypergeometric distribution is 0.00013, but this requires computer software to compute. The result agrees well with the binomial approximation.

Exercises FOR SECTION 3-8

⊕ Problem available in *WileyPLUS* at instructor's discretion.

⊕ **Go Tutorial** Tutoring problem available in *WileyPLUS* at instructor's discretion

3-141. ⊕ Suppose that X has a hypergeometric distribution with $N = 100$, $n = 4$, and $K = 20$. Determine the following:
(a) $P(X = 1)$ (b) $P(X = 6)$
(c) $P(X = 4)$ (d) Mean and variance of X

3-142. ⊕ Suppose that X has a hypergeometric distribution with $N = 20$, $n = 4$, and $K = 4$. Determine the following:
(a) $P(X = 1)$ (b) $P(X = 4)$
(c) $P(X \le 2)$ (d) Mean and variance of X.

3-143. ⊕ Suppose that X has a hypergeometric distribution with $N = 10$, $n = 3$, and $K = 4$. Sketch the probability mass function of X. Determine the cumulative distribution function for X.

3-144. ⊕ A batch contains 36 bacteria cells and 12 of the cells are not capable of cellular replication. Suppose that you examine three bacteria cells selected at random without replacement.

(a) What is the probability mass function of the number of cells in the sample that can replicate?
(b) What are the mean and variance of the number of cells in the sample that can replicate?
(c) What is the probability that at least one of the selected cells cannot replicate?

3-145. ⊕ A research study uses 800 men under the age of 55. Suppose that 30% carry a marker on the male chromosome that indicates an increased risk for high blood pressure.
(a) If 10 men are selected randomly and tested for the marker, what is the probability that exactly 1 man has the marker?
(b) If 10 men are selected randomly and tested for the marker, what is the probability that more than 1 has the marker?

3-146. ⊕ Printed circuit cards are placed in a functional test after being populated with semiconductor chips. A lot contains 140 cards, and 20 are selected without replacement for functional testing.
(a) If 20 cards are defective, what is the probability that at least 1 defective card is in the sample?
(b) If 5 cards are defective, what is the probability that at least 1 defective card appears in the sample?

3-147. ⊕ The analysis of results from a leaf transmutation experiment (turning a leaf into a petal) is summarized by the type of transformation completed:

		Total Textural Transformation	
		Yes	No
Total Color	Yes	243	26
Transformation	No	13	18

A naturalist randomly selects three leaves from this set without replacement. Determine the following probabilities.
(a) Exactly one has undergone both types of transformations.
(b) At least one has undergone both transformations.
(c) Exactly one has undergone one but not both transformations.
(d) At least one has undergone at least one transformation.

3-148. ⊕ A state runs a lottery in which six numbers are randomly selected from 40 without replacement. A player chooses six numbers before the state's sample is selected.
(a) What is the probability that the six numbers chosen by a player match all six numbers in the state's sample?
(b) What is the probability that five of the six numbers chosen by a player appear in the state's sample?
(c) What is the probability that four of the six numbers chosen by a player appear in the state's sample?
(d) If a player enters one lottery each week, what is the expected number of weeks until a player matches all six numbers in the state's sample?

3-149. ⊕ A slitter assembly contains 48 blades. Five blades are selected at random and evaluated each day for sharpness. If any dull blade is found, the assembly is replaced with a newly sharpened set of blades.
(a) If 10 of the blades in an assembly are dull, what is the probability that the assembly is replaced the first day it is evaluated?
(b) If 10 of the blades in an assembly are dull, what is the probability that the assembly is not replaced until the third day of evaluation? [*Hint:* Assume that the daily decisions are independent, and use the geometric distribution.]
(c) Suppose that on the first day of evaluation, 2 of the blades are dull; on the second day of evaluation, 6 are dull; and on the third day of evaluation, 10 are dull. What is the probability that the assembly is not replaced until the third day of evaluation? [*Hint:* Assume that the daily decisions are independent. However, the probability of replacement changes every day.]

3-150. Calculate the finite population corrections
(a) For Exercises 3-141 and 3-142, for which exercise should the binomial approximation to the distribution of X be better?
(b) For Exercise 3-141, calculate $P(X=1)$ and $P(X=4)$, assuming that X has a binomial distribution, and compare these results to results derived from the hypergeometric distribution.
(c) For Exercise 3-142, calculate $P(X=1)$ and $P(X=4)$, assuming that X has a binomial distribution, and compare these results to the results derived from the hypergeometric distribution.
(d) Use the binomial approximation to the hypergeometric distribution to approximate the probabilities in Exercise 3-146. What is the finite population correction in this exercise?

3-151. Consider the visits that result in leave without being seen (LWBS) at an emergency department in Example 2-8. Assume that four visits that result in LWBS are to be randomly selected (without replacement) for a follow-up interview.
(a) What is the probability that all selected visits are from hospital 4?
(b) What is the probability that no selected visits are from hospital 4?
(c) What is the probability that all selected visits are from the same hospital?

3-152. Consider the nonfailed wells in Exercises 3-35. Assume that four wells are selected randomly (without replacement) for inspection.
(a) What is the probability that exactly two are selected from the Loch Raven Schist?
(b) What is the probability that one or more is selected from the Loch Raven Schist?
(c) What is the expected number selected from the Loch Raven Schist?

3-153. Consider the semiconductor wafer data in Table 2-1. Suppose that 10 wafers are selected randomly (without replacement) for an electrical test. Determine the following:
(a) Probability that exactly 4 wafers have high contamination.
(b) Probability that at least 1 is from the center of the sputtering tool and has high contamination.
(c) Probability that exactly 3 have high contamination or are from the edge of the sputtering tool.
(d) Instead of 10 wafers, what is the minimum number of wafers that need to be selected so that the probability that at least 1 wafer has high contamination is greater than or equal to 0.9?

3-154. Suppose that a healthcare provider selects 20 patients randomly (without replacement) from among 500 to evaluate adherence to a medication schedule. Suppose that 10% of the 500 patients fail to adhere with the schedule. Determine the following:

(a) Probability that exactly 10% of the patients in the sample fail to adhere.
(b) Probability that fewer than 10% of the patients in the sample fail to adhere.
(c) Probability that more than 10% of the patients in the sample fail to adhere.
(d) Mean and variance of the number of patients in the sample who fail to adhere.

3-155. Suppose that lesions are present at 5 sites among 50 in a patient. A biopsy selects 8 sites randomly (without replacement).
(a) What is the probability that lesions are present in at least one selected site?
(b) What is the probability that lesions are present in two or more selected sites?
(c) Instead of eight sites, what is the minimum number of sites that need to be selected to meet the following objective? The probability that at least one site has lesions present is greater than or equal to 0.9.

3-156. A utility company might offer electrical rates based on time-of-day consumption to decrease the peak demand in a day. Enough customers need to accept the plan for it to be successful. Suppose that among 50 major customers, 15 would accept the plan. The utility selects 10 major customers randomly (without replacement) to contact and promote the plan.
(a) What is the probability that exactly two of the selected major customers accept the plan?
(b) What is the probability that at least one of the selected major customers accepts the plan?
(c) Instead of 15 customers, what is the minimum number of major customers that would need to accept the plan to meet the following objective? The probability that at least 1 selected major customer accepts the plan is greater than or equal to 0.95.

3-9 Poisson Distribution

A widely-used distribution emerges from the concept that events occur randomly in an interval (or, more generally, in a region). The random variable of interest is the count of events that occur within the interval. Consider the following example.

Example 3-30 **Wire Flaws** Flaws occur at random along the length of a thin copper wire. Let X denote the random variable that counts the number of flaws in a length of T millimeters of wire and suppose that the average number of flaws per millimeter is λ.

We expect $E(X) = \lambda T$ from the definition of λ. The probability distribution of X is determined as follows. Partition the length of wire into n subintervals of small length $\Delta t = T/n$ (say, one micrometer each). If the subintervals are chosen small enough, the probability that more than one flaw occurs in a subinterval is negligible. Furthermore, we can interpret the assumption that flaws occur at random to imply that every subinterval has the same probability of containing a flaw, say p. Also, the occurrence of a flaw in a subinterval is assumed to be independent of flaws in other subintervals.

Then we can model the distribution of X as approximately a binomial random variable. Each subinterval generates an event (flaw) or not. Therefore,

$$E(X) = \lambda T = np$$

and one can solve for p to obtain

$$p = \frac{\lambda T}{n}$$

From the approximate binomial distribution

$$P(X = x) \approx \binom{n}{x} p^x (1-p)^{n-x}$$

With small enough subintervals, n is large and p is small. Basic properties of limits can be used to show that as n increases

$$\binom{n}{x}\left(\frac{\lambda T}{n}\right)^x \to \frac{(\lambda T)^x}{x!} \quad \left(1-\frac{\lambda T}{n}\right)^{-x} \to 1 \quad \left(1-\frac{\lambda T}{n}\right)^n \to e^{-\lambda T}$$

Therefore,

$$\lim_{n\to\infty} P(X = x) = \frac{e^{-\lambda T}(\lambda T)^x}{x!}, \quad x = 0, 1, 2, \ldots$$

Because the number of subintervals tends to infinity, the range of X (the number of flaws) can equal any nonnegative integer.

Example 3-32 **Magnetic Storage and Contamination** Contamination is a problem in the manufacture of magnetic storage disks. Assume that the number of particles of contamination that occur on a disk surface has a Poisson distribution, and the average number of particles per square centimeter of media surface is 0.1. The area of a disk under study is 100 square centimeters. Determine the probability that 12 particles occur in the area of a disk under study.

Let X denote the number of particles in the area of a disk under study. Here the mean number of particles per cm^2 is $\lambda = 0.1$ and $T = 100\,\text{cm}^2$ so that $\lambda T = 0.1(100) = 10$ particles. Therefore,

$$P(X = 12) = \frac{e^{-10}10^{12}}{12!} = 0.095$$

The probability that zero particles occur in the area of the disk under study is

$$P(X = 0) = e^{-10} = 4.54 \times 10^{-5}$$

Determine the probability that 12 or fewer particles occur in the area of the disk under study. The probability is

$$P(X \le 12) = P(X = 0) + P(X = 1) + \cdots + P(X = 12)$$

$$= \sum_{x=0}^{12} \frac{e^{-10}(10)^x}{x!}$$

Because this sum is tedious to compute, many computer programs calculate cumulative Poisson probabilities. From one such program, $P(X \le 12) = 0.792$.

The mean of a Poisson random variable is

$$E(X) = \sum_{x=1}^{\infty} x \frac{e^{-\lambda T}(\lambda T)^x}{x!} = \lambda T \sum_{x=1}^{\infty} \frac{e^{-\lambda T}(\lambda T)^{x-1}}{(x-1)!}$$

where the summation can start at $x = 1$ because the $x = 0$ term is zero. If a change of variable $y = x - 1$ is used, the summation on the right-hand side of the previous equation is recognized to be the sum of the probabilities of a Poisson random variable and this equals 1. Therefore, the previous equation simplifies to

$$E(X) = \lambda T$$

To obtain the variance of a Poisson random variable, we can start with $E(X^2)$ and this equals

$$E(X^2) = \sum_{x=1}^{\infty} x^2 \frac{e^{-\lambda T}(\lambda T)^x}{x!} = \lambda T \sum_{x=1}^{\infty} x \frac{e^{-\lambda T}(\lambda T)^{x-1}}{(x-1)!}$$

Write $x = (x-1) + 1$ to obtain

$$E(X^2) = \lambda T \sum_{x=1}^{\infty} (x-1) \frac{e^{-\lambda T}(\lambda T)^{x-1}}{(x-1)!} + \lambda T \sum_{x=1}^{\infty} \frac{e^{-\lambda T}(\lambda T)^{x-1}}{(x-1)!}$$

The summation in the first term on the right-hand side of the previous equation is recognized to be the mean of X, which equals λT so that the first term is $(\lambda T)^2$. The summation in the second term on the right-hand side is recognized to be the sum of the probabilities, which equals 1. Therefore, the previous equation simplifies to $E(X^2) = (\lambda T)^2 + \lambda T$. Because the $V(X) = E(X^2) - (EX)^2$, we have

$$V(X) = (\lambda T)^2 + \lambda T - (\lambda T)^2 = \lambda T$$

and the variance is derived.

Mean and Variance

If X is a Poisson random variable over an interval of length T with parameter λ, then

$$\mu = E(X) = \lambda T \quad \text{and} \quad \sigma^2 = V(X) = \lambda T \qquad (3\text{-}16)$$

The mean and variance of a Poisson random variable are equal. For example, if particle counts follow a Poisson distribution with a mean of 25 particles per square centimeter, the variance is also 25 and the standard deviation of the counts is 5 per square centimeter. Consequently, information on the variability is very easily obtained. Conversely, if the variance of count data is much greater than the mean of the same data, the Poisson distribution is not a good model for the distribution of the random variable.

Exercises FOR SECTION 3-9

⊕ Problem available in *WileyPLUS* at instructor's discretion.

⊕ Go Tutorial Tutoring problem available in *WileyPLUS* at instructor's discretion

3-157. ⊕ Suppose that X has a Poisson distribution with a mean of 4. Determine the following probabilities:
(a) $P(X=0)$ (b) $P(X \le 2)$
(c) $P(X=4)$ (d) $P(X=8)$

3-158. ⊕ Suppose that X has a Poisson distribution with a mean of 0.4. Determine the following probabilities:
(a) $P(X=0)$ (b) $P(X \le 2)$
(c) $P(X=4)$ (d) $P(X=8)$

3-159. ⊕ Suppose that the number of customers who enter a bank in an hour is a Poisson random variable, and suppose that $P(X=0)=0.05$. Determine the mean and variance of X.

3-160. The number of telephone calls that arrive at a phone exchange is often modeled as a Poisson random variable. Assume that on the average there are 10 calls per hour.
(a) What is the probability that there are exactly 5 calls in one hour?
(b) What is the probability that there are 3 or fewer calls in one hour?
(c) What is the probability that there are exactly 15 calls in two hours?
(d) What is the probability that there are exactly 5 calls in 30 minutes?

3-161. ⊕ Astronomers treat the number of stars in a given volume of space as a Poisson random variable. The density in the Milky Way Galaxy in the vicinity of our solar system is one star per 16 cubic light-years.
(a) What is the probability of two or more stars in 16 cubic light-years?
(b) How many cubic light-years of space must be studied so that the probability of one or more stars exceeds 0.95?

3-162. Data from www.centralhudsonlabs.com determined the mean number of insect fragments in 225-gram chocolate bars was 14.4, but three brands had insect contamination more than twice the average. See the U.S. Food and Drug Administration–Center for Food Safety and Applied Nutrition for Defect Action Levels for food products. Assume that the number of fragments (contaminants) follows a Poisson distribution.
(a) If you consume a 225-gram bar from a brand at the mean contamination level, what is the probability of no insect contaminants?
(b) Suppose that you consume a bar that is one-fifth the size tested (45 grams) from a brand at the mean contamination level. What is the probability of no insect contaminants?
(c) If you consume seven 28.35-gram (one-ounce) bars this week from a brand at the mean contamination level, what is the probability that you consume one or more insect fragments in more than one bar?
(d) Is the probability of contamination more than twice the mean of 14.4 unusual, or can it be considered typical variation? Explain.

3-163. ⊕ In 1898, L. J. Bortkiewicz published a book entitled *The Law of Small Numbers*. He used data collected over 20 years to show that the number of soldiers killed by horse kicks each year in each corps in the Prussian cavalry followed a Poisson distribution with a mean of 0.61.
(a) What is the probability of more than one death in a corps in a year?
(b) What is the probability of no deaths in a corps over five years?

3-164. The number of flaws in bolts of cloth in textile manufacturing is assumed to be Poisson distributed with a mean of 0.1 flaw per square meter.
(a) What is the probability that there are two flaws in one square meter of cloth?
(b) What is the probability that there is one flaw in 10 square meters of cloth?
(c) What is the probability that there are no flaws in 20 square meters of cloth?
(d) What is the probability that there are at least two flaws in 10 square meters of cloth?

3-165. ⊕ When a computer disk manufacturer tests a disk, it writes to the disk and then tests it using a certifier. The certifier counts the number of missing pulses or errors. The number of errors on a test area on a disk has a Poisson distribution with $\lambda = 0.2$.
(a) What is the expected number of errors per test area?
(b) What percentage of test areas have two or fewer errors?

3-166. The number of cracks in a section of interstate highway that are significant enough to require repair is assumed to follow a Poisson distribution with a mean of two cracks per mile.
(a) What is the probability that there are no cracks that require repair in 5 miles of highway?
(b) What is the probability that at least one crack requires repair in 1/2 mile of highway?
(c) If the number of cracks is related to the vehicle load on the highway and some sections of the highway have a heavy load of vehicles whereas other sections carry a light load, what do you think about the assumption of a Poisson distribution for the number of cracks that require repair?

3-167. ⊕ The number of surface flaws in plastic panels used in the interior of automobiles has a Poisson distribution with a mean of 0.05 flaw per square foot of plastic panel. Assume that an automobile interior contains 10 square feet of plastic panel.

(a) What is the probability that there are no surface flaws in an auto's interior?
(b) If 10 cars are sold to a rental company, what is the probability that none of the 10 cars has any surface flaws?
(c) If 10 cars are sold to a rental company, what is the probability that at most 1 car has any surface flaws?

3-168. ⊕ The number of failures of a testing instrument from contamination particles on the product is a Poisson random variable with a mean of 0.02 failures per hour.
(a) What is the probability that the instrument does not fail in an 8-hour shift?
(b) What is the probability of at least one failure in a 24-hour day?

3-169. The number of content changes to a Web site follows a Poisson distribution with a mean of 0.25 per day.
(a) What is the probability of two or more changes in a day?
(b) What is the probability of no content changes in five days?
(c) What is the probability of two or fewer changes in five days?

3-170. The number of views of a page on a Web site follows a Poisson distribution with a mean of 1.5 per minute.
(a) What is the probability of no views in a minute?
(b) What is the probability of two or fewer views in 10 minutes?
(c) Does the answer to the previous part depend on whether the 10-minute period is an uninterrupted interval? Explain.

3-171. Cabs pass your workplace according to a Poisson process with a mean of five cabs per hour. Suppose that you exit the workplace at 6:00 P.M. Determine the following:
(a) Probability that you wait more than 10 minutes for a cab.
(b) Probability that you wait less than 20 minutes for a cab.
(c) Mean number of cabs per hour so that the probability that you wait more than 10 minutes is 0.1.

3-172. Orders arrive at a Web site according to a Poisson process with a mean of 12 per hour. Determine the following:
(a) Probability of no orders in five minutes.
(b) Probability of 3 or more orders in five minutes.
(c) Length of a time interval such that the probability of no orders in an interval of this length is 0.001.

3-173. The article "An Association Between Fine Particles and Asthma Emergency Department Visits for Children in Seattle" [*Environmental Health Perspectives* June, 1999 107(6)] used Poisson models for the number of asthma emergency department (ED) visits per day. For the zip codes studied, the mean ED visits were 1.8 per day. Determine the following:
(a) Probability of more than five visits in a day.
(b) Probability of fewer than five visits in a week.
(c) Number of days such that the probability of at least one visit is 0.99.
(d) Instead of a mean of 1.8 per day, determine the mean visits per day such that the probability of more than five visits in a day is 0.1.

3-174. *Inclusions* are defects in poured metal caused by contaminants. The number of (large) inclusions in cast iron follows a Poisson distribution with a mean of 2.5 per cubic millimeter. Determine the following:
(a) Probability of at least one inclusion in a cubic millimeter.
(b) Probability of at least five inclusions in 5.0 cubic millimeters.
(c) Volume of material to inspect such that the probability of at least one inclusion is 0.99.
(d) Instead of a mean of 2.5 per cubic millimeters, the mean inclusions per cubic millimeter such that the probability of at least one inclusion is 0.95.

Supplemental Exercises

⊕ Problem available in *WileyPLUS* at instructor's discretion.
⊕ **Go Tutorial** Tutoring problem available in *WileyPLUS* at instructor's discretion

3-175. ⊕ Let the random variable X be equally likely to assume any of the values $1/8$, $1/4$, or $3/8$. Determine the mean and variance of X.

3-176. ⊕ Let X denote the number of bits received in error in a digital communication channel, and assume that X is a binomial random variable with $p = 0.001$. If 1000 bits are transmitted, determine the following:
(a) $P(X = 1)$
(b) $P(X \geq 1)$
(c) $P(X \leq 2)$
(d) mean and variance of X

3-177. Batches that consist of 50 coil springs from a production process are checked for conformance to customer requirements. The mean number of nonconforming coil springs in a batch is five. Assume that the number of nonconforming springs in a batch, denoted as X, is a binomial random variable.
(a) What are n and p?
(b) What is $P(X \leq 2)$?
(c) What is $P(X \geq 49)$?

3-178. An automated egg carton loader has a 1% probability of cracking an egg, and a customer will complain if more than one egg per dozen is cracked. Assume that each egg load is an independent event.
(a) What is the distribution of cracked eggs per dozen? Include parameter values.
(b) What is the probability that a carton of a dozen eggs results in a complaint?
(c) What are the mean and standard deviation of the number of cracked eggs in a carton of a dozen eggs?

3-179. ⊕ A total of 12 cells are replicated. Freshly synthesized DNA cannot be replicated again until mitosis is completed. Two control mechanisms have been identified—one positive and one negative—that are used with equal probability. Assume that each cell independently uses a control mechanism. Determine the following probabilities.
(a) All cells use a positive control mechanism.
(b) Exactly half the cells use a positive control mechanism.
(c) More than four but fewer than seven cells use a positive control mechanism.

3-180. A congested computer network has a 1% chance of losing a data packet, and packet losses are independent events. An e-mail message requires 100 packets.
(a) What distribution of data packets must be re-sent? Include the parameter values.

(b) What is the probability that at least one packet must be re-sent?
(c) What is the probability that two or more packets must be re-sent?
(d) What are the mean and standard deviation of the number of packets that must be re-sent?
(e) If there are 10 messages and each contains 100 packets, what is the probability that at least one message requires that two or more packets be re-sent?

3-181. A particularly long traffic light on your morning commute is green on 20% of the mornings. Assume that each morning represents an independent trial.
(a) What is the probability that the first morning that the light is green is the fourth morning?
(b) What is the probability that the light is not green for 10 consecutive mornings?

3-182. The probability is 0.6 that a calibration of a transducer in an electronic instrument conforms to specifications for the measurement system. Assume that the calibration attempts are independent. What is the probability that at most three calibration attempts are required to meet the specifications for the measurement system?

3-183. An electronic scale in an automated filling operation stops the manufacturing line after three underweight packages are detected. Suppose that the probability of an underweight package is 0.001 and each fill is independent.
(a) What is the mean number of fills before the line is stopped?
(b) What is the standard deviation of the number of fills before the line is stopped?

3-184. The probability that an eagle kills a rabbit in a day of hunting is 10%. Assume that results are independent for each day.
(a) What is the distribution of the number of days until a successful hunt?
(b) What is the probability that the first successful hunt occurs on day five?
(c) What is the expected number of days until a successful hunt?
(d) If the eagle can survive up to 10 days without food (it requires a successful hunt on the 10th day), what is the probability that the eagle is still alive 10 days from now?

3-185. Traffic flow is traditionally modeled as a Poisson distribution. A traffic engineer monitors the traffic flowing through an intersection with an average of six cars per minute. To set the timing of a traffic signal, the following probabilities are used.
(a) What is the probability that no cars pass through the intersection within 30 seconds?
(b) What is the probability that three or more cars pass through the intersection within 30 seconds?
(c) Calculate the minimum number of cars through the intersection so that the probability of this number or fewer cars in 30 seconds is at least 90%.
(d) If the variance of the number of cars through the intersection per minute is 20, is the Poisson distribution appropriate? Explain.

3-186. A shipment of chemicals arrives in 15 totes. Three of the totes are selected at random without replacement for an inspection of purity. If two of the totes do not conform to purity requirements, what is the probability that at least one of the nonconforming totes is selected in the sample?

3-187. The probability that your call to a service line is answered in less than 30 seconds is 0.75. Assume that your calls are independent.
(a) If you call 10 times, what is the probability that exactly nine of your calls are answered within 30 seconds?
(b) If you call 20 times, what is the probability that at least 16 calls are answered in less than 30 seconds?
(c) If you call 20 times, what is the mean number of calls that are answered in less than 30 seconds?

3-188. The probability that your call to a service line is answered in less than 30 seconds is 0.75. Assume that your calls are independent.
(a) What is the probability that you must call four times to obtain the first answer in less than 30 seconds?
(b) What is the mean number of calls until you are answered in less than 30 seconds?

3-189. The probability that your call to a service line is answered in less than 30 seconds is 0.75. Assume that your calls are independent.
(a) What is the probability that you must call six times in order for two of your calls to be answered in less than 30 seconds?
(b) What is the mean number of calls to obtain two answers in less than 30 seconds?

3-190. The number of messages that arrive at a Web site is a Poisson random variable with a mean of five messages per hour.
(a) What is the probability that five messages are received in 1.0 hour?
(b) What is the probability that 10 messages are received in 1.5 hours?
(c) What is the probability that fewer than two messages are received in 0.5 hour?

3-191. Four identical computer servers operate a Web site. Only one is used to operate the site; the others are spares that can be activated in case the active server fails. The probability that a request to the Web site generates a failure in the active server is 0.0001. Assume that each request is an independent trial. What is the mean time until all four computers fail?

3-192. The number of errors in a textbook follows a Poisson distribution with a mean of 0.01 error per page. What is the probability that there are three or fewer errors in 100 pages?

3-193. The probability that an individual recovers from an illness in a one-week time period without treatment is 0.1. Suppose that 20 independent individuals suffering from this illness are treated with a drug and 4 recover in a one-week time period. If the drug has no effect, what is the probability that 4 or more people recover in a one-week time period?

3-194. Patient response to a generic drug to control pain is scored on a 5-point scale where a 5 indicates complete relief. Historically, the distribution of scores is

1	2	3	4	5
0.05	0.1	0.2	0.25	0.4

Two patients, assumed to be independent, are each scored.

(a) What is the probability mass function of the total score?

(b) What is the probability mass function of the average score?

3-195. ⊕ In a manufacturing process that laminates several ceramic layers, 1% of the assemblies are defective. Assume that the assemblies is independent.

(a) What is the mean number of assemblies that need to be checked to obtain five defective assemblies?

(b) What is the standard deviation of the number of assemblies that need to be checked to obtain five defective assemblies?

(c) Determine the minimum number of assemblies that need to be checked so that the probability that at least one defective assembly is obtained exceeds 0.95.

3-196. Consider the circuit in Example 2-35. Assume that devices fail independently. What is the probability of two or fewer failed devices?

3-197. Determine the constant c so that the following function is a probability mass function: $f(x) = cx$ for $x = 1, 2, 3, 4$.

3-198. ⊕ A manufacturer of a consumer electronics product expects 2% of units to fail during the warranty period. A sample of 500 independent units is tracked for warranty performance.

(a) What is the probability that none fails during the warranty period?

(b) What is the expected number of failures during the warranty period?

(c) What is the probability that more than two units fail during the warranty period?

3-199. ⊕ Messages that arrive at a service center for an information systems manufacturer have been classified on the basis of the number of keywords (used to help route messages) and the type of message, either e-mail or voice. Also, 70% of the messages arrive via e-mail and the rest are voice.

Number of keywords	0	1	2	3	4
E-mail	0.1	0.1	0.2	0.4	0.2
Voice	0.3	0.4	0.2	0.1	0

Determine the probability mass function of the number of keywords in a message.

3-200. ⊕ The random variable X has the following probability distribution:

x	2	3	5	8
Probability	0.2	0.4	0.3	0.1

Determine the following:

(a) $P(X \le 3)$ (b) $P(X > 2.5)$

(c) $P(2.7 < X < 5.1)$ (d) $E(X)$ (e) $V(X)$

3-201. Determine the probability mass function for the random variable with the following cumulative distribution function:

$$F(x) = \begin{cases} 0 & x < 2 \\ 0.2 & 2 \le x < 5.7 \\ 0.5 & 5.7 \le x < 6.5 \\ 0.8 & 6.5 \le x < 8.5 \\ 1 & 8.5 \le x \end{cases}$$

3-202. ⊕ Each main bearing cap in an engine contains 4 bolts. The bolts are selected at random without replacement from a parts bin that contains 30 bolts from one supplier and 70 bolts from another.

(a) What is the probability that a main bearing cap contains all bolts from the same supplier?

(b) What is the probability that exactly 3 bolts are from the same supplier?

3-203. Assume that the number of errors along a magnetic recording surface is a Poisson random variable with a mean of one error every 10^5 bits. A sector of data consists of 4096 eight-bit bytes.

(a) What is the probability of more than one error in a sector?

(b) What is the mean number of sectors until an error occurs?

3-204. ⊕ An installation technician for a specialized communication system is dispatched to a city only when three or more orders have been placed. Suppose that orders follow a Poisson distribution with a mean of 0.25 per week for a city with a population of 100,000, and suppose that your city contains a population of 800,000.

(a) What is the probability that a technician is required after a one-week period?

(b) If you are the first one in the city to place an order, what is the probability that you have to wait more than two weeks from the time you place your order until a technician is dispatched?

3-205. From 500 customers, a major appliance manufacturer randomly selects a sample without replacement. The company estimates that 25% of the customers will reply to the survey. If this estimate is correct, what is the probability mass function of the number of customers that will reply?

(a) Assume that the company samples 5 customers.

(b) Assume that the company samples 10 customers.

3-206. ⊕ It is suspected that some of the totes containing chemicals purchased from a supplier exceed the moisture content target. Samples from 30 totes are to be tested for moisture content. Assume that the totes are independent. Determine the proportion of totes from the supplier that must exceed the moisture content target so that the probability is 0.90 that at least 1 tote in the sample of 30 fails the test.

3-207. ⊕ Messages arrive to a computer server according to a Poisson distribution with a mean rate of 10 per hour. Determine the length of an interval of time such that the probability that no messages arrive during this interval is 0.90.

3-208. ⊕ Flaws occur in the interior of plastic used for automobiles according to a Poisson distribution with a mean of 0.02 flaw per panel.

(a) If 50 panels are inspected, what is the probability that there are no flaws?

(b) What is the expected number of panels that need to be inspected before a flaw is found?

(c) If 50 panels are inspected, what is the probability that the number of panels that have one or more flaws is fewer than or equal to 2?

3-209. Saguaro cacti are large cacti indigenous to the southwestern United States and Mexico. Assume that the number of saguaro cacti in a region follows a Poisson distribution with a mean of 280 per square kilometer. Determine the following:
(a) Mean number of cacti per 10,000 square meters.
(b) Probability of no cacti in 10,000 square meters.
(c) Area of a region such that the probability of at least two cacti in the region is 0.9.

3-210. Suppose that 50 sites on a patient might contain lesions. A biopsy selects 8 sites randomly (without replacement). What is the minimum number of sites with lesions so that the probability of at least one selected site contains lesions is greater than or equal to 0.95? Rework for greater than or equal to 0.99.

Mind-Expanding Exercises

3-211. Derive the convergence results used to obtain a Poisson distribution as the limit of a binomial distribution.

3-212. Show that the function $f(x)$ in Example 3-5 satisfies the properties of a probability mass function by summing the infinite series.

3-213. Derive the formula for the mean and standard deviation of a discrete uniform random variable over the range of integers $a, a+1, \ldots b$.

3-214. Derive the expression for the variance of a geometric random variable with parameter p.

3-215. An air flight can carry 120 passengers. A passenger with a reserved seat arrives for the flight with probability 0.95. Assume that the passengers behave independently. (Use of computer software is expected.)
(a) What is the minimum number of seats the airline should reserve for the probability of a full flight to be at least 0.90?
(b) What is the maximum number of seats the airline should reserve for the probability that more passengers arrive than the flight can seat to be less than 0.10?
(c) Discuss some reasonable policies the airline could use to reserve seats based on these probabilities.

3-216. A company performs inspection on shipments from suppliers to detect nonconforming products. Assume that a lot contains 1000 items and 1% are nonconforming. What sample size is needed so that the probability of choosing at least one nonconforming item in the sample is at least 0.90? Assume that the binomial approximation to the hypergeometric distribution is adequate.

3-217. A company performs inspection on shipments from suppliers to detect nonconforming products. The company's policy is to use a sample size that is always 10% of the lot size. Comment on the effectiveness of this policy as a general rule for all sizes of lots.

3-218. A manufacturer stocks components obtained from a supplier. Suppose that 2% of the components are defective and that the defective components occur independently. How many components must the manufacturer have in stock so that the probability that 100 orders can be completed without reordering components is at least 0.95?

3-219. A large bakery can produce rolls in lots of either 0, 1000, 2000, or 3000 per day. The production cost per item is \$0.10. The demand varies randomly according to the following distribution:

Demand for rolls	0	1000	2000	3000
Probability of demand	0.3	0.2	0.3	0.2

Every roll for which there is a demand is sold for \$0.30. Every roll for which there is no demand is sold in a secondary market for \$0.05. How many rolls should the bakery produce each day to maximize the mean profit?

Important Terms and Concepts

Bernoulli trial
Binomial distribution
Cumulative distribution function-discrete random variable
Discrete uniform distribution
Expected value of a function of a discrete random variable
Finite population correction factor
Geometric distribution
Hypergeometric distribution
Lack of memory property-discrete random variable
Mean—discrete random variable
Negative binomial distribution
Poisson distribution
Poisson process
Probability distribution-discrete random variable
Probability mass function
Standard deviation-discrete random variable
Variance—discrete random variable