12 Review of Calculus and Probability

We review in this chapter some basic topics in calculus and probability, which will be useful in later chapters.

12.1 Review of Integral Calculus

In our study of random variables, we often require a knowledge of the basics of integral calculus, which will be briefly reviewed in this section.

Consider two functions: $f(x)$ and $F(x)$. If $F'(x) = f(x)$, we say that $F(x)$ is the **indefinite integral** of $f(x)$. The fact that $F(x)$ is the indefinite integral of $f(x)$ is written

$$F(x) = \int f(x)\,dx$$

The following rules may be used to find the indefinite integrals of many functions (C is an arbitrary constant):

$$\int (1)\,dx = x + C$$

$$\int af(x)\,dx = a\int f(x)\,dx \qquad (a \text{ is any constant})$$

$$\int [f(x) + g(x)]\,dx = \int f(x)\,dx + \int g(x)\,dx$$

$$\int x^n\,dx = \frac{x^{n+1}}{n+1} + C \qquad (n \neq -1)$$

$$\int x^{-1}\,dx = \ln x + C$$

$$\int e^x\,dx = e^x + C$$

$$\int a^x\,dx = \frac{a^x}{\ln a} + C \qquad (a > 0,\ a \neq 1)$$

$$\int [f(x)]^n f'(x)\,dx = \frac{[f(x)]^{n+1}}{n+1} + C \qquad (n \neq -1)$$

$$\int f(x)^{-1} f'(x)\,dx = \ln f(x) + C$$

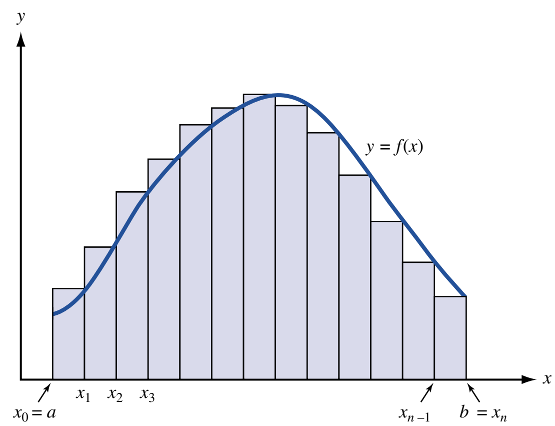


FIGURE 1
Relation of Area and Definite Integral

For two functions $u(x)$ and $v(x)$,

$$\int u(x)v'(x)\,dx = u(x)v(x) - \int v(x)u'(x)\,dx \qquad \text{(Integration by parts)}$$

$$\int e^{f(x)}f'(x)\,dx = e^{f(x)} + C$$

$$\int a^{f(x)}f'(x)\,dx = \frac{a^{f(x)}}{\ln a} + C \qquad (a > 0,\ a \neq 1)$$

The concept of an integral is important for the following reasons. Consider a function $f(x)$ that is continuous for all points satisfying $a \le x \le b$. Let $x_0 = a$, $x_1 = x_0 + \Delta$, $x_2 = x_1 + \Delta, \ldots, x_i = x_{i-1} + \Delta$, $x_n = x_{n-1} + \Delta = b$, where $\Delta = \frac{b-a}{n}$. From Figure 1, we see that as Δ approaches zero (or equivalently, as n grows large),

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

will closely approximate the area under the curve $y = f(x)$ between $x = a$ and $x = b$. If $f(x)$ is continuous for all x satisfying $a \le x \le b$, it can be shown that the area under the curve $y = f(x)$ between $x = a$ and $x = b$ is given by

$$\lim_{\Delta \to 0} \sum_{i=1}^{i=n} f(x_i)\Delta$$

which is written as

$$\int_a^b f(x)\,dx$$

or the **definite integral** of $f(x)$ from $x = a$ to $x = b$. The **Fundamental Theorem of Calculus** states that if $f(x)$ is continuous for all x satisfying $a \le x \le b$, then

$$\int_a^b f(x)\,dx = F(b) - F(a)$$

where $F(x)$ is any indefinite integral of $f(x)$. $F(b) - F(a)$ is often written as $[F(x)]_a^b$. Example 1 illustrates the use of the definite integral.

EXAMPLE 1

Suppose that at time t (measured in hours, and the present $t = 0$), the rate $a(t)$ at which customers enter a bank is $a(t) = 100t$. During the next 2 hours, how many customers will enter the bank?

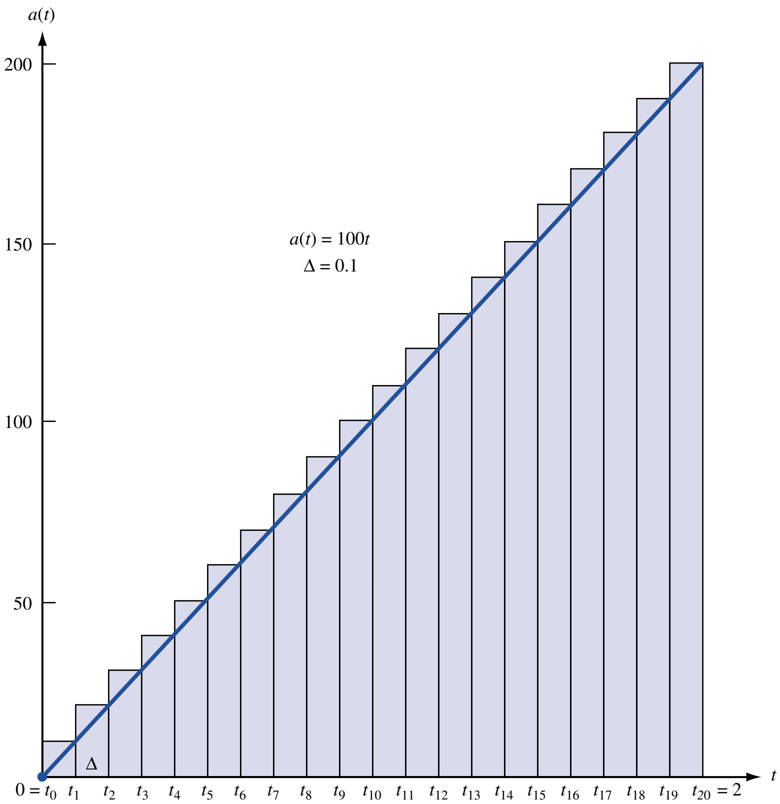


FIGURE 2
Relation of Total Arrivals in Next 2 Hours to Area under $a(t)$ Curve

Solution Let $t_0 = 0$, $t_1 = t_0 + \Delta$, $t_2 = t_1 + \Delta, \ldots, t_n = t_{n-1} + \Delta = 2$ (of course, $\Delta = \frac{2}{n}$). Between time t_{i-1} and time t_i, approximately $100t_i\Delta$ customers will arrive. Therefore, the total number of customers to arrive during the next 2 hours will equal

$$\lim_{\Delta\to 0} \sum_{i=1}^{i=n} 100t_i\Delta$$

(see Figure 2). From the Fundamental Theorem of Calculus,

$$\lim_{\Delta\to 0} \sum_{i=1}^{i=n} 100t_i\Delta = \int_0^2 (100t)\, dt = [50t^2]_0^2 = 200 - 0 = 200$$

Thus, 200 customers will arrive during the next 2 hours.

PROBLEMS

Group A

1 The present is $t = 0$. At a time t years from now, I earn income at a rate e^{2t}. How much money do I earn during the next 5 years?

2 If money is continuously discounted at a rate of r% per year, then \$1 earned t years in the future is equivalent to e^{-rt} dollars earned at the present time. Use this fact to determine the present value of the income earned in Problem 1.

3 At time 0, a company has I units of inventory in stock. Customers demand the product at a constant rate of d units per year (assume that $I \geq d$). The cost of holding 1 unit of stock in inventory for a time Δ is \$$h\Delta$. Determine the total holding cost incurred during the next year.

12.2 Differentiation of Integrals

In our study of inventory theory in Chapter 16, we will have to differentiate a function whose value depends on an integral. Let $f(x, y)$ be a function of variables x and y, and let $g(y)$ and $h(y)$ be functions of y. Then

$$F(y) = \int_{g(y)}^{h(y)} f(x, y)\, dx$$

is a function only of y. **Leibniz's rule for differentiating an integral** states that

$$\text{If} \qquad F(y) = \int_{g(y)}^{h(y)} f(x, y)\, dx, \qquad \text{then}$$

$$F'(y) = h'(y)f(h(y), y) - g'(y)f(g(y), y) + \int_{g(y)}^{h(y)} \frac{\partial f(x, y)}{\partial y}\, dx$$

Example 2 illustrates Leibniz's rule.

EXAMPLE 2

For

$$F(y) = \int_{1}^{y^2} \frac{y\, dx}{x}$$

find $F'(y)$.

Solution We have that $f(x, y) = \frac{y}{x}$, $h(y) = y^2$, $h'(y) = 2y$, $\frac{\partial f}{\partial y} = \frac{1}{x}$, $g(y) = 1$, $g'(y) = 0$. Then

$$F'(y) = 2y\left(\frac{y}{y^2}\right) - 0\left(\frac{y}{1}\right) + \int_{1}^{y^2} \frac{dx}{x}$$

$$= 2 + [\ln x]_1^{y^2} = 2 + \ln y^2 - 0 = 2 + 2 \ln y$$

PROBLEMS

Group A

For each of the following functions, use Leibniz's rule to find $F'(y)$:

1 $F(y) = \int_{y}^{y^2} (2y + x)\, dx$

2 $F(y) = \int_{0}^{y} yx^2\, dx$

3 $F(y) = \int_{0}^{y} 6(5 - x)f(x)\, dx + \int_{y}^{\infty} 4(x - 5)f(x)\, dx$

12.3 Basic Rules of Probability

In this section, we review some basic rules and definitions that you may have encountered during your previous study of probability.

DEFINITION ■ Any situation where the outcome is uncertain is called an **experiment.** ■

For example, drawing a card from a deck of cards would be an experiment.

DEFINITION ■ For any experiment, the **sample space** S of the experiment consists of all possible outcomes for the experiment. ■

For example, if we toss a die and are interested in the number of dots showing, then $S = \{1, 2, 3, 4, 5, 6\}$.

DEFINITION ■ An **event** E consists of any collection of points (set of outcomes) in the sample space. ■

A collection of events $E_1, E_2, \ldots, E_n$ is said to be a **mutually exclusive** collection of events if for $i \neq j$ ($i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, n$), E_i and E_j have no points in common. ■

With each event E, we associate an event $\bar{E}$. $\bar{E}$ consists of the points in the sample space that are not in E. With each event E, we also associate a number $P(E)$, which is the probability that event E will occur when we perform the experiment. The probabilities of events must satisfy the following rules of probability:

Rule 1 For any event E, $P(E) \geq 0$.

Rule 2 If $E = S$ (that is, if E contains all points in the sample space), then $P(E) = 1$.

Rule 3 If $E_1, E_2, \ldots, E_n$ is a mutually exclusive collection of events, then

$$P(E_1 \cup E_2 \cup \cdots \cup E_n) = \sum_{k=1}^{k=n} P(E_k)$$

Rule 4 $P(\bar{E}) = 1 - P(E)$.

DEFINITION ■ For two events E_1 and E_2, $P(E_2|E_1)$ (the **conditional probability** of E_2 given E_1) is the probability that the event E_2 will occur given that event E_1 has occurred. Then

$$P(E_2|E_1) = \frac{P(E_1 \cap E_2)}{P(E_1)} \quad ■ \tag{1}$$

Suppose events E_1 and E_2 both occur with positive probability. Events E_1 and E_2 are **independent** if and only if $P(E_2|E_1) = P(E_2)$ (or equivalently, $P(E_1|E_2) = P(E_1)$). ■

Thus, events E_1 and E_2 are independent if and only if knowledge that E_1 has occurred does not change the probability that E_2 has occurred, and vice versa. From (1), E_1 and E_2 are independent if and only if

$$\frac{P(E_1 \cap E_2)}{P(E_1)} = P(E_2) \qquad \text{or} \qquad P(E_1 \cap E_2) = P(E_1)\,P(E_2) \tag{2}$$

EXAMPLE 3

Suppose we draw a single card from a deck of 52 cards.

1 What is the probability that a heart or spade is drawn?

2 What is the probability that the drawn card is not a 2?

3 Given that a red card has been drawn, what is the probability that it is a diamond? Are the events

$$E_1 = \text{red card is drawn}$$
$$E_2 = \text{diamond is drawn}$$

independent events?

4 Show that the events

$$E_1 = \text{spade is drawn}$$
$$E_2 = 2 \text{ is drawn}$$

are independent events.

Solution **1** Define the events

$$E_1 = \text{heart is drawn}$$
$$E_2 = \text{spade is drawn}$$

E_1 and E_2 are mutually exclusive events with $P(E_1) = P(E_2) = \frac{1}{4}$. We seek $P(E_1 \cup E_2)$. From probability rule 3,

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) = (\tfrac{1}{4}) + (\tfrac{1}{4}) = \tfrac{1}{2}$$

2 Define event E = a 2 is drawn. Then $P(E) = \frac{4}{52} = \frac{1}{13}$. We seek $P(\bar{E})$. From probability rule 4, $P(\bar{E}) = 1 - \frac{1}{13} = \frac{12}{13}$.

3 From (1),

$$P(E_2|E_1) = \frac{P(E_1 \cap E_2)}{P(E_1)}$$
$$P(E_1 \cap E_2) = P(E_2) = \frac{13}{52} = \frac{1}{4}$$
$$P(E_1) = \frac{26}{52} = \frac{1}{2}$$

Thus,

$$P(E_2|E_1) = \frac{\frac{1}{4}}{\frac{1}{2}} = \frac{1}{2}$$

Since $P(E_2) = \frac{1}{4}$, we see that $P(E_2|E_1) \neq P(E_2)$. Thus, E_1 and E_2 are not independent events. (This is because knowing that a red card was drawn increases the probability that a diamond was drawn.)

4 $P(E_1) = \frac{13}{52} = \frac{1}{4}$, $P(E_2) = \frac{4}{52} = \frac{1}{13}$, and $P(E_1 \cap E_2) = \frac{1}{52}$. Since $P(E_1)\,P(E_2) = P(E_1 \cap E_2)$, E_1 and E_2 are independent events. Intuitively, since $\frac{1}{4}$ of all cards in the deck are spades and $\frac{1}{4}$ of all 2's in the deck are spades, knowing that a 2 has been drawn does not change the probability that the card drawn was a spade.

PROBLEMS

Group A

1 Suppose two dice are tossed (for each die, it is equally likely that 1, 2, 3, 4, 5, or 6 dots will show).

a What is the probability that the total of the two dice will add up to 7 or 11?

b What is the probability that the total of the two dice will add up to a number other than 2 or 12?

c Are the events

E_1 = first die shows a 3

E_2 = total of the two dice is 6

independent events?

d Are the events

E_1 = first die shows a 3

E_2 = total of the two dice is 7

independent events?

e Given that the total of the two dice is 5, what is the probability that the first die showed 2 dots?

f Given that the first die shows 5, what is the probability that the total of the two dice is even?

12.4 Bayes' Rule

An important decision often depends on the "state of the world." For example, we may want to know whether a person has tuberculosis. Then we would be concerned with the probability of the following states of the world:

S_1 = person has tuberculosis

S_2 = person does not have tuberculosis

More generally, n mutually exclusive states of the world ($S_1, S_2, \ldots, S_n$) may occur. The states of the world are **collectively exhaustive:** $S_1, S_2, \ldots, S_n$ include all possibilities. Suppose a decision maker assigns a probability $P(S_i)$ to S_i. $P(S_i)$ is the **prior probability** of S_i. To obtain more information about the state of the world, the decision maker may observe the outcome of an experiment. Suppose that for each possible outcome O_j and each possible state of the world S_i, the decision maker knows $P(O_j|S_i)$, the **likelihood** of the outcome O_j given state of the world S_i. Bayes' rule combines prior probabilities and likelihoods with the experimental outcomes to determine a post-experimental probability, or **posterior probability,** for each state of the world. To derive Bayes' rule, observe that (1) implies that

$$P(S_i|O_j) = \frac{P(S_i \cap O_j)}{P(O_j)} \qquad (3)$$

From (1), it also follows that

$$P(S_i \cap O_j) = P(O_j|S_i)P(S_i) \qquad (4)$$

The states of the world $S_1, S_2, \ldots, S_n$ are collectively exhaustive, so the experimental outcome O_j (if it occurs) must occur with one of the S_i (see Figure 3). Since $S_1 \cap O_j$, $S_2 \cap O_j, \ldots, S_n \cap O_j$ are mutually exclusive events, probability rule 3 implies that

$$P(O_j) = P(S_1 \cap O_j) + P(S_2 \cap O_j) + \cdots + P(S_n \cap O_j) \qquad (5)$$

The probabilities of the form $P(S_i \cap O_j)$ are often referred to as **joint probabilities,** and the probabilities $P(O_j)$ are called **marginal probabilities.** Substituting (4) into (5), we obtain

$$P(O_j) = \sum_{k=1}^{k=n} P(O_j|S_k)P(S_k) \qquad (6)$$

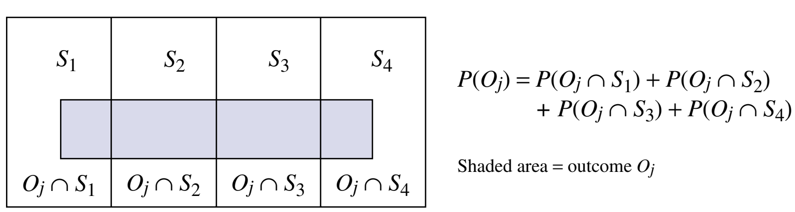


FIGURE 3 Illustration of Equation (5)

Substituting (4) and (6) into (3) yields **Bayes' rule:**

$$P(S_i|O_j) = \frac{P(O_j|S_i)P(S_i)}{\sum_{k=1}^{k=n} P(O_j|S_k)P(S_k)} \tag{7}$$

The following example illustrates the use of Bayes' rule.

EXAMPLE 4 Bayes' Rule

Suppose that 1% of all children have tuberculosis (TB). When a child who has TB is given the Mantoux test, a positive test result occurs 95% of the time. When a child who does not have TB is given the Mantoux test, a positive test result occurs 1% of the time. Given that a child is tested and a positive test result occurs, what is the probability that the child has TB?

Solution The states of the world are

$$S_1 = \text{child has TB}$$
$$S_2 = \text{child does not have TB}$$

The possible experimental outcomes are

$$O_1 = \text{positive test result}$$
$$O_2 = \text{nonpositive test result}$$

We are given the prior probabilities $P(S_1) = .01$ and $P(S_2) = .99$ and the likelihoods $P(O_1|S_1) = .95$, $P(O_1|S_2) = .01$, $P(O_2|S_1) = .05$, and $P(O_2|S_2) = .99$. We seek $P(S_1|O_1)$. From (7),

$$P(S_1|O_1) = \frac{P(O_1|S_1)P(S_1)}{P(O_1|S_1)P(S_1) + P(O_1|S_2)P(S_2)}$$

$$= \frac{.95(.01)}{.95(.01) + .01(.99)} = \frac{95}{194} = .49$$

The reason a positive test result implies only a 49% chance that the child has TB is that many of the 99% of all children who do not have TB will test positive. For example, in a typical group of 10,000 children, 9,900 will not have TB and .01(9,900) = 99 children will yield a positive test result. In the same group of 10,000 children, .01(10,000) = 100 children will have TB and .95(100) = 95 children will yield a positive test result. Thus, the probability that a positive test result indicates TB is $\frac{95}{95+99} = \frac{95}{194}$.

PROBLEMS

Group A

1 A desk contains three drawers. Drawer 1 contains two gold coins. Drawer 2 contains one gold coin and one silver coin. Drawer 3 contains two silver coins. I randomly choose a drawer and then randomly choose a coin. If a silver coin is chosen, what is the probability that I chose drawer 3?

2 Cliff Colby wants to determine whether his South Japan oil field will yield oil. He has hired geologist Digger Barnes to run tests on the field. If there is oil in the field, there is a 95% chance that Digger's tests will indicate oil. If the field contains no oil, there is a 5% chance that Digger's tests will indicate oil. If Digger's tests indicate that there is no oil in the field, what is the probability that the field contains oil? Before Digger conducts the test, Cliff believes that there is a 10% chance that the field will yield oil.

3 A customer has approached a bank for a loan. Without further information, the bank believes there is a 4% chance that the customer will default on the loan. The bank can run a credit check on the customer. The check will yield either a favorable or an unfavorable report. From past experience, the bank believes that P(favorable report being received)| customer will default) $= \frac{1}{40}$, and P(favorable report| customer will not default) $= \frac{99}{100}$. If a favorable report is received, what is the probability that the customer will default on the loan?

4 Of all 40-year-old women, 1% have breast cancer. If a woman has breast cancer, a mammogram will give a positive indication for cancer 90% of the time. If a woman does not have breast cancer, a mammogram will give a positive indication for cancer 9% of the time. If a 40-year-old woman's mammogram gives a positive indication for cancer, what is the probability that she has cancer?

5 Three out of every 1,000 low-risk 50-year-old males have colon cancer. If a man has colon cancer, a test for hidden blood in the stool will indicate hidden blood half the time. If he does not have colon cancer, a test for hidden blood in the stool will indicate hidden blood 3% of the time. If the hidden-blood test turns out positive for a low-risk 50-year-old male, what is the chance that he has colon cancer?

Group B

6 You have made it to the final round of "Let's Make a Deal." You know there is \$1 million behind either door 1, door 2, or door 3. It is equally likely that the prize is behind any of the three. The two doors without a prize have nothing behind them. You randomly choose door 2, but before door 2 is opened Monte reveals that there is no prize behind door 3. You now have the opportunity to switch and choose door 1. Should you switch? Assume that Monte plays as follows: Monte knows where the prize is and will open an empty door, but he cannot open door 2. If the prize is really behind door 2, Monte is equally likely to open door 1 or door 3. If the prize is really behind door 1, Monte must open door 3. If the prize is really behind door 3, Monte must open door 1. What is your decision?

12.5 Random Variables, Mean, Variance, and Covariance

The concepts of random variables, mean, variance, and covariance are employed in several later chapters.

DEFINITION ■ A **random variable** is a function that associates a number with each point in an experiment's sample space. We denote random variables by boldface capital letters (usually **X**, **Y**, or **Z**). ■

Discrete Random Variables

DEFINITION ■ A random variable is **discrete** if it can assume only discrete values $x_1, x_2, \ldots$. A discrete random variable **X** is characterized by the fact that we know the probability that $\mathbf{X} = x_i$ (written $P(\mathbf{X} = x_1)$). ■

$P(\mathbf{X} = x_i)$ is the **probability mass function** (pmf) for the random variable **X**.

DEFINITION ■ The **cumulative distribution function** $F(x)$ for any random variable **X** is defined by $F(x) = P(\mathbf{X} \le x)$. For a discrete random variable **X**,

$$F(x) = \sum_{\substack{\text{all } x \\ \text{having } x_k \le x}} P(\mathbf{X} = x_k) \quad ■$$

An example of a discrete random variable follows.

EXAMPLE 5 **Tossing a Die**

Let $\mathbf{X}$ be the number of dots that show when a die is tossed. Then for $i = 1, 2, 3, 4, 5, 6$, $P(\mathbf{X} = i) = \frac{1}{6}$. The cumulative distribution function (cdf) for $\mathbf{X}$ is shown in Figure 4.

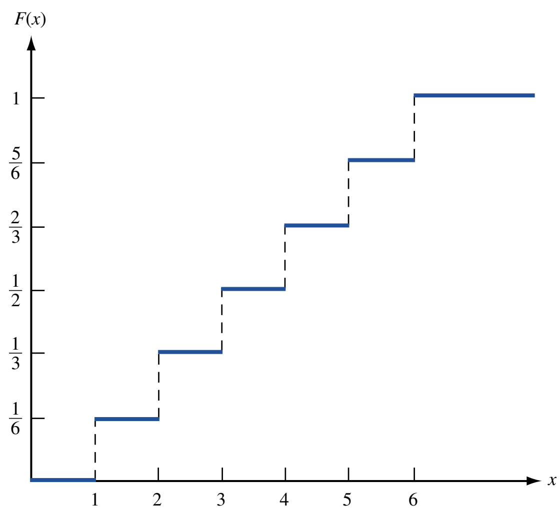


FIGURE 4
Cumulative Distribution Function for Example 5

Continuous Random Variables

If, for some interval, the random variable $\mathbf{X}$ can assume all values on the interval, then $\mathbf{X}$ is a **continuous** random variable. Probability statements about a continuous random variable $\mathbf{X}$ require knowing $\mathbf{X}$'s **probability density function** (pdf). The probability density function $f(x)$ for a random variable $\mathbf{X}$ may be interpreted as follows: For Δ small,

$$P(x \le \mathbf{X} \le x + \Delta) \cong \Delta f(x)$$

From Figure 5, we see that for a random variable $\mathbf{X}$ having density function $f(x)$,

$$\text{Area } 1 = P(a \le \mathbf{X} \le a + \Delta) \cong \Delta f(a)$$

and

$$\text{Area } 2 = P(b \le \mathbf{X} \le b + \Delta) \cong \Delta f(b)$$

Thus, for a random variable $\mathbf{X}$ with density function $f(x)$ as given in Figure 5, values of $\mathbf{X}$ near a are much more likely to occur than values of $\mathbf{X}$ near b.

From our previous discussion of the Fundamental Theorem of Calculus, it follows that

$$P(a \le \mathbf{X} \le b) = \int_a^b f(x)\,dx$$

Thus, for a continuous random variable, any area under the random variable's pdf corresponds to a probability. Using the concept of area as probability, we see that the cdf for a continuous random variable $\mathbf{X}$ with density $f(x)$ is given by

$$F(a) = P(\mathbf{X} \le a) = \int_{-\infty}^{a} f(x)\,dx$$

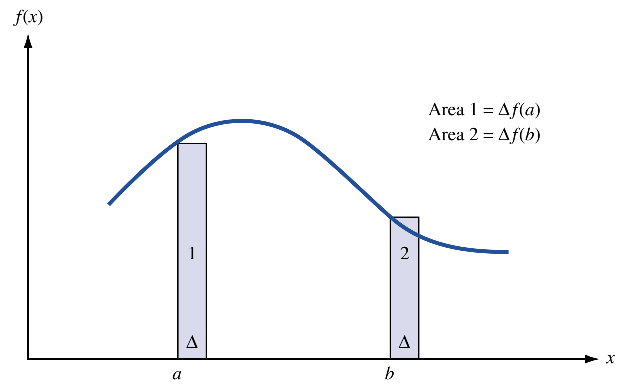


FIGURE 5
Illustration of Probability Density Function

EXAMPLE 6

Consider a continuous random variable **X** having a density function $f(x)$ given by

$$f(x) = \begin{cases} 2x & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Find the cdf for **X**. Also find $P(\frac{1}{4} \le \mathbf{X} \le \frac{3}{4})$.

Solution For $a \le 0$, $F(a) = 0$. For $0 \le a \le 1$,

$$F(a) = \int_0^a 2x\, dx = a^2$$

For $a \ge 1$, $F(a) = 1$. $F(a)$ is graphed in Figure 6.

$$P(\tfrac{1}{4} \le \mathbf{X} \le \tfrac{3}{4}) = \int_{1/4}^{3/4} 2x\, dx = [x^2]_{1/4}^{3/4} = (\tfrac{9}{16}) - (\tfrac{1}{16}) = \tfrac{1}{2}$$

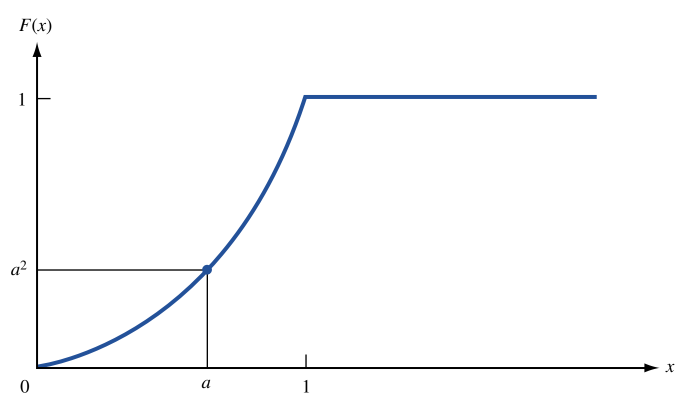


FIGURE 6
Cumulative Distribution Function for Example 6

Mean and Variance of a Random Variable

The **mean** (or expected value) and **variance** are two important measures that are often used to summarize information contained in a random variable's probability distribution. The mean of a random variable **X** (written $E(\mathbf{X})$) is a measure of central location for the random variable.

Mean of a Discrete Random Variable

For a discrete random variable **X**,

$$E(\mathbf{X}) = \sum_{\text{all } k} x_k P(\mathbf{X} = x_k) \tag{8}$$

Mean of a Continuous Random Variable

For a continuous random variable,

$$E(\mathbf{X}) = \int_{-\infty}^{\infty} xf(x)\,dx \tag{9}$$

Observe that in computing $E(\mathbf{X})$, each possible value of a random variable is weighted by its probability of occurring. Thus, the mean of a random variable is essentially the random variable's center of mass.

For a function $h(\mathbf{X})$ of a random variable **X** (such as $\mathbf{X}^2$ and $e^{\mathbf{X}}$), $E[h(\mathbf{X})]$ may be computed as follows. If **X** is a discrete random variable,

$$E[h(\mathbf{X})] = \sum_{\text{all } k} h(x_k)\, P(\mathbf{X} = x_k) \tag{8$'$}$$

If **X** is a continuous random variable,

$$E[h(\mathbf{X})] = \int_{-\infty}^{\infty} h(x)f(x)\,dx \tag{9$'$}$$

The variance of a random variable **X** (written as var **X**) measures the dispersion or spread of **X** about $E(\mathbf{X})$. Then var **X** is defined to be $E[\mathbf{X} - E(\mathbf{X})]^2$.

Variance of a Discrete Random Variable

For a discrete random variable **X**, (8′) yields

$$\text{var } \mathbf{X} = \sum_{\text{all } k} [x_k - E(\mathbf{X})]^2\, P(\mathbf{X} = x_k) \tag{10}$$

Variance of a Continuous Random Variable

For a continuous random variable **X**, (9′) yields

$$\text{var } \mathbf{X} = \int_{-\infty}^{\infty} [x - E(\mathbf{X})]^2 f(x)\,dx \tag{11}$$

Also, var **X** may be found from the relation

$$\text{var } \mathbf{X} = E(\mathbf{X}^2) - E(\mathbf{X})^2 \tag{12}$$

For any random variable **X**, $(\text{var } \mathbf{X})^{1/2}$ is the **standard deviation** of **X** (written σ_x).

Examples 7 and 8 illustrate the computation of mean and variance for a discrete and a continuous random variable.

EXAMPLE 7

Consider the discrete random variable **X** having $P(\mathbf{X} = i) = \frac{1}{6}$ for $i = 1, 2, 3, 4, 5, 6$. Find $E(\mathbf{X})$ and var **X**.

Solution

$$E(\mathbf{X}) = (\tfrac{1}{6})(1 + 2 + 3 + 4 + 5 + 6) = \tfrac{21}{6} = \tfrac{7}{2}$$
$$\text{var } \mathbf{X} = (\tfrac{1}{6})[(1 - 3.5)^2 + (2 - 3.5)^2 + (3 - 3.5)^2 + (4 - 3.5)^2 + (5 - 3.5)^2 + (6 - 3.5)^2] = \tfrac{35}{12}$$

EXAMPLE 8 Continuous Random Variable

Find the mean and variance for the continuous random variable **X** having the following density function:

$$f(x) = \begin{cases} 2x & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Solution

$$E(\mathbf{X}) = \int_0^1 x(2x)\,dx = \left[\frac{2x^3}{3}\right]_0^1 = \frac{2}{3}$$
$$\text{var } \mathbf{X} = \int_0^1 \left(x - \frac{2}{3}\right)^2 2x\,dx = \int_0^1 \left(x^2 - \frac{4x}{3} + \frac{4}{9}\right) 2x\,dx$$
$$= \left[\frac{2x^4}{4} - \frac{8x^3}{9} + \frac{8x^2}{18}\right]_0^1 = \frac{1}{18}$$

Independent Random Variables

DEFINITION ■ Two random variables **X** and **Y** are **independent** if and only if for any two sets A and B,

$$P(\mathbf{X} \in A \text{ and } \mathbf{Y} \in B) = P(\mathbf{X} \in A)P(\mathbf{Y} \in B) \quad ■$$

From this definition, it can be shown that **X** and **Y** are independent random variables if and only if knowledge about the value of **Y** does not change the probability of any event involving **X**. For example, suppose **X** and **Y** are independent random variables. This implies that where $\mathbf{Y} = 8$, $\mathbf{Y} = 10$, $\mathbf{Y} = 0$, or $\mathbf{Y} =$ anything else, $P(\mathbf{X} \ge 10)$ will be the same. If **X** and **Y** are independent, then $E(\mathbf{XY}) = E(\mathbf{X})E(\mathbf{Y})$. (The random variable **XY** has an expected value equal to the product of the expected value of **X** and the expected value of **Y**.)

The definition of independence generalizes to situations where more than two random variables are of interest. Loosely speaking, a group of n random variables is independent if knowledge of the values of any subset of the random variables does not change our view of the distribution of any of the other random variables. (See Problem 5 at the end of this section.)

Covariance of Two Random Variables

An important concept in the study of financial models is covariance. For two random variables **X** and **Y**, the **covariance** of **X** and **Y** (written cov(**X**, **Y**)) is defined by

$$\text{cov}(\mathbf{X}, \mathbf{Y}) = E\{[\mathbf{X} - E(\mathbf{X})][\mathbf{Y} - E(\mathbf{Y})]\} \quad (13)$$

If $\mathbf{X} > E(\mathbf{X})$ tends to occur when $\mathbf{Y} > E(\mathbf{Y})$, and $\mathbf{X} < E(\mathbf{X})$ tends to occur when $\mathbf{Y} < E(\mathbf{Y})$, then cov(**X**, **Y**) will be positive. On the other hand, if $\mathbf{X} > E(\mathbf{X})$ tends to occur when $\mathbf{Y} < E(\mathbf{Y})$, and $\mathbf{X} < E(\mathbf{X})$ tends to occur when $\mathbf{Y} > E(\mathbf{Y})$, then cov(**X**, **Y**) will be negative. The value of cov(**X**, **Y**) measures the association (actually, linear association) between random variables **X** and **Y**. It can be shown that if **X** and **Y** are independent random variables, then cov(**X**, **Y**) = 0. (However, cov(**X**, **Y**) = 0 can hold even if **X** and **Y** are not independent random variables. See Problem 6 at the end of this section for an example.)

EXAMPLE 9 Gotham City Summers

Each summer in Gotham City is classified as being either a rainy summer or a sunny summer. The profits earned by Gotham City's two leading industries (the Gotham City Hotel and the Gotham City Umbrella Store) depend on the summer's weather, as shown in Table 1. Of all summers, 20% are rainy, and 80% are sunny. Let **H** and **U** be the following random variables:

$$\mathbf{H} = \text{profit earned by Gotham City Hotel during a summer}$$
$$\mathbf{U} = \text{profit earned by Gotham City Umbrella Store during a summer}$$

Find cov(**H**,**U**).

Solution We find that

$$E(\mathbf{H}) = .2(-1{,}000) + .8(2{,}000) = \$1{,}400$$
$$E(\mathbf{U}) = .2(4{,}500) + .8(-500) = \$500$$

With probability .20, Gotham City has a rainy summer. Then

$$[\mathbf{H} - E(\mathbf{H})][\mathbf{U} - E(\mathbf{U})] = (-1{,}000 - 1{,}400)(4{,}500 - 500) = -9{,}600{,}000(\text{dollars})^2$$

With probability .80, Gotham City has a sunny summer. Then

$$[\mathbf{H} - E(\mathbf{H})][\mathbf{U} - E(\mathbf{U})] = (2{,}000 - 1{,}400)(-500 - 500) = -600{,}000(\text{dollars})^2$$

Thus,

$$\text{cov}(\mathbf{H},\mathbf{U}) = E\{[\mathbf{H} - E(\mathbf{H})][\mathbf{U} - E(\mathbf{U})]\} = .20(-9{,}600{,}000) + .80(-600{,}000)$$
$$= -2{,}400{,}000(\text{dollars})^2$$

The fact that cov(**H**,**U**) is negative indicates that when one industry does well, the other industry tends to do poorly.

TABLE 1
Profits for Gotham City Covariance

Type of Summer	Hotel Profit	Umbrella Profit
Rainy	−$1,000	$4,500
Sunny	$2,000	−$500

Mean, Variance, and Covariance for Sums of Random Variables

From given random variables $\mathbf{X}_1$ and $\mathbf{X}_2$, we often create new random variables (c is a constant): $c\mathbf{X}_1$, $\mathbf{X}_1 + c$, $\mathbf{X}_1 + \mathbf{X}_2$. The following rules can be used to express the mean, variance, and covariance of these random variables in terms of $E(\mathbf{X}_1)$, $E(\mathbf{X}_2)$, var $\mathbf{X}_1$, var $\mathbf{X}_2$, and cov($\mathbf{X}_1$, $\mathbf{X}_2$). Examples 10 and 11 illustrate the use of these rules.

$$E(c\mathbf{X}_1) = cE(\mathbf{X}_1) \tag{14}$$

$$E(\mathbf{X}_1 + c) = E(\mathbf{X}_1) + c \tag{15}$$

$$E(\mathbf{X}_1 + \mathbf{X}_2) = E(\mathbf{X}_1) + E(\mathbf{X}_2) \tag{16}$$

$$\text{var } c\mathbf{X}_1 = c^2\text{var } \mathbf{X}_1 \tag{17}$$

$$\text{var}(\mathbf{X}_1 + c) = \text{var } \mathbf{X}_1 \tag{18}$$

If $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent random variables,

$$\text{var}(\mathbf{X}_1 + \mathbf{X}_2) = \text{var } \mathbf{X}_1 + \text{var } \mathbf{X}_2 \tag{19}$$

In general,

$$\text{var}(\mathbf{X}_1 + \mathbf{X}_2) = \text{var } \mathbf{X}_1 + \text{var } \mathbf{X}_2 + 2\text{cov}(\mathbf{X}_1, \mathbf{X}_2) \tag{20}$$

For random variables $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$,

$$\text{var}(\mathbf{X}_1 + \mathbf{X}_2 + \cdots + \mathbf{X}_n) = \text{var } \mathbf{X}_1 + \text{var } \mathbf{X}_2 + \cdots + \text{var } \mathbf{X}_n + \sum_{i \neq j} \text{cov}(\mathbf{X}_i, \mathbf{X}_j) \tag{21}$$

Finally, for constants a and b,

$$\text{cov}(a\mathbf{X}_1, b\mathbf{X}_2) = ab\,\text{cov}(\mathbf{X}_1, \mathbf{X}_2) \tag{22}$$

EXAMPLE 10

I pay \$1 to play the following game: I toss a die and receive \$3 for each dot that shows. Determine the mean and variance of my profit.

Solution Let $\mathbf{X}$ be the random variable representing the number of dots that show when the die is tossed. Then my profit is given by the value of the random variable $3\mathbf{X} - 1$. From Example 7, we know that $E(\mathbf{X}) = \frac{7}{2}$ and var $\mathbf{X} = \frac{35}{12}$. In turn, Equations (15) and (14) yield

$$\text{E}(3\mathbf{X} - 1) = E(3\mathbf{X}) - 1 = 3E(\mathbf{X}) - 1 = 3(\tfrac{7}{2}) - 1 = \tfrac{19}{2}$$

From Equations (18) and (17), respectively,

$$\text{var}(3\mathbf{X} - 1) = \text{var}(3\mathbf{X}) = 9(\text{var } \mathbf{X}) = 9(\tfrac{35}{12}) = \tfrac{315}{12}$$

EXAMPLE 11

In Example 9, suppose I owned both the hotel and the umbrella store. Find the mean and the variance of the total profit I would earn during a summer.

Solution My total profits are given by the random variable $\mathbf{H} + \mathbf{U}$. From Equation (16) and Example 9,

$$E(\mathbf{H} + \mathbf{U}) = E(\mathbf{H}) + E(\mathbf{U}) = 1{,}400 + 500 = \$1{,}900$$

Now

$$\text{var } \mathbf{H} = .2(-1{,}000 - 1{,}400)^2 + .8(2{,}000 - 1{,}400)^2 = 1{,}440{,}000(\text{dollars})^2$$
$$\text{var } \mathbf{U} = .2(4{,}500 - 500)^2 + .8(-500 - 500)^2 = 4{,}000{,}000(\text{dollars})^2$$

From Example 9, $\text{cov}(\mathbf{H}, \mathbf{U}) = -2{,}400{,}000$ (dollars)^2. Then Equation (20) yields

$$\begin{aligned}\text{var}(\mathbf{H} + \mathbf{U}) &= \text{var } \mathbf{H} + \text{var } \mathbf{U} + 2\text{cov}(\mathbf{H}, \mathbf{U}) \\ &= 1{,}440{,}000(\text{dollars})^2 + 4{,}000{,}000(\text{dollars})^2 - 2(2{,}400{,}000)(\text{dollars})^2 \\ &= 640{,}000(\text{dollars})^2\end{aligned}$$

Thus, $\mathbf{H} + \mathbf{U}$ has a smaller variance than either **H** or **U**. This is because by owning both the hotel and umbrella store, we will always have, regardless of the weather, one industry that does well and one that does poorly. This reduces the spread, or variability, of our profits.

PROBLEMS

Group A

1 I have 100 items of a product in stock. The probability mass function for the product's demand **D** is $P(\mathbf{D} = 90) = P(\mathbf{D} = 100) = P(\mathbf{D} = 110) = \frac{1}{3}$.

a Find the mass function, mean, and variance of the number of items sold.

b Find the mass function, mean, and variance of the amount of demand that will be unfilled because of lack of stock.

2 I draw 5 cards from a deck (replacing each card immediately after it is drawn). I receive $4 for each heart that is drawn. Find the mean and variance of my total payoff.

3 Consider a continuous random variable **X** with the density function (called the *exponential density*)

$$f(x) = \begin{cases} e^{-x} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

a Find and sketch the cdf for **X**.

b Find the mean and variance of **X**. (*Hint:* Use integration by parts.)

c Find $P(1 \leq \mathbf{X} \leq 2)$.

4 I have 100 units of a product in stock. The demand **D** for the item is a continuous random variable with the following density function:

$$f(d) = \begin{cases} \frac{1}{40} & \text{if } 80 \leq d \leq 120 \\ 0 & \text{otherwise} \end{cases}$$

a Find the probability that supply is insufficient to meet demand.

b What is the expected number of items sold? What is the variance of the number of items sold?

5 An urn contains 10 red balls and 30 blue balls.

a Suppose you draw 4 balls from the urn. Let $\mathbf{X}_i$ be the number of red balls drawn on the ith ball ($\mathbf{X}_i = 0$ or 1). After each ball is drawn, it is put back into the urn. Are the random variables $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{X}_3$, and $\mathbf{X}_4$ independent random variables?

b Repeat part (a) for the case in which the balls are not put back in the urn after being drawn.

Group B

6 Let **X** be the following discrete random variable: $P(\mathbf{X} = -1) = P(\mathbf{X} = 0) = P(\mathbf{X} = 1) = \frac{1}{3}$. Let $\mathbf{Y} = \mathbf{X}^2$. Show that $\text{cov}(\mathbf{X}, \mathbf{Y}) = 0$, but **X** and **Y** are not independent random variables.

12.6 The Normal Distribution

The most commonly used probability distribution in this book is the normal distribution. In this section, we discuss some useful properties of the normal distribution.

DEFINITION ■ A continuous random variable **X** has a normal distribution if for some μ and $\sigma > 0$, the random variable has the following density function:

$$f(x) = \frac{1}{\sigma(2\pi)^{1/2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \quad ■$$

If a random variable **X** is normally distributed with a mean μ and variance σ^2, we write that **X** is $N(\mu, \sigma^2)$. It can be shown that for a normal random variable, $E(\mathbf{X}) = \mu$ and $\text{var}\ \mathbf{X} = \sigma^2$ (the standard deviation of **X** is σ). The normal density functions for several values of σ and a single value of μ are shown in Figure 7.

For any normal distribution, the normal density is symmetric about μ (that is, $f(\mu + a) = f(\mu - a)$). Also, as σ increases, the probability that the random variable assumes a value within c of μ (for any $c > 0$) decreases. Thus, as σ increases, the normal distribution becomes more spread out. The properties are illustrated in Figure 7.

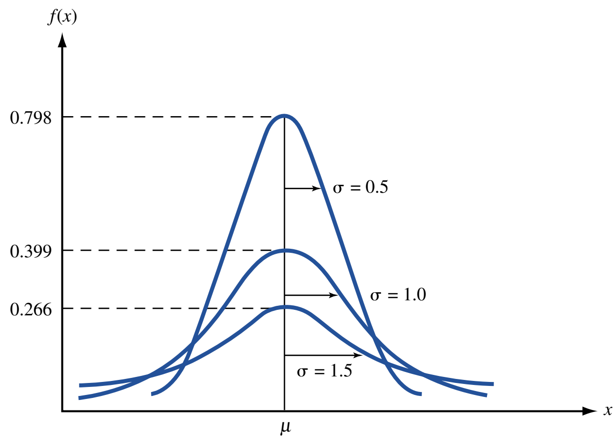


FIGURE 7
Some Examples of Normal Distributions

Useful Properties of Normal Distributions

Property 1 If $\mathbf{X}$ is $N(\mu, \sigma^2)$, then $c\mathbf{X}$ is $N(c\mu, c^2\sigma^2)$.

Property 2 If $\mathbf{X}$ is $N(\mu, \sigma^2)$, then $\mathbf{X} + c$ (for any constant c) is $N(\mu + c, \sigma^2)$.

Property 3 If $\mathbf{X}_1$ is $N(\mu_1, \sigma_1^2)$, $\mathbf{X}_2$ is $N(\mu_2, \sigma_2^2)$, and $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent, then $\mathbf{X}_1 + \mathbf{X}_2$ is $N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

Finding Normal Probabilities via Standardization

If $\mathbf{Z}$ is a random variable that is $N(0, 1)$, then $\mathbf{Z}$ is said to be a standardized normal random variable. In Table 2, $F(z) = P(\mathbf{Z} \le z)$ is tabulated. For example,

$$P(\mathbf{Z} \le -1) = F(-1) = .1587$$

and

$$P(\mathbf{Z} \ge 2) = 1 - P(\mathbf{Z} \le 2) = 1 - F(2) = 1 - .9772 = .0228.$$

If $\mathbf{X}$ is $N(\mu, \sigma^2)$, then $(\mathbf{X} - \mu)/\sigma$ is $N(0, 1)$. This follows, because by property 2 of the normal distribution, $\mathbf{X} - \mu$ is $N(\mu - \mu, \sigma^2) = N(0, \sigma^2)$. Then by property 1, $\frac{x-\mu}{\sigma}$ is $N(\frac{0}{\sigma}, \frac{\sigma^2}{\sigma^2}) = N(0, 1)$. The last equality enables us to use Table 2 to find probabilities for any normal random variable, not just an $N(0, 1)$ random variable. Suppose $\mathbf{X}$ is $N(\mu, \sigma^2)$ and we want to find $P(a \le \mathbf{X} \le b)$. To find this probability from Table 2, we use the following relations (this procedure is called **standardization**):

$$\begin{aligned} P(a \le \mathbf{X} \le b) &= P\left(\frac{a-\mu}{\sigma} \le \frac{\mathbf{X}-\mu}{\sigma} \le \frac{b-\mu}{\sigma}\right) \\ &= P\left(\frac{a-\mu}{\sigma} \le \mathbf{Z} \le \frac{b-\mu}{\sigma}\right) \\ &= F\left(\frac{b-\mu}{\sigma}\right) - F\left(\frac{a-\mu}{\sigma}\right) \end{aligned}$$

The Central Limit Theorem

If $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ are independent random variables, then for n sufficiently large (usually $n \ge 30$ will do, but the actual size of n depends on the distributions of $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$),

TABLE 2

Standard Normal Cumulative Probabilities†

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
−3.8	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
−3.7	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
−3.6	0.0002	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
−3.5	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
−3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
−3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
−3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
−3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
−3.0	0.0014	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
−2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
−2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
−2.4	0.0082	0.0080	0.0078	0.0076	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
−2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
−2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
−2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
−1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
−1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
−1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
−1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
−1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
−1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
−1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1057	0.1038	0.1020	0.1003	0.0985
−1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
−1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
−0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
−0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
−0.7	0.2420	0.2389	0.2358	0.2327	0.2297	0.2266	0.2236	0.2206	0.2177	0.2148
−0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
−0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
−0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
−0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
−0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
−0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
−0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

Source: Reprinted by permission from David E. Kleinbaum, Lawrence L. Kupper, and Keith E. Muller, *Applied Regression Analysis and Other Multivariable Methods,* 2nd edition. Copyright © 1988 PWS-KENT Publishing Company.

†*Note:* Table entry is the area under the standard normal curve to the left of the indicated z-value, thus giving $P(Z \le z)$.

TABLE 2

Standard Normal Cumulative Probabilities (Continued)

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8943	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9673	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9683	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9762	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9924	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9986	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000									

the random variable $\mathbf{X} = \mathbf{X}_1 + \mathbf{X}_2 + \cdots + \mathbf{X}_n$ may be closely approximated by a normal random variable $\mathbf{X}'$ that has $E(\mathbf{X}') = E(\mathbf{X}_1) + E(\mathbf{X}_2) + \cdots + E(\mathbf{X}_n)$ and var $\mathbf{X}' =$ var $\mathbf{X}_1 +$ var $\mathbf{X}_2 + \cdots +$ var $\mathbf{X}_n$. This result is known as the Central Limit Theorem. When we say that $\mathbf{X}'$ closely approximates $\mathbf{X}$, we mean that $P(a \le \mathbf{X} \le b)$ is close to $P(a \le \mathbf{X}' \le b)$.

Finding Normal Probabilities with Excel

Probabilities involving a standard normal variable can be determined with Excel, using the =NORMSDIST function. The S in NORMSDIST stands for *standardized normal.* For example, $P(\mathbf{Z} \le -1)$ can be found by entering the formula

=NORMSDIST(−1)

Normal.xls

Excel returns the value .1587. See Figure 8 and file Normal.xls.

The =NORMDIST function can be used to determine a normal probability for any normal (not just a standard normal) random variable. If $\mathbf{X}$ is N(μ, σ^2), then entering the formula

=NORMSDIST(a,μ,σ,1)

will return $P(\mathbf{X} \le a)$. The "1" ensures that Excel returns the cumulative normal probability. Changing the last argument to "0" causes Excel to return the height of the normal density function for $\mathbf{X} = a$. As an example, we know that IQs follow N(100, 225). The fraction of people with IQs of 90 or less is computed with the formula

=NORMDIST(90,100,15,1)

Excel yields .2525. See Figure 8 and file Normal.xls.

The height of the density for N(100, 225) for $\mathbf{X} = 100$ is computed with the formula

=NORMDIST(100,100,15,0)

Excel yields .026596.

By varying the first argument in the =NORMDIST function, we may graph a normal density. See Figure 9 and sheet density of file Normal.xls.

Consider a given normal random variable $\mathbf{X}$, with mean μ and standard deviation σ. In many situations, we want to answer questions such as the following. (1) Eli Lilly believes that the year's demand for Prozac will be normally distributed, with $\mu = 60$ million d.o.t. (days of therapy) and $\sigma = 5$ million d.o.t. How many units should be produced this year if Lilly wants to have only a 1% chance of running out of Prozac? (2) Family income in Bloomington is normally distributed, with $\mu = \$30{,}000$ and $\sigma = \$8{,}000$. The poorest 10% of all families in Bloomington are eligible for federal aid. What should the aid cutoff be?

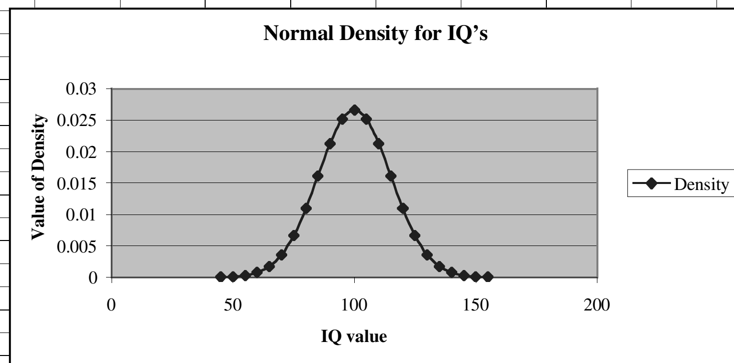
In the first example, we want the 99th percentile of Prozac demand. That is, we seek the number $\mathbf{X}$ such that there is only a 1% chance that demand will exceed $\mathbf{X}$ and a 99% chance

	E	F	G	H
7				
8				
9	P(Z<=-1)	0.158655	normsdist(-1)	
10	P(IQ<90)	0.252492	normdist(90,100,15,1)	
11	density for IQ=100	0.026596	normdist(100,100,15,0)	

FIGURE 8

FIGURE 9

	D	E	F	G	H	I	J	K	L	M	N	O	P
1													
2		IQ	Density										
3		45	3.2018E-05										
4		50	0.00010282										
5		55	0.00029546										
6		60	0.00075973										
7		65	0.00174813										
8		70	0.0035994										
9		75	0.00663181										
10		80	0.010934										
11		85	0.01613138										
12		90	0.02129653										
13		95	0.02515888										
14		100	0.02659615										
15		105	0.02515888										
16		110	0.02129653										
17		115	0.01613138										
18		120	0.010934										
19		125	0.00663181										
20		130	0.0035994										
21		135	0.00174813										
22		140	0.00075973										
23		145	0.00029546										
24		150	0.00010282										
25		155	3.2018E-05										



that it will be less than **X**. In the second example, we want the 10th percentile of family income in Bloomington. That is, we seek the number **X** such that there is only a 10% chance that family income will be less than **X** and a 90% chance that it will exceed **X**.

Suppose we want to find the pth percentile (expressed as a decimal) of a normal random variable **X** with mean μ and standard deviation σ. Simply enter the following formula into Excel:

$$=\text{NORMINV}(p,\mu,\sigma)$$

This will return the number x having the property that $P(\mathbf{X} \leq x) = p$, as desired. We now can solve the two examples described above.

EXAMPLE 12

Eli Lilly believes that the year's demand for Prozac will be normally distributed, with $\mu = 60$ million d.o.t. (days of therapy) and $\sigma = 5$ million d.o.t. How many units should be produced this year if Lilly wants to have only a 1% chance of running out of Prozac?

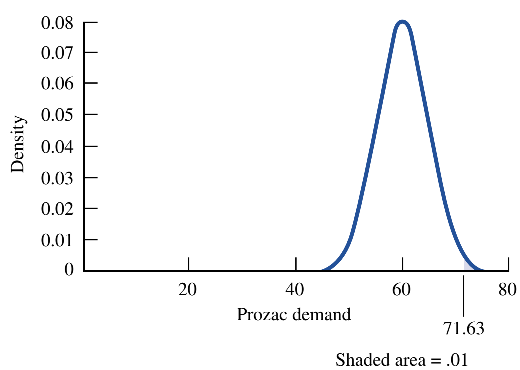


FIGURE 10
99th Percentile of Prozac Demand

Solution Letting $\mathbf{X}$ = annual demand for Prozac, we seek a value x such that $P(\mathbf{X} \geq x) = .01$ or $P(\mathbf{X} \leq x) = .99$. Thus, we seek the 99th percentile of Prozac demand, which we find (in millions) with the formula

=NORMINV(.99,60,5)

Excel returns 71.63, so Lilly must produce 71,630,000 d.o.t. This assumes, of course, that Lilly begins the year with no Prozac on hand. If the company had a beginning inventory of 10 million d.o.t., it would need to produce 61,630,000 d.o.t. during the current year. Figure 10 displays the 99th percentile of Prozac demand.

EXAMPLE 13 Family Income

Family income in Bloomington is normally distributed, with μ = \$30,000 and σ = \$8,000. The poorest 10% of all families in Bloomington are eligible for federal aid. What should the aid cutoff be?

Solution If $\mathbf{X}$ = income of a Bloomington family, we seek an x such that $P(\mathbf{X} \leq x) = .10$. Thus, we seek the 10th percentile of Bloomington family income, which we find with the statement

=NORMINV(.10,30000,8000)

Excel returns \$19,747.59. Thus, aid should be given to all families with incomes smaller than \$19,749.59. Figure 11 displays the 10th percentile of family income.

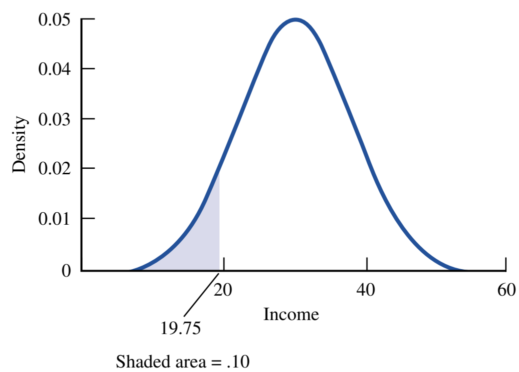


FIGURE 11
10th Percentile of Family Income

EXAMPLE 14

Daily demand for chocolate bars at the Gillis Grocery has a mean of 100 and a variance of 3,000 (chocolate bars)^2. At present, the store has 3,500 chocolate bars in stock. What is the probability that the store will run out of chocolate bars during the next 30 days? Also, how many should Gillis have on hand at the beginning of a 30-day period if the store wants to have only a 1% chance of running out during the 30-day period? Assume that the demands on different days are independent random variables.

Solution Let

$$\mathbf{X}_i = \text{demand for chocolate bars on day } i \qquad (i = 1, 2, \ldots, 30)$$
$$\mathbf{X} = \text{number of chocolate bars demanded in next 30 days}$$

Gillis will run out of stock during the next 30 days if $\mathbf{X} \geq 3{,}500$. The Central Limit Theorem implies that $\mathbf{X} = \mathbf{X}_1 + \mathbf{X}_2 + \cdots + \mathbf{X}_{30}$ can be closely approximated by a normal distribution $\mathbf{X}'$ with $E(\mathbf{X}') = 30(100) = 3{,}000$ and var $\mathbf{X}' = 30(3{,}000) = 90{,}000$ and

$\sigma_{X'} = (90{,}000)^{1/2} = 300$. Then we approximate the probability that Gillis will run out of stock during the next 30 days by

$$P(\mathbf{X}' \geq 3{,}500) = P\left(\frac{\mathbf{X}' - 3000}{300} \geq \frac{3{,}500 - 3{,}000}{300}\right)$$
$$= P(\mathbf{Z} \geq 1.67) = 1 - P(\mathbf{Z} \leq 1.67)$$
$$= 1 - F(1.67) = 1 - .9525 = .0475$$

Let c = number of chocolate bars that should be stocked to have only a 1% chance of running out of chocolate bars within the next 30 days. We seek c satisfying $P(\mathbf{X}' \geq c) = .01$, or

$$P\left(\frac{\mathbf{X}' - 3{,}000}{300} \geq \frac{c - 3{,}000}{300}\right) = .01$$

This is equivalent to

$$P\left(\mathbf{Z} \geq \frac{c - 3{,}000}{300}\right) = .01$$

Since $F(2.33) = P(\mathbf{Z} \leq 2.33) = .99$,

$$\frac{c - 3{,}000}{300} = 2.33 \qquad \text{or} \qquad c = 3{,}699$$

Thus, if Gillis has 3,699 chocolate bars in stock, there is a 1% probability that the store will run out during the next 30 days. (We have defined running out of chocolate bars as having no chocolate bars left at the end of 30 days.)

Alternatively, we could find the probability that the demand is at least 3,500 with the Excel formula

=1 − NORMDIST(3500,3000,300,1)

This formula returns .0475.

We could also have used Excel to determine the level that must be stocked to have a 1% chance of running out as the 99th percentile of the demand distribution. Simply use the formula

=NORMINV(.99,3000,300)

This formula returns the value 3,699.

PROBLEMS

Group A

1 The daily demand for milk (in gallons) at Gillis Grocery is $N(1{,}000, 100)$. How many gallons must be in stock at the beginning of the day if Gillis is to have only a 5% chance of running out of milk by the end of the day?

2 Before burning out, a light bulb gives $\mathbf{X}$ hours of light, where $\mathbf{X}$ is $N(500, 400)$. If we have 3 bulbs, what is the probability that they will give a total of at least 1,460 hours of light?

Group B

3 The number of traffic accidents occurring in Bloomington in a single day has a mean and a variance of 3. What is the probability that during a given year (365-day period), there will be at least 1,000 traffic accidents in Bloomington?

4 Suppose that the number of ounces of soda put into a Pepsi can is normally distributed, with $\mu = 12.05$ oz and $\sigma = .03$ oz.

a Legally, a can must contain at least 12 oz of soda. What fraction of cans will contain at least 12 oz of soda?

b What fraction of cans will contain under 11.9 oz of soda?

c What fraction of cans will contain between 12 and 12.08 oz of soda?

d 1% of all cans will contain more than ______ oz.

e 10% of all cans will contain less than ______ oz.

f Pepsi controls the mean content in a can by setting a timer. For what mean should the timer be set so that only 1 in 1,000 cans will be underfilled?

g Every day, Pepsi produces 10,000 cans. The government inspects 10 randomly chosen cans per day. If at least two are underfilled, Pepsi is fined $10,000. Given that $\mu = 12.05$ oz and $\sigma = .03$ oz, what is the chance that Pepsi will be fined on a given day?

5 Suppose the annual return on Disney stock follows a normal distribution, with mean .12 and standard deviation .30.

a What is the probability that Disney's value will decrease during a year?

b What is the probability that the return on Disney during a year will be at least 20%?

c What is the probability that the return on Disney during a year will be between -6% and 9%?

d There is a 5% chance that the return on Disney during a year will be greater than or equal to ______ .

e There is a 1% chance that the return on Disney during a year will be less than ______ .

f There is a 95% chance that the return on Disney during a year will be between ______ and ______ .

6 The daily demand for six-packs of Coke at Mr. D's follows a normal distribution, with a mean of 120 and a standard deviation of 30. Every Monday, the delivery driver delivers Coke to Mr. D's. If the store wants to have only a 1% chance of running out of Coke by the end of the week, how many six-packs should be ordered for the week? (Assume that orders can be placed Sunday at midnight.)

7 The Coke factory fills bottles of soda by setting a timer on a filling machine. It has been observed that the number of ounces the machine puts in a bottle has a standard deviation of .05 oz. If 99.9% of all bottles are to have at least 16 oz of soda, to what amount should the average amount be set? (*Hint:* Use the Excel Goal Seek feature.)

8 We assemble a large part by joining two smaller parts together. In the past, the smaller parts we have produced have had a mean length of 1″ and a standard deviation of .01″. Assume that the lengths of the smaller parts are normally distributed and are independent.

a What fraction of the larger parts are more than 2.05″ in diameter?

b What fraction of the larger parts are between 1.96″ and 2.02″ in diameter?

9 Weekly Ford sales follow a normal distribution, with a mean of 50,000 cars and a standard deviation of 14,000 cars.

a There is a 1% chance that Ford will sell more than ______ cars during the next year.

b The chance that Ford will sell between 2.4 and 2.7 million cars during the next year is ______ .

10 Warren Dinner has invested in nine different investments. The profits earned on the different investments are independent. The return on each investment follows a normal distribution, with a mean of $500 and a standard deviation of $100.

a There is a 1% chance that the total return on the nine investment is less than ______ .

b The probability that Warren's total return is between $4,000 and $5,200 is ______ .

12.7 *z*-Transforms

Consider a discrete random variable **X** whose only possible values are nonnegative integers. For $n = 0, 1, 2, \ldots$, let $P(\mathbf{X} = n) = a_n$. We define (for $|z| \le 1$) the ***z*-transform of X** (call it $p^T_{\mathbf{X}}(z)$) to be

$$E(z^{\mathbf{X}}) = \sum_{n=0}^{n=\infty} a_n z^n$$

To see why *z*-transforms are useful, note that

$$\left[\frac{dp^T_{\mathbf{X}}(z)}{dz}\right]_{z=1} = \left[\sum_{n=1}^{n=\infty} nz^{n-1}a_n\right]_{z=1} = E(\mathbf{X})$$

Also note that

$$\left[\frac{d^2p^T_{\mathbf{X}}(z)}{dz^2}\right]_{z=1} = \left[\sum_{n=1}^{n=\infty} n(n-1)z^{n-2}a_n\right]_{z=1} = E(\mathbf{X}^2) - E(\mathbf{X})$$