CHAPTER 12

Review of Basic Probability

Chapter Guide. This chapter provides a review of probability laws, random variables, and probability distributions. If you already have had a course in basic probability and statistics, you may skip this chapter. Nevertheless, the chapter provides a useful summary of five common distributions that are used frequently in the book: binomial, Poisson, uniform, exponential, and normal. We have also developed a spreadsheet-based statistical table (file StatTables.xls) that automates the computations of the mean, standard deviation, probabilities, and percentiles of 16 different distributions. Another spreadsheet is provided for histogramming empirical data (file excelMeanVar.xls).

This chapter includes 12 solved examples, 2 spreadsheets, and 44 end-of-section problems. The AMPL/Excel/Solver/TORA programs are in folder ch12Files.

12.1 LAWS OF PROBABILITY

Probability deals with random outcomes of an **experiment**. The conjunction of all the outcomes is referred to as the **sample space**, and a subset of the sample space is known as an **event**. As an illustration, the outcomes of rolling a (six-faced) die are 1, 2, 3, 4, 5, and 6. The set $\{1, 2, 3, 4, 5, 6\}$ defines the associated sample space. An example of an event is that a roll turns up an even value (2, 4, or 6).

An experiment may deal with a continuous sample space as well. For example, the time between failures of an electronic component may assume any nonnegative value.

If an event E occurs m times in an n-trial experiment, then the probability, $P\{E\}$, of realizing the event E is defined as

$$P\{E\} = \lim_{n \to \infty} \frac{m}{n}$$

The definition implies that if the experiment is repeated *indefinitely* $(n \to \infty)$, then the desired probability is represented by $\frac{m}{n}$. You can verify this definition by flipping a coin and observing its outcome: head (H) or tail (T). The longer you repeat the experiment, the closer will be the estimate of $P\{H\}$ (or $P\{T\}$) to the theoretical value of 0.5.

By definition,

$$0 \leq P\{E\} \leq 1$$

An event E is impossible if $P\{E\} = 0$, and certain if $P\{E\} = 1$. For example, in a six-faced die experiment, rolling a 7 is impossible, whereas rolling an integer value from 1 to 6, inclusive, is certain.

PROBLEM SET 12.1A

*1. In a survey conducted in the State of Arkansas high schools to study the correlation between senior year scores in mathematics and enrollment in engineering colleges, 400 out of 1000 surveyed seniors have studied mathematics. Engineering enrollment shows that, of the 1000 seniors, 150 students have studied mathematics and 29 have not. Determine the probabilities of the following events:
   - **(a)** A student who studied mathematics is enrolled in engineering. Is not enrolled in engineering.
   - **(b)** A student neither studied mathematics nor enrolled in engineering.
   - **(c)** A student is not studying engineering.

*2. Consider a random gathering of n persons. Determine the smallest n such that it is more likely than not that two persons or more have the same birthday. (*Hint*: Assume no leap years and that all days of the year are equally likely to be a person's birthday.)

*3. Answer Problem 2 assuming that two or more persons share your birthday.

12.1.1 Addition Law of Probability

For two events, E and F, $E + F$ (or $E \cup F$) represents the **union** of E and F, and EF (or $E \cap F$) represents their **intersection**. The events E and F are **mutually exclusive** if they do not intersect—that is, if the occurrence of one event precludes the occurrence of the other. Based on these definitions, the addition law of probability can be stated as

$$P\{E + F\} = \begin{cases} P\{E\} + P\{F\}, & E \text{ and } F \text{ mutually exclusive} \\ P\{E\} + P\{F\} - P\{EF\}, & \text{otherwise} \end{cases}$$

$P\{EF\}$ is the probability that events E and F occur simultaneously.

Example 12.1-1

Consider the experiment of rolling a die. The sample space of the experiment is $\{1, 2, 3, 4, 5, 6\}$. For a fair die, we have

$$P\{1\} = P\{2\} = P\{3\} = P\{4\} = P\{5\} = P\{6\} = \tfrac{1}{6}$$

Define

$$E = \{1, 2, 3, \text{or } 4\}$$
$$F = \{3, 4, \text{or } 5\}$$

The outcomes 3 and 4 are common between E and F—hence, $EF = \{3 \text{ or } 4\}$. Thus,

$$P\{E\} = P\{1\} + P\{2\} + P\{3\} + P\{4\} = \tfrac{1}{6} + \tfrac{1}{6} + \tfrac{1}{6} + \tfrac{1}{6} = \tfrac{2}{3}$$

$$P\{F\} = P\{3\} + P\{4\} + P\{5\} = \tfrac{1}{2}$$

$$P\{EF\} = P\{3\} + P\{4\} = \tfrac{1}{3}$$

It then follows that

$$P\{E + F\} = P\{E\} + P\{F\} - P\{EF\} = \tfrac{2}{3} + \tfrac{1}{2} - \tfrac{1}{3} = \tfrac{5}{6}$$

Intuitively, the result makes sense because $(E + F) = \{1, 2, 3, 4, 5\}$, whose probability of occurrence is $\frac{5}{6}$.

PROBLEM SET 12.1B

1. A fair 6-faced die is tossed twice. Letting E and F represent the outcomes of the two tosses, compute the following probabilities:
   (a) The sum of E and F is 11.
   (b) The sum of E and F is even.
   (c) The sum of E and F is odd and greater than 3.
   (d) E is even less than 6 and F is odd greater than 1.
   (e) E is greater than 2 and F is less than 4.
   (f) E is 4 and the sum of E and F is odd.
2. Suppose that you roll two dice independently and record the number that turns up for each die. Determine the following:
   (a) The probability that both numbers are even.
   (b) The probability that the sum of the two numbers is 10.
   (c) The probability that the two numbers differ by at least 3.

*3. You can toss a fair coin up to 7 times. You will win $100 if three tails appear before a head is encountered. What are your chances of winning?

*4. Ann, Jim, John, and Liz are scheduled to compete in a racquetball tournament. Ann is twice as likely to beat Jim, and Jim is at the same level as John. Liz's past winning record against John is one out of three. Determine the following:
   (a) The probability that Jim will win the tournament.
   (b) The probability that a woman will win the tournament.
   (c) The probability that no woman will win.

12.1.2 Conditional Law of Probability

Given the two events E and F with $P\{F\} > 0$, the conditional probability of E given F, $P\{E|F\}$, is defined as

$$P\{E|F\} = \frac{P\{EF\}}{P\{F\}}, \quad P\{F\} > 0$$

If E is a subset of (i.e., contained in) F, then $P\{EF\} = P\{E\}$.

The two events, E and F, are *independent* if, and only if,

$$P\{E|F\} = P\{E\}$$

In this case, the conditional probability law reduces to

$$P\{EF\} = P\{E\}P\{F\}$$

Example 12.1-2

You are playing a game in which another person is rolling a die. You cannot see the die, but you are given information about the outcomes. Your job is to predict the outcome of each roll. Determine the probability that the outcome is a 6, given that you are told that the roll has turned up an even number.

Let $E = \{6\}$, and define $F = \{2, 4, \text{or } 6\}$. Thus,

$$P\{E|F\} = \frac{P\{EF\}}{P\{F\}} = \frac{P\{E\}}{P\{F\}} = \left(\frac{1/6}{1/2}\right) = \frac{1}{3}$$

Note that $P\{EF\} = P\{E\}$ because E is a subset of F.

PROBLEM SET 12.1C

1. In Example 12.1-2, suppose that you are told that the outcome is less than 6.
   (a) Determine the probability of getting an even number.
   (b) Determine the probability of getting an odd number larger than one.
2. The stock of WalMark Stores, Inc., trades on the New York Stock Exchange under the symbol WMS. Historically, the price of WMS goes up with the increase in the Dow average 60% of the time and goes down with the Dow 25% of the time. There is also a 5% chance that WMS will go up when the Dow goes down and 10% that it will go down when the Dow goes up.
   (a) Determine the probability that WMS will go up regardless of the Dow.
   (b) Find the probability that WMS goes up given that the Dow is up.
   (c) What is the probability WMS goes down given that Dow is down?

*3. Graduating high school seniors with an ACT score of at least 26 can apply to two universities, A and B, for admission. The probability of being accepted in A is .4 and in B .25. The chance of being accepted in both universities is only 15%.
   (a) Determine the probability that the student is accepted in B given that A has granted admission as well.
   (b) What is the probability that admission will be granted in A given that the student was accepted in B?

4. Prove that if the probability $P\{A|B\} = P\{A\}$, then A and B must be independent.

5. *Bayes' theorem.*[1] Given the two events A and B, show that

$$P\{A|B\} = \frac{P\{B|A\}P\{A\}}{P\{B\}}, P\{B\} > 0$$

6. A retailer receives 75% of its batteries from Factory A and 25% from Factory B. The percentages of defectives produced by A and B are known to be 1% and 2%, respectively. A customer has just bought a battery randomly from the retailer.
   (a) What is the probability that the battery is defective?
   (b) If the battery you bought is defective, what is the probability that it came from Factory A? (*Hint*: Use Bayes' theorem in Problem 5.)

*7. Statistics show that 70% of all men have some form of prostate cancer. The PSA test will show positive 90% of the time for afflicted men and 10% of the time for healthy men. What is the probability that a man who tested positive does have prostate cancer?

12.2 RANDOM VARIABLES AND PROBABILITY DISTRIBUTIONS

The outcomes of an experiment either are naturally numeric or can be coded numerically. For example, the outcomes of rolling a die are naturally numeric—namely, 1, 2, 3, 4, 5, or 6. Conversely, the testing of an item produces two outcomes: bad and good. In such a case, we can use the numeric code (0, 1) to represent (bad, good). The numeric representation of the outcomes produces what is known as a **random variable**.

A random variable, x, may be **discrete** or **continuous**. For example, the random variable associated with the die-rolling experiment is discrete with $x = 1, 2, 3, 4, 5$, or 6, whereas the interarrival time at a service facility is continuous with $x \geq 0$.

Each continuous or discrete random variable x is quantified by a **probability density function** (pdf), $f(x)$ or $p(x)$. These functions must satisfy the conditions in the following table:

	Random variable, x	
Characteristic	*Discrete*	*Continuous*
Applicability range	$x = a, a + 1, \ldots, b$	$a \leq x \leq b$
Conditions for the pdf	$p(x) \geq 0, \sum_{x=a}^{b} p(x) = 1$	$f(x) \geq 0, \int_a^b f(x)dx = 1$

A pdf, $p(x)$ or $f(x)$, must be nonnegative (otherwise, the probability of some event may be negative!). Also, the probability of the entire sample space must equal 1.

An important probability measure is the **cumulative distribution function** (CDF), defined as

$$P\{x \leq X\} = \begin{cases} P(X) = \sum_{x=a}^{X} p(x), & x \text{ discrete} \\ F(X) = \int_a^X f(x)\,dx, & x \text{ continuuous} \end{cases}$$

[1]Section 13.2.2 provides a more detailed presentation of Bayes' theorem.

Example 12.2-1

Consider the case of rolling a fair die. The random variable $x = \{1, 2, 3, 4, 5, 6\}$ represents the face of the die that turns up. The associated pdf and CDF are

$$p(x) = \tfrac{1}{6}, x = 1, 2, \ldots, 6$$
$$P(X) = \tfrac{X}{6}, X = 1, 2, \ldots, 6$$

Figure 12.1 graphs the two functions. The pdf $p(x)$ is a **uniform discrete function** because all the values of the random variables occur with equal probabilities.

The continuous counterpart of uniform $p(x)$ is illustrated by the following experiment. A needle of length l is pivoted in the center of a circle whose diameter also equals l. After marking an arbitrary reference point on the circumference, we spin the needle clockwise and measure the circumference distance x from where the pointer stops to the marked point. Thus, the random variable x is continuous in the range $0 \le x \le \pi l$. There is no reason to believe that the needle will tend to stop more often in a specific region of the circumference. Hence, all the values of x in the specified range are equally likely to occur, and the distribution of x must be uniform.

The pdf of x, $f(x)$, is defined as

$$f(x) = \frac{1}{\pi l}, 0 \le x \le \pi l$$

The associated CDF, $F(X)$, is computed as

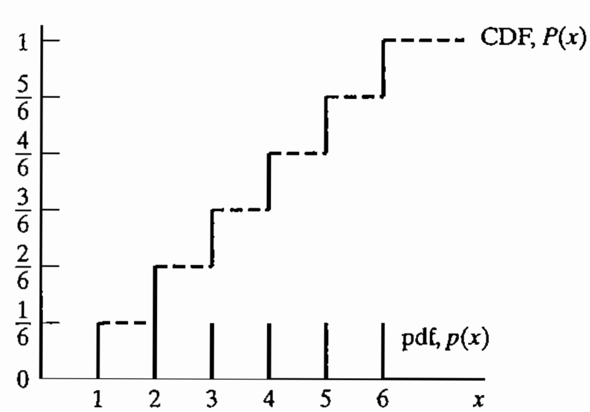
$$F(X) = P\{x \le X\} = \int_0^X f(x)\,dx = \int_0^X \frac{1}{\pi l}dx = \frac{X}{\pi l}, 0 \le X \le \pi l$$

Figure 12.2 graphs the two functions.

PROBLEM SET 12.2A

1. The number of units, x, needed of an item is discrete from 1 to 5. The probability, $p(x)$, is directly proportional to the number of units needed. The constant of proportionality is K.
   (a) Determine the pdf and CDF of x, and graph the resulting functions.
   (b) Find the probability that x is an even value.

FIGURE 12.1
CDF and pdf for rolling a fair die



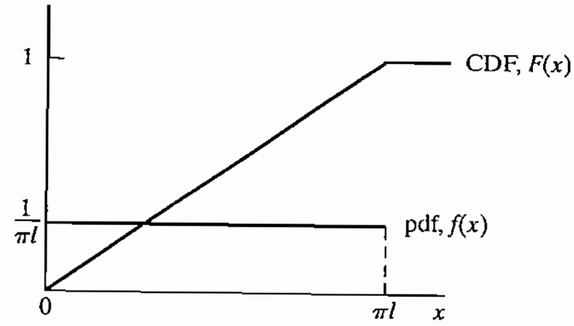


FIGURE 12.2
CDF and pdf for spinning a needle

2. Consider the following function:

$$f(x) = \frac{k}{x^2},\ 10 \le x \le 20$$

*(a) Find the value of the constant k that will make $f(x)$ a pdf.

(b) Determine the CDF, and find the probability that x is (i) larger than 12, and (ii) between 13 and 15.

*3. The daily demand for unleaded gasoline is uniformly distributed between 750 and 1250 gallons. The gasoline tank, with a capacity of 1100 gallons, is refilled daily at midnight. What is the probability that the tank will be empty just before a refill?

12.3 EXPECTATION OF A RANDOM VARIABLE

Given that $h(x)$ is a real function of a random variable x, we define the **expected value** of $h(x)$, $E\{h(x)\}$, as the (long-run) weighted average with respect to the pdf of x. Mathematically, given that $p(x)$ and $f(x)$ are, respectively, the discrete and continuous pdfs of x, $E\{h(x)\}$ is computed as

$$E\{h(x)\} = \begin{cases} \sum_{x=a}^{b} h(x)p(x), & x \text{ discrete} \\ \int_a^b h(x)f(x)\,dx, & x \text{ continuous} \end{cases}$$

Example 12.3-1

During the first week of each month, I (like many people) pay all my bills and answer a few letters. I usually buy 20 first-class mail stamps each month for this purpose. The number of stamps I will be using varies randomly between 10 and 24, with equal probabilities. What is the average number of stamps left?

The pdf of the number of stamps used is

$$p(x) = \tfrac{1}{15},\ x = 10, 11, \ldots, 24.$$

The number of stamps left is given as

$$h(x) = \begin{cases} 20 - x, & x = 10, 11, \ldots, 19 \\ 0, & \text{otherwise} \end{cases}$$

Thus,

$$E\{h(x)\} = \tfrac{1}{15}[(20 - 10) + (20 - 11) + (20 - 12) + \cdots + (20 - 19)] + \tfrac{5}{15}(0)$$
$$= 3\tfrac{2}{3}$$

The product $\frac{5}{15}(0)$ is needed to complete the expected value of $h(x)$. Specifically, the probability of being left with *zero* extra stamps equals the probability of needing 20 stamps or more—that is,

$$P\{x \geq 20\} = p(20) + p(21) + p(22) + p(23) + p(24) = 5\left(\tfrac{1}{15}\right) = \tfrac{5}{15}$$

PROBLEM SET 12.3A

1. In Example 12.3-1, compute the average number of extra stamps needed to meet your maximum possible demand.
2. The results of Example 12.3-1 and of Problem 1 show *positive* averages for *both* the surplus and shortage of stamps. Are these results inconsistent? Explain.

*3. The owner of a newspaper stand receives 50 copies of *Al Ahram* newspaper every morning. The number of copies sold daily, x, varies randomly according to the following probability distribution:

$$p(x) = \begin{cases} \frac{1}{45}, x = 35, 36, \ldots, 49 \\ \frac{1}{30}, x = 50, 51, \ldots, 59 \\ \frac{1}{33}, x = 60, 61, \ldots, 70 \end{cases}$$

(a) Determine the probability that the owner will sell out completely.

(b) Determine the expected number of unsold copies per day.

(c) If the owner pays 50 cents a copy and sells it for \$1.00. Determine the owner's expected net income per day.

12.3.1 Mean and Variance (Standard Deviation) of a Random Variable

The **mean** of x, $E\{x\}$, is a numeric measure of the central tendency (or weighted sum) of the random variable. The **variance**, var$\{x\}$, is a measure of the dispersion or deviation of x around the mean $E\{x\}$. Its square root is known as the **standard deviation** of x, stdDev $\{x\}$. A larger standard deviation means a higher degree of uncertainty regarding the random variable. Specifically, when the value of a variable is known with certainty, its standard deviation is zero.

The formulas for the mean and variance can be derived from the general definition of $E\{h(x)\}$ as follows: For $E\{x\}$, use $h(x) = x$, and for var$\{x\}$ use $h(x) = (x - E\{x\})^2$. Thus,

$$E\{x\} = \begin{cases} \displaystyle\sum_{x=a}^{b} xp(x), & x \text{ discrete} \\ \displaystyle\int_a^b xf(x)\,dx, & x \text{ continuous} \end{cases}$$

$$\text{var}\{x\} = \begin{cases} \displaystyle\sum_{x=a}^{b}(x - E\{x\})^2 p(x), & x \text{ discrete} \\ \displaystyle\int_a^b (x - E\{x\})^2 f(x)\,dx, & x \text{ continuous} \end{cases}$$

$$\text{stdDev}\{x\} = \sqrt{\text{var}\{x\}}$$

We can see the basis for the development of the formulas more readily by examining the discrete case. Here, $E\{x\}$ is the *weighted sum* of the discrete values of x. Also, $\text{var}\{x\}$ is the *weighted sum* of the square of the deviation around $E\{x\}$. The continuous case can be interpreted similarly, with integration replacing summation.

Example 12.3-2

We compute the mean and variance for each of the two experiments in Example 12.2-1.

Case 1 (Die Rolling). The pdf is $p(x) = \frac{1}{6}, x = 1, 2, \ldots, 6$. Thus,

$$E\{x\} = 1\left(\tfrac{1}{6}\right) + 2\left(\tfrac{1}{6}\right) + 3\left(\tfrac{1}{6}\right) + 4\left(\tfrac{1}{6}\right) + 5\left(\tfrac{1}{6}\right) + 6\left(\tfrac{1}{6}\right) = 3.5$$

$$\text{var}\{x\} = \left(\tfrac{1}{6}\right)\{(1 - 3.5)^2 + (2 - 3.5)^2 + (3 - 3.5)^2 + (4 - 3.5)^2 + (5 - 3.5)^2 + (6 - 3.5)^2\} = 2.917$$

$$\text{stdDev}(x) = \sqrt{2.917} = 1.708$$

Case 2 (Needle Spinning). Suppose that the length of the needle is 1 inch. Then,

$$f(x) = \frac{1}{3.14}, \quad 0 \le x \le 3.14$$

The mean and variance are computed as

$$E(x) = \int_0^{3.14} x\left(\tfrac{1}{3.14}\right) dx = 1.57 \text{ inch}$$

$$\text{var}(x) = \int_0^{3.14} (x - 1.57)^2\left(\tfrac{1}{3.14}\right) dx = .822 \text{ inch}^2$$

$$\text{stdDev}(x) = \sqrt{.822} = .906 \text{ inch}$$

Excel Moment

Template exelStatTables.xls is designed to compute the mean, standard deviation, probabilities, and percentiles for 16 common pdfs, including the discrete and continuous uniform distributions of Example 12.3-2. The use of the template is self-explanatory.

PROBLEM SET 12.3B

*1. Compute the mean and variance of the random variable defined in Problem 1, Set 12.2a.

2. Compute the mean and variance of the random variable in Problem 2, Set 12.2a.

3. Show that the mean and variance of a uniform random variable x, $a \le x \le b$, are

$$E\{x\} = \frac{b + a}{2}$$
$$\text{var}\{x\} = \frac{(b - a)^2}{12}$$

4. Given the pdf $f(x)$, $a \le x \le b$, prove that

$$\text{var}\{x\} = E\{x^2\} - (E\{x\})^2$$

5. Given the pdf $f(x)$, $a \le x \le b$, and $y = cx + d$, where c and d are constants. Prove that

$$E\{y\} = cE\{x\} + d$$
$$\text{var}\{y\} = c^2\,\text{var}\{x\}$$

12.3.2 Mean and Variance of Joint Random Variables

Consider the two continuous random variables x_1, $a_1 \le x_1 \le b_1$, and x_2, $a_2 \le x_2 \le b_2$. Define $f(x_1, x_2)$ as the **joint pdf** of x_1 and x_2 and $f_1(x_1)$ and $f_2(x_2)$ as the **marginal pdfs** of x_1 and x_2, respectively. Then

$$f(x_1, x_2) \ge 0,\ a_1 \le x_1 \le b_1,\ a_2 \le x_2 \le b_2$$
$$\int_{a_1}^{b_1} dx_1 \int_{a_2}^{b_2} dx_2 f(x_1, x_2) = 1$$
$$f_1(x_1) = \int_{a_2}^{b_2} f(x_1, x_2)\, dx_2$$
$$f_2(x_2) = \int_{a_1}^{b_1} f(x_1, x_2)\, dx_1$$
$$f(x_1, x_2) = f_1(x_1) f_2(x_2), \text{ if } x_1 \text{ and } x_2 \text{ are independent}$$

The same formulas apply to discrete pdfs, replacing integration with summation.

For the special case $y = c_1x_1 + c_2x_2$, where the random variables x_1 and x_2 are jointly distributed according to the pdf $f(x_1, x_2)$, we can prove that

$$E\{c_1x_1 + c_2x_2\} = c_1E\{x_1\} + c_2E\{x_2\}$$
$$\text{var}\{c_1x_1 + c_2x_2\} = c_1^2\,\text{var}\{x_1\} + c_2^2\,\text{var}\{x_2\} + 2c_1c_2\,\text{cov}\{x_1, x_2\}$$

where

$$\begin{aligned}\text{cov}\{x_1, x_2\} &= E\{(x_1 - E\{x_1\})(x_2 - E\{x_2\})\\ &= E(x_1x_2 - x_1E\{x_2\} - x_2E\{x_1\} + E\{x_1\}E\{x_2\})\\ &= E\{x_1x_2\} - E\{x_1\}E\{x_2\}\end{aligned}$$

If x_1 and x_2 are *independent*, then $E\{x_1 x_2\} = E\{x_1\}E\{x_2\}$ and $\text{cov}\{x_1, x_2\} = 0$. The converse is not true, in the sense that two *dependent* variables may have zero covariance.

Example 12.3-3

A lot includes four defective (D) items and six good (G) ones. You select one item randomly and test it. Then, without replacement, you test a second item. Let the random variables x_1 and x_2 represent the outcomes for the first and second items, respectively.

a. Determine the joint and marginal pdfs of x_1 and x_2.
b. Suppose that you get \$5 for each good item you select but must pay \$6 if it is defective. Determine the mean and variance of your revenue after two items have been selected.

Let $p(x_1, x_2)$ be the joint pdf of x_1 and x_2, and define $p_1(x_1)$ and $p_2(x_2)$ as the respective marginal pdfs. First, we determine $p_1(x_1)$ as

$$p_1(G) = \tfrac{6}{10} = .6,\ p_1(D) = \tfrac{4}{10} = .4$$

Next, we know that x_2, the second outcome, depends on x_1. Hence, to determine $p_2(x_2)$, we first determine the joint pdf $p(x_1, x_2)$, from which we can determine the marginal distribution $p_2(x_2)$.

$$\begin{aligned}
P\{x_2 = G|x_1 = G\} &= \tfrac{5}{9}\\
P\{x_2 = G|x_1 = B\} &= \tfrac{6}{9}\\
P\{x_2 = B|x_1 = G\} &= \tfrac{4}{9}\\
P\{x_2 = B|x_1 = B\} &= \tfrac{3}{9}
\end{aligned}$$

To determine $p(x_1, x_2)$, we use the formula $P\{AB\} = P\{A|B\}P\{B\}$ (see Section 12.1.2).

$$\begin{aligned}
p\{x_2 = G, x_1 = G\} &= \tfrac{5}{9} \times \tfrac{6}{10} = \tfrac{5}{15}\\
p\{x_2 = G, x_1 = B\} &= \tfrac{6}{9} \times \tfrac{4}{10} = \tfrac{4}{15}\\
p\{x_2 = B, x_1 = G\} &= \tfrac{4}{9} \times \tfrac{6}{10} = \tfrac{4}{15}\\
p\{x_2 = B, x_1 = B\} &= \tfrac{3}{9} \times \tfrac{4}{10} = \tfrac{2}{15}
\end{aligned}$$

The marginal distributions, $p_1(x_1)$ and $p_2(x_2)$, can be determined by first summarizing the joint distribution, $p(x_1, x_2)$, in a table format and then adding the respective rows and columns, as the following table shows.

	$x_2 = G$	$x_2 = B$	$p_1(x_1)$
$x_1 = G$	$\frac{5}{15}$	$\frac{4}{15}$	$\frac{9}{15} = .6$
$x_1 = B$	$\frac{4}{15}$	$\frac{2}{15}$	$\frac{6}{15} = .4$
$p_2(x_2)$	$\frac{9}{15} = .6$	$\frac{6}{15} = .4$	

It is interesting that, contrary to intuition, $p_1(x_1) = p_2(x_2)$.

The expected revenue can be determined from the joint distribution by recognizing that G produces \$5 and B yields $-$\$6. Thus,

$$\text{Expected revenue} = (5 + 5)\tfrac{5}{15} + (5 - 6)\tfrac{4}{15} + (-6 + 5)\tfrac{4}{15} + (-6 - 6)\tfrac{2}{15} = \$1.20$$

The same result can be determined by recognizing that the expected revenue for both selections equals the sum of the expected revenue for each individual selection (even though the two variables are *not* independent). This means that 12

$$\begin{aligned}\text{Expected revenue} &= \text{Selection 1 expected revenue} + \text{Selection 2 expected revenue}\\ &= (5 \times .6 - 6 \times .4) + (5 \times .6 - 6 \times .4) = \$1.20\end{aligned}$$

To compute the variance of the total revenue, we note that 12

$$\text{var}\{\text{revenue}\} = \text{var}\{\text{revenue 1}\} + \text{var}\{\text{revenue 2}\} + 2\,\text{cov}\{\text{revenue 1, revenue 2}\}$$

Because $p_1(x_1) = p_2(x_2)$, then $\text{var}\{\text{revenue 1}\} = \text{var}\{\text{revenue 2}\}$. To compute the variance, we use the formula

$$\text{var}\{x\} = E\{x^2\} - (E\{x\})^2$$

(See Problem 4, Set 12.3b.) Thus,

$$\text{var}\{\text{revenue 1}\} = [5^2 \times .6 + (-6)^2 \times .4] - .6^2 = 29.04$$

Next, to compute the covariance, we use the formula

$$\text{cov}\{x_1, x_2) = E\{x_1x_2\} - E\{x_1\}E\{x_2\}$$

The term $E\{x_1x_2\}$ can be computed from the joint pdf of x_1 and x_2. Thus, we have

$$\begin{aligned}\text{Convariance} = {} & [(5 \times 5)\left(\tfrac{5}{15}\right) + (5 \times -6)\left(\tfrac{4}{15}\right) + (-6 \times 5)\left(\tfrac{4}{15}\right)\\ & + (-6 \times -6)\left(\tfrac{2}{15}\right)] - .6 \times .6 = -3.23\end{aligned}$$

Thus,

$$\text{Variance} = 29.04 + 29.04 + 2(-3.23) = 51.62$$

PROBLEM SET 12.3C

1. The joint pdf of x_1 and x_2, $p(x_1, x_2)$, is

	$x_2 = 3$	$x_2 = 5$	$x_2 = 7$
$x_1 = 1$	.2	0	.2
$x_1 = 2$	0	.2	0
$x_1 = 3$	.2	0	.2

*(a) Find the marginal pdfs $p_1(x_1)$ and $p_2(x_2)$.
*(b) Are x_1 and x_2 independent?
(c) Compute $E\{x_1 + x_2\}$.
(d) Compute $\text{cov}\{x_1, x_2\}$.
(e) Compute $\text{var}\{5x_1 - 6x_2\}$.

12.4 FOUR COMMON PROBABILITY DISTRIBUTIONS

In Sections 12.2 and 12.3 we discussed the (discrete and continuous) uniform distribution. This section presents four additional pdfs that are encountered often in operations research studies: the discrete binomial and Poisson, and the continuous exponential and normal.

12.4.1 Binomial Distribution

Suppose that a manufacturer produces a certain product in lots of n items each. The fraction of defective items in each lot, p, is estimated from historical data. We are interested in determining the pdf of the number of defectives in a lot.

There are $C_x^n = \frac{n!}{x!(n-x)!}$ distinct combinations of x defectives in a lot of n items, and the probability of getting each combination is $p^x(1-p)^{n-x}$. It follows (from the addition law of probability) that the probability of k defectives in a lot of n items is

$$P\{x = k\} = C_k^n p^k (1-p)^{n-k}, k = 0, 1, 2, \ldots, n$$

This is the binomial distribution with parameters n and p. Its mean and variance are given by

$$E\{x\} = np$$
$$\text{var}\{x\} = np(1-p)$$

Example 12.4-1

John Doe's daily chores require making 10 round trips by car between two towns. Once through with all 10 trips, Mr. Doe can take the rest of the day off, a good enough motivation to drive above the speed limit. Experience shows that there is a 40% chance of getting a speeding ticket on any round trip.

a. What is the probability that the day will end without a speeding ticket?
b. If each speeding ticket costs \$80, what is the average daily fine?

The probability of getting a ticket on any one trip is $p = .4$. Thus, the probability of not getting a ticket in any one day is

$$P\{x = 0\} = C_0^{10}(.4)^0(.6)^{10} = .006$$

This means that there is less than 1% chance of finishing the day without a fine. In fact, the average fine per day can be computed as

$$\text{Average fine} = \$80E\{x\} = \$80(np) = 80 \times 10 \times .4 = \$320$$

Remark. $P\{x = 0\}$ can be computed using excelStatTables.xls. Enter 10 in F7, .4 in G7, and 0 in J7. The answer, $P\{x = 0\} = .006047$, is given in M7.

PROBLEM SET 12.4A

*1. A fair die is rolled 10 times. What is the probability that the rolled die will not show an even number?

2. Suppose that five fair coins are tossed independently. What is the probability that exactly one of the coins will be different from the remaining four?

*3. A fortune teller claims to predict whether or not people will amass financial wealth in their lifetime by examining their handwriting. To verify this claim, 10 millionaires and 10 university professors were asked to provide samples of their handwriting. The samples are then paired, one millionaire and one professor, and presented to the fortune teller. We say that the claim is true if the fortune teller makes at least eight correct predictions. What is the probability that the claim is proved true by a "fluke"?

4. In a gambling casino game you are required to select a number from 1 to 6 before the operator rolls three fair dice simultaneously. The casino pays you as many dollars as the number of dice that match your selection. If there is no match, you pay the casino only $1. What is your long-run expected payoff from this game?

5. Suppose that you play the following game: You throw 2 fair dice. If there is no match, you pay 10 cents. If there is a match, you get 50 cents. What is the expected payoff for the game?

6. Prove the formulas for the mean and variance of the binomial distribution.

12.4.2 Poisson Distribution

Customers arrive at a bank or a grocery store in a "totally random" fashion, meaning that we cannot predict when someone will arrive. The pdf describing the *number* of such arrivals during a specified period is the Poisson distribution.

Let x be the number of events (e.g., arrivals) that take place during a specified time unit (e.g., a minute or an hour). Given that λ is a known constant, the Poisson pdf is defined as

$$P\{x = k\} = \frac{\lambda^k e^{-\lambda}}{k!}, k = 0, 1, 2, \ldots$$

The mean and variance of the Poisson are

$$E\{x\} = \lambda$$
$$\text{var}\{x\} = \lambda$$

The formula for the mean reveals that λ must represent the rate at which events occur.

The Poisson distribution figures prominently in the study of queues (see Chapter 15).

Example 12.4-2

Repair jobs arrive at a small-engine repair shop in a totally random fashion at the rate of 10 per day.

a. What is the average number of jobs that are received daily at the shop?

b. What is the probability that no jobs will arrive during any 1 hour, assuming that the shop is open 8 hours a day?

The average number of jobs received per day equals $\lambda = 10$ jobs per day. To compute the probability of no arrivals per *hour*, we need to compute the arrival rate per hour—namely, $\lambda_{\text{hour}} = \frac{10}{8} = 1.25$ jobs per hour. Thus

$$\begin{aligned} P\{\text{no arrivals per hour}\} &= \frac{(\lambda_{\text{hour}})^0 e^{-\lambda_{\text{hour}}}}{0!} \\ &= \frac{1.25^0 e^{-1.25}}{0!} = .2865 \end{aligned}$$

Remark. The probability above can be computed with excelStatTables.xls. Enter 1.25 in F16 and 0 in J16. The answer, .286505, appears in M16.

PROBLEM SET 12.4B

*1. Customers arrive at a service facility according to a Poisson distribution at the rate of four per minute. What is the probability that at least one customer will arrive in any given 30-second interval?

2. The Poisson distribution with parameter λ approximates the binomial distribution with parameters (n, p) when $n \to \infty$, $p \to 0$, and $np \to \lambda$. Demonstrate this result for the situation where a manufactured lot is known to contain 1% defective items. If a sample of 10 items is taken from the lot, compute the probability of at most one defective item in a sample, first by using the (exact) binomial distribution and then by using the (approximate) Poisson distribution. Show that the approximation will not be acceptable if the value of p is increased to, say, 0.5.

*3. Customers arrive randomly at a checkout counter at the average rate of 20 per hour.
   (a) Determine the probability that the counter is idle.
   (b) What is the probability that at least two people are in line awaiting service?

4. Prove the formulas for the mean and variance of the Poisson distribution.

12.4.3 Negative Exponential Distribution

If the *number* of arrivals at a service facility during a specified period follows the Poisson distribution (Section 12.4.2), then, automatically, the distribution of the *time interval* between successive arrivals must follow the negative exponential (or, simply, the exponential) distribution. Specifically, if λ is the rate at which Poisson events occur, then the distribution of time between successive arrivals, x, is

$$f(x) = \lambda e^{-\lambda x}, x > 0$$

Figure 12.3 graphs $f(x)$.

The mean and variance of the exponential distribution are

$$E\{x\} = \frac{1}{\lambda}$$

$$\text{var}\{x\} = \frac{1}{\lambda}$$

The mean $E\{x\}$ is consistent with the definition of λ. If λ is the *rate* at which events occur, then $\frac{1}{\lambda}$ is the average time interval between successive events.

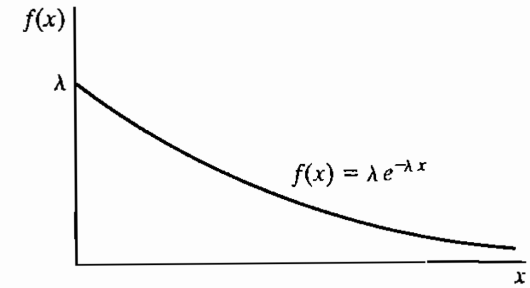


FIGURE 12.3

Probability density function of the exponential distribution

Example 12.4-3

Cars arrive at a gas station randomly every 2 minutes, on the average. Determine the probability that the interarrival time of cars does not exceed 1 minute.

The desired probability is of the form $P\{x \le A\}$, where $A = 1$ minute in the present example. The determination of this probability is the same as computing the CDF of x—namely,

$$\begin{aligned} P\{x \le A\} &= \int_0^A \lambda e^{-\lambda x}\,dx \\ &= -e^{-\lambda x}\big|_o^A \\ &= 1 - e^{-\lambda A} \end{aligned}$$

The arrival rate for the example is computed as

$$\lambda = \tfrac{1}{2} \text{ arrival per minute}$$

Thus,

$$P\{x \le 1\} = 1 - e^{-\left(\frac{1}{2}\right)(1)} = .3934$$

Remark. You can use excelStatTables.xls to compute the probability above. Enter .5 in F9, 1 in J9. The answer, .393468, appears in O9.

PROBLEM SET 12.4C

*1. Walmark Store gets its customers from within town and the surrounding rural areas. Town customers arrive at the rate of 5 per minute, and rural customers arrive at the rate of 7 per minute. Arrivals are totally random. Determine the probability that the interarrival time for all customers is less than 5 seconds.

2. Prove the formulas for the mean and variance of the exponential distribution.

12.4.4 Normal Distribution

The normal distribution describes many random phenomena that occur in everyday life, including test scores, weights, heights, and many others. The pdf of the normal distribution is defined as

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

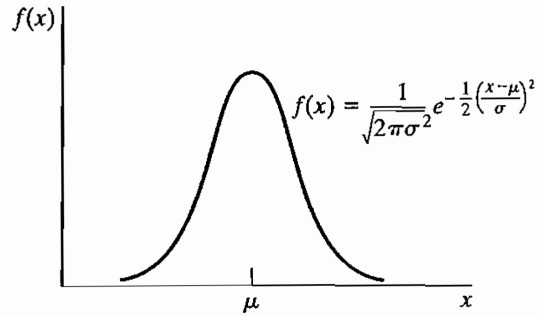


FIGURE 12.4
Probability density function of the normal random variable

The mean and variance are

$$E\{x\} = \mu$$
$$\text{var}\{x\} = \sigma^2$$

The notation $N(\mu, \sigma)$ is usually used to represent a normal distribution with mean μ and standard deviation σ.

Figure 12.4 graphs the normal pdf, $f(x)$. The function is always symmetrical around the mean μ.

An important property of the normal random variable is that it approximates the distribution of the average of a sample taken from *any* distribution. This remarkable result is based on the following theorem:

Central Limit Theorem. *Let $x_1, x_2, \ldots,$ and x_n be independent and identically distributed random variables, each with mean μ and standard deviation σ, and define*

$$s_n = x_1 + x_2 + \cdots + x_n$$

As n becomes large $(n \rightarrow \infty)$, the distribution of s_n becomes asymptotically normal with mean $n\mu$ and variance $n\sigma^2$, regardless of the original distribution of $x_1, x_2, \ldots,$ and x_n.

The central limit theorem particularly tells us that the distribution of the *average* of a sample of size n drawn from any distribution is asymptotically normal with mean μ and variance $\frac{\sigma^2}{n}$. This result has important applications in statistical quality control.

The CDF of the normal random variable cannot be determined in a closed form. As a result, normal tables (Table 1 in Appendix B or excelStatTables.xls) have been prepared for this purpose. These tables apply to the **standard normal** with mean zero and standard deviation 1—that is, $N(0, 1)$. Any normal random variable, x (with mean μ and standard deviation σ), can be converted to a standard normal, z, by using the transformation

$$z = \frac{x - \mu}{\sigma}$$

Over 99% of the area under any normal distribution is enclosed in the range $\mu - 3\sigma \leq x \leq \mu + 3\sigma$, known as the **6-sigma limits**.

Example 12.4-4

The inside diameter of a cylinder has the specifications $1 \pm .03$ in. The machining process output follows a normal distribution with mean 1 cm and standard deviation .1 cm. Determine the percentage of production that will meet the specifications.

Let x represent the output of the process. The probability that a cylinder will meet specifications is

$$P\{1 - .03 \le x \le 1 + .03\} = P\{.97 \le x \le 1.03\}$$

Given $\mu = 1$ and $\sigma = .1$, the equivalent standard normal probability statement is

$$\begin{aligned}
P\{.97 \le x \le 1.03\} &= P\left\{\tfrac{.97-1}{.1} \le z \le \tfrac{1.03-1}{.1}\right\} \\
&= P\{-.3 \le z \le .3\} \\
&= P\{z \le .3\} - P\{z \le -.3\} \\
&= P\{z \le .3\} - P\{z \ge .3\} \\
&= P\{z \le .3\} - [1 - P\{z \le .3\}] \\
&= 2P\{z \le .3\} - 1 \\
&= 2 \times .6179 - 1 \\
&= .2358
\end{aligned}$$

The given probability statements can be justified by picturing the shaded area in Figure 12.5. Notice that $P\{z \le -.3\} = 1 - P\{z \le .3\}$ because of the symmetry of the pdf. The value .6179 ($= P\{z \le .3\}$ is obtained from the standard normal table (Table 1 in Appendix B).

Remark. $P\{a \le x \le b\}$ can be computed directly from excelStatTables.xls. Enter 1 in F15, .1 in G15, .97 in J15, and 1.03 in K15. The answer, .235823, appears in Q15.

PROBLEM SET 12.4D

1. The college of engineering at U of A requires a minimum ACT score of 26. The test score among high school seniors in a given school district is normally distributed with mean 22 and standard deviation 2.
   - **(a)** Determine the percentage of high school seniors who are potential engineering recruits.
   - **(b)** If U of A does not accept any student with an ACT score less than 17, determine the percentage of students that will not be eligible for admission at U of A.

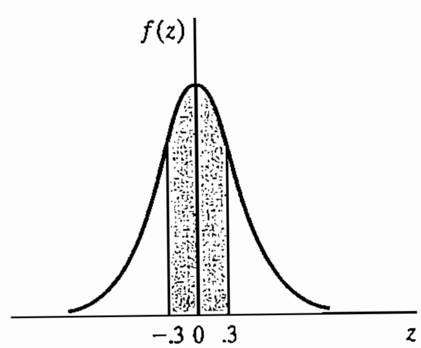


FIGURE 12.5
Calculation of $P\{-.3 \le z \le .3\}$ in a standard normal distribution