Chapter 6

Some Continuous Probability Distributions

6.1 Continuous Uniform Distribution

One of the simplest continuous distributions in all of statistics is the **continuous uniform distribution**. This distribution is characterized by a density function that is "flat," and thus the probability is uniform in a closed interval, say $[A, B]$. Although applications of the continuous uniform distribution are not as abundant as those for other distributions discussed in this chapter, it is appropriate for the novice to begin this introduction to continuous distributions with the uniform distribution.

Uniform Distribution

The density function of the continuous uniform random variable X on the interval $[A, B]$ is

$$f(x; A, B) = \begin{cases} \frac{1}{B-A}, & A \leq x \leq B, \\ 0, & \text{elsewhere.} \end{cases}$$

The density function forms a rectangle with base $B-A$ and **constant height** $\frac{1}{B-A}$. As a result, the uniform distribution is often called the **rectangular distribution**. Note, however, that the interval may not always be closed: $[A, B]$. It can be (A, B) as well. The density function for a uniform random variable on the interval [1, 3] is shown in Figure 6.1.

Probabilities are simple to calculate for the uniform distribution because of the simple nature of the density function. However, note that the application of this distribution is based on the assumption that the probability of falling in an interval of fixed length within $[A, B]$ is constant.

Example 6.1: Suppose that a large conference room at a certain company can be reserved for no more than 4 hours. Both long and short conferences occur quite often. In fact, it can be assumed that the length X of a conference has a uniform distribution on the interval [0, 4].

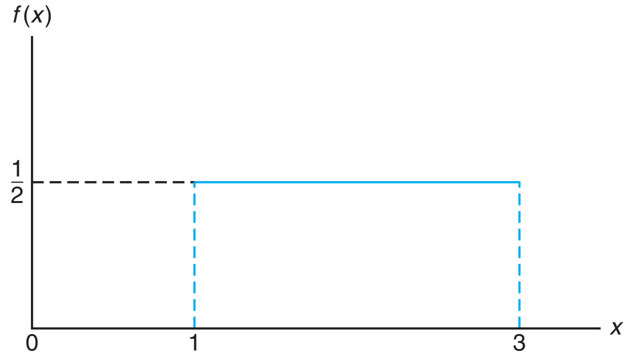

Figure 6.1: The density function for a random variable on the interval $[1, 3]$.

(a) What is the probability density function?

(b) What is the probability that any given conference lasts at least 3 hours?

Solution: (a) The appropriate density function for the uniformly distributed random variable X in this situation is

$$f(x) = \begin{cases} \frac{1}{4}, & 0 \leq x \leq 4, \\ 0, & \text{elsewhere.} \end{cases}$$

(b) $P[X \geq 3] = \int_3^4 \frac{1}{4}\, dx = \frac{1}{4}$.

Theorem 6.1: The mean and variance of the uniform distribution are

$$\mu = \frac{A+B}{2} \text{ and } \sigma^2 = \frac{(B-A)^2}{12}.$$

The proofs of the theorems are left to the reader. See Exercise 6.1 on page 185.

6.2 Normal Distribution

The most important continuous probability distribution in the entire field of statistics is the **normal distribution**. Its graph, called the **normal curve**, is the bell-shaped curve of Figure 6.2, which approximately describes many phenomena that occur in nature, industry, and research. For example, physical measurements in areas such as meteorological experiments, rainfall studies, and measurements of manufactured parts are often more than adequately explained with a normal distribution. In addition, errors in scientific measurements are extremely well approximated by a normal distribution. In 1733, Abraham DeMoivre developed the mathematical equation of the normal curve. It provided a basis from which much of the theory of inductive statistics is founded. The normal distribution is often referred to as the **Gaussian distribution**, in honor of Karl Friedrich Gauss

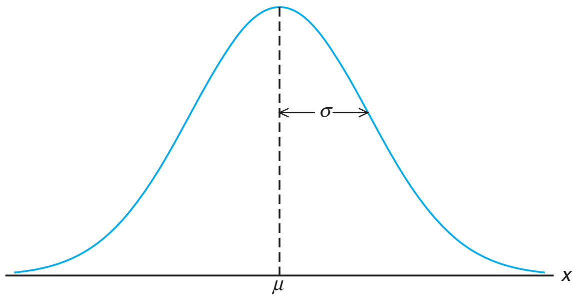

Figure 6.2: The normal curve.

(1777–1855), who also derived its equation from a study of errors in repeated measurements of the same quantity.

A continuous random variable X having the bell-shaped distribution of Figure 6.2 is called a **normal random variable**. The mathematical equation for the probability distribution of the normal variable depends on the two parameters μ and σ, its mean and standard deviation, respectively. Hence, we denote the values of the density of X by $n(x; \mu, \sigma)$.

Normal Distribution

The density of the normal random variable X, with mean μ and variance σ^2, is

$$n(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \quad -\infty < x < \infty,$$

where $\pi = 3.14159\ldots$ and $e = 2.71828\ldots$.

Once μ and σ are specified, the normal curve is completely determined. For example, if $\mu = 50$ and $\sigma = 5$, then the ordinates $n(x; 50, 5)$ can be computed for various values of x and the curve drawn. In Figure 6.3, we have sketched two normal curves having the same standard deviation but different means. The two curves are identical in form but are centered at different positions along the horizontal axis.

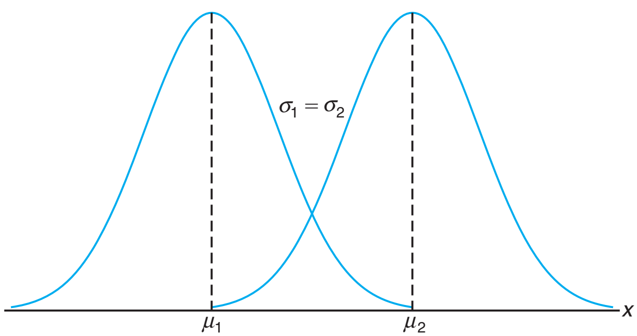

Figure 6.3: Normal curves with $\mu_1 < \mu_2$ and $\sigma_1 = \sigma_2$.

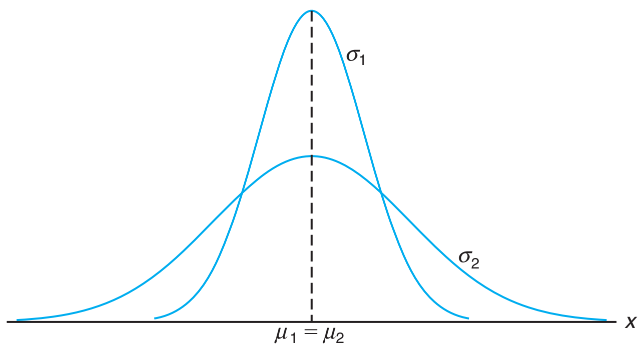

Figure 6.4: Normal curves with $\mu_1 = \mu_2$ and $\sigma_1 < \sigma_2$.

In Figure 6.4, we have sketched two normal curves with the same mean but different standard deviations. This time we see that the two curves are centered at exactly the same position on the horizontal axis, but the curve with the larger standard deviation is lower and spreads out farther. Remember that the area under a probability curve must be equal to 1, and therefore the more variable the set of observations, the lower and wider the corresponding curve will be.

Figure 6.5 shows two normal curves having different means and different standard deviations. Clearly, they are centered at different positions on the horizontal axis and their shapes reflect the two different values of σ.

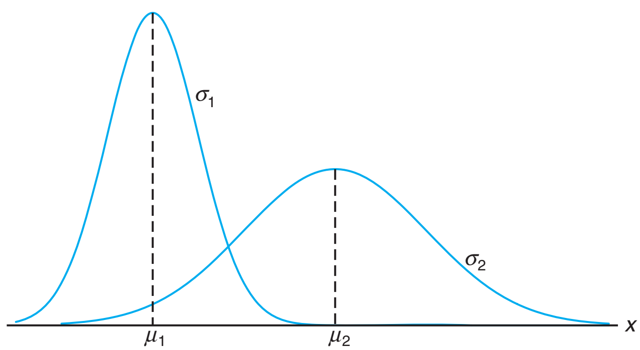

Figure 6.5: Normal curves with $\mu_1 < \mu_2$ and $\sigma_1 < \sigma_2$.

Based on inspection of Figures 6.2 through 6.5 and examination of the first and second derivatives of $n(x; \mu, \sigma)$, we list the following properties of the normal curve:

1. The mode, which is the point on the horizontal axis where the curve is a maximum, occurs at $x = \mu$.
2. The curve is symmetric about a vertical axis through the mean μ.
3. The curve has its points of inflection at $x = \mu \pm \sigma$; it is concave downward if $\mu - \sigma < X < \mu + \sigma$ and is concave upward otherwise.

4. The normal curve approaches the horizontal axis asymptotically as we proceed in either direction away from the mean.
5. The total area under the curve and above the horizontal axis is equal to 1.

Theorem 6.2: The mean and variance of $n(x; \mu, \sigma)$ are μ and σ^2, respectively. Hence, the standard deviation is σ.

***Proof*:** To evaluate the mean, we first calculate

$$E(X - \mu) = \int_{-\infty}^{\infty} \frac{x - \mu}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx.$$

Setting $z = (x - \mu)/\sigma$ and $dx = \sigma\ dz$, we obtain

$$E(X - \mu) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} ze^{-\frac{1}{2}z^2} dz = 0,$$

since the integrand above is an odd function of z. Using Theorem 4.5 on page 128, we conclude that

$$E(X) = \mu.$$

The variance of the normal distribution is given by

$$E[(X - \mu)^2] = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (x - \mu)^2 e^{-\frac{1}{2}[(x-\mu)/\sigma]^2}\ dx.$$

Again setting $z = (x - \mu)/\sigma$ and $dx = \sigma\ dz$, we obtain

$$E[(X - \mu)^2] = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2 e^{-\frac{z^2}{2}}\ dz.$$

Integrating by parts with $u = z$ and $dv = ze^{-z^2/2}\ dz$ so that $du = dz$ and $v = -e^{-z^2/2}$, we find that

$$E[(X - \mu)^2] = \frac{\sigma^2}{\sqrt{2\pi}} \left(-ze^{-z^2/2}\Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-z^2/2}\ dz \right) = \sigma^2(0 + 1) = \sigma^2.$$

Many random variables have probability distributions that can be described adequately by the normal curve once μ and σ^2 are specified. In this chapter, we shall assume that these two parameters are known, perhaps from previous investigations. Later, we shall make statistical inferences when μ and σ^2 are unknown and have been estimated from the available experimental data.

We pointed out earlier the role that the normal distribution plays as a reasonable approximation of scientific variables in real-life experiments. There are other applications of the normal distribution that the reader will appreciate as he or she moves on in the book. The normal distribution finds enormous application as a *limiting distribution*. Under certain conditions, the normal distribution provides a good continuous approximation to the binomial and hypergeometric distributions. The case of the approximation to the binomial is covered in Section 6.5. In Chapter 8, the reader will learn about **sampling distributions**. It turns out that the limiting distribution of sample averages is normal. This provides a broad base for statistical inference that proves very valuable to the data analyst interested in estimation and hypothesis testing. Theory in the important areas such as analysis of variance (Chapters 13, 14, and 15) and quality control (Chapter 17) is based on assumptions that make use of the normal distribution.

In Section 6.3, examples demonstrate the use of tables of the normal distribution. Section 6.4 follows with examples of applications of the normal distribution.

6.3 Areas under the Normal Curve

The curve of any continuous probability distribution or density function is constructed so that the area under the curve bounded by the two ordinates $x = x_1$ and $x = x_2$ equals the probability that the random variable X assumes a value between $x = x_1$ and $x = x_2$. Thus, for the normal curve in Figure 6.6,

$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} n(x; \mu, \sigma)\, dx = \frac{1}{\sqrt{2\pi}\sigma} \int_{x_1}^{x_2} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx$$

is represented by the area of the shaded region.

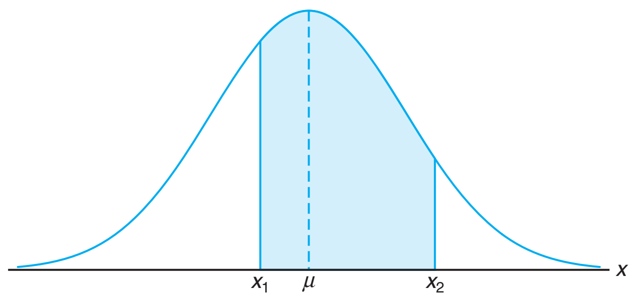

Figure 6.6: $P(x_1 < X < x_2)$ = area of the shaded region.

In Figures 6.3, 6.4, and 6.5 we saw how the normal curve is dependent on the mean and the standard deviation of the distribution under investigation. The area under the curve between any two ordinates must then also depend on the values μ and σ. This is evident in Figure 6.7, where we have shaded regions corresponding to $P(x_1 < X < x_2)$ for two curves with different means and variances. $P(x_1 < X < x_2)$, where X is the random variable describing distribution A, is indicated by the shaded area below the curve of A. If X is the random variable describing distribution B, then $P(x_1 < X < x_2)$ is given by the entire shaded region. Obviously, the two shaded regions are different in size; therefore, the probability associated with each distribution will be different for the two given values of X.

There are many types of statistical software that can be used in calculating areas under the normal curve. The difficulty encountered in solving integrals of normal density functions necessitates the tabulation of normal curve areas for quick reference. However, it would be a hopeless task to attempt to set up separate tables for every conceivable value of μ and σ. Fortunately, we are able to transform all the observations of any normal random variable X into a new set of observations

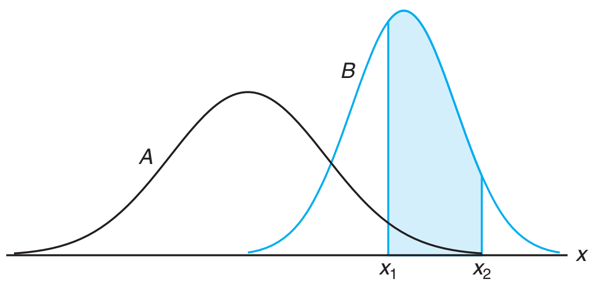

Figure 6.7: $P(x_1 < X < x_2)$ for different normal curves.

of a normal random variable Z with mean 0 and variance 1. This can be done by means of the transformation

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

Whenever X assumes a value x, the corresponding value of Z is given by $z = (x - \mu)/\sigma$. Therefore, if X falls between the values $x = x_1$ and $x = x_2$, the random variable Z will fall between the corresponding values $z_1 = (x_1 - \mu)/\sigma$ and $z_2 = (x_2 - \mu)/\sigma$. Consequently, we may write

$$\begin{aligned} P(x_1 < X < x_2) &= \frac{1}{\sqrt{2\pi}\sigma} \int_{x_1}^{x_2} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx = \frac{1}{\sqrt{2\pi}} \int_{z_1}^{z_2} e^{-\frac{1}{2}z^2}\, dz \\ &= \int_{z_1}^{z_2} n(z; 0, 1)\, dz = P(z_1 < Z < z_2), \end{aligned}$$

where Z is seen to be a normal random variable with mean 0 and variance 1.

Definition 6.1: The distribution of a normal random variable with mean 0 and variance 1 is called a **standard normal distribution**.

The original and transformed distributions are illustrated in Figure 6.8. Since all the values of X falling between x_1 and x_2 have corresponding z values between z_1 and z_2, the area under the X-curve between the ordinates $x = x_1$ and $x = x_2$ in Figure 6.8 equals the area under the Z-curve between the transformed ordinates $z = z_1$ and $z = z_2$.

We have now reduced the required number of tables of normal-curve areas to one, that of the standard normal distribution. Table A.3 indicates the area under the standard normal curve corresponding to $P(Z < z)$ for values of z ranging from -3.49 to 3.49. To illustrate the use of this table, let us find the probability that Z is less than 1.74. First, we locate a value of z equal to 1.7 in the left column; then we move across the row to the column under 0.04, where we read 0.9591. Therefore, $P(Z < 1.74) = 0.9591$. To find a z value corresponding to a given probability, the process is reversed. For example, the z value leaving an area of 0.2148 under the curve to the left of z is seen to be -0.79.

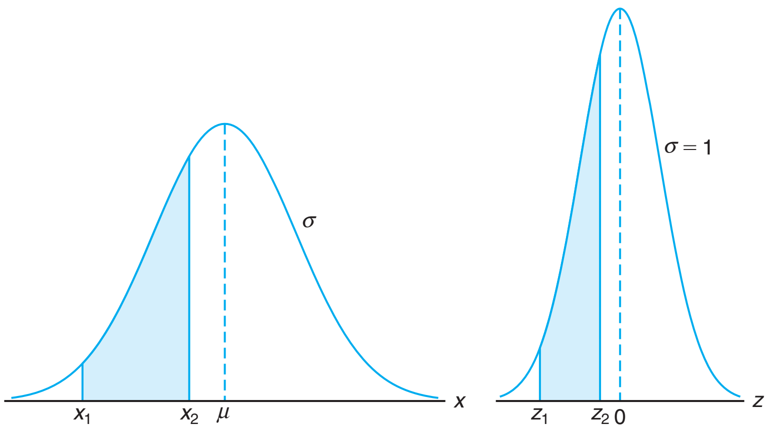

Figure 6.8: The original and transformed normal distributions.

Example 6.2: Given a standard normal distribution, find the area under the curve that lies

(a) to the right of $z = 1.84$ and

(b) between $z = -1.97$ and $z = 0.86$.

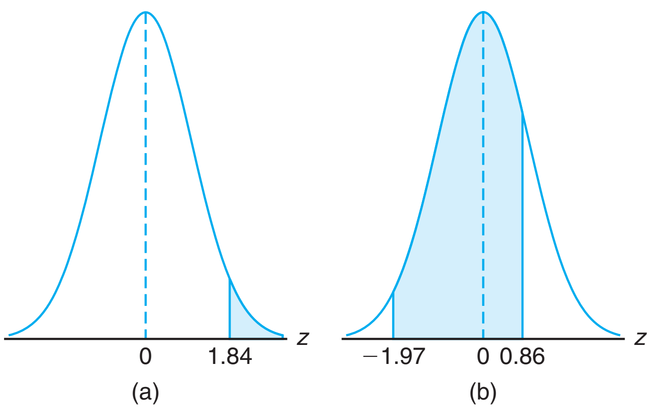

Figure 6.9: Areas for Example 6.2.

***Solution*:** See Figure 6.9 for the specific areas.

(a) The area in Figure 6.9(a) to the right of $z = 1.84$ is equal to 1 minus the area in Table A.3 to the left of $z = 1.84$, namely, $1 - 0.9671 = 0.0329$.

(b) The area in Figure 6.9(b) between $z = -1.97$ and $z = 0.86$ is equal to the area to the left of $z = 0.86$ minus the area to the left of $z = -1.97$. From Table A.3 we find the desired area to be $0.8051 - 0.0244 = 0.7807$.

Example 6.3: Given a standard normal distribution, find the value of k such that

(a) $P(Z > k) = 0.3015$ and

(b) $P(k < Z < -0.18) = 0.4197$.

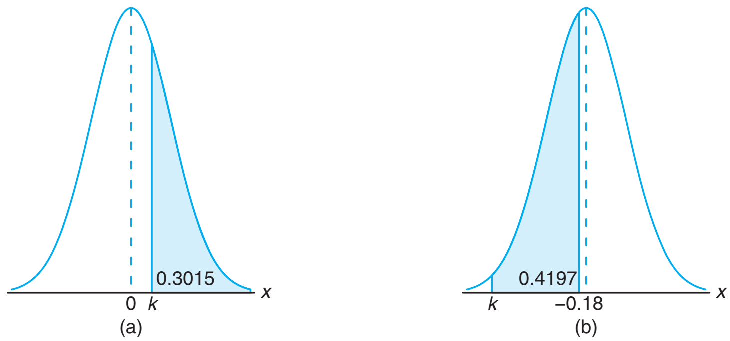

Figure 6.10: Areas for Example 6.3.

***Solution*:** Distributions and the desired areas are shown in Figure 6.10.

(a) In Figure 6.10(a), we see that the k value leaving an area of 0.3015 to the right must then leave an area of 0.6985 to the left. From Table A.3 it follows that $k = 0.52$.

(b) From Table A.3 we note that the total area to the left of -0.18 is equal to 0.4286. In Figure 6.10(b), we see that the area between k and -0.18 is 0.4197, so the area to the left of k must be $0.4286 - 0.4197 = 0.0089$. Hence, from Table A.3, we have $k = -2.37$.

Example 6.4: Given a random variable X having a normal distribution with $\mu = 50$ and $\sigma = 10$, find the probability that X assumes a value between 45 and 62.

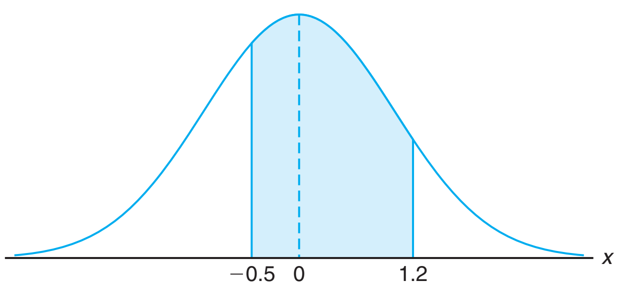

Figure 6.11: Area for Example 6.4.

***Solution*:** The z values corresponding to $x_1 = 45$ and $x_2 = 62$ are

$$z_1 = \frac{45 - 50}{10} = -0.5 \text{ and } z_2 = \frac{62 - 50}{10} = 1.2.$$

Therefore,

$$P(45 < X < 62) = P(-0.5 < Z < 1.2).$$

$P(-0.5 < Z < 1.2)$ is shown by the area of the shaded region in Figure 6.11. This area may be found by subtracting the area to the left of the ordinate $z = -0.5$ from the entire area to the left of $z = 1.2$. Using Table A.3, we have

$$\begin{aligned} P(45 < X < 62) &= P(-0.5 < Z < 1.2) = P(Z < 1.2) - P(Z < -0.5) \\ &= 0.8849 - 0.3085 = 0.5764. \end{aligned}$$

Example 6.5: Given that X has a normal distribution with $\mu = 300$ and $\sigma = 50$, find the probability that X assumes a value greater than 362.

***Solution*:** The normal probability distribution with the desired area shaded is shown in Figure 6.12. To find $P(X > 362)$, we need to evaluate the area under the normal curve to the right of $x = 362$. This can be done by transforming $x = 362$ to the corresponding z value, obtaining the area to the left of z from Table A.3, and then subtracting this area from 1. We find that

$$z = \frac{362 - 300}{50} = 1.24.$$

Hence,

$$P(X > 362) = P(Z > 1.24) = 1 - P(Z < 1.24) = 1 - 0.8925 = 0.1075.$$

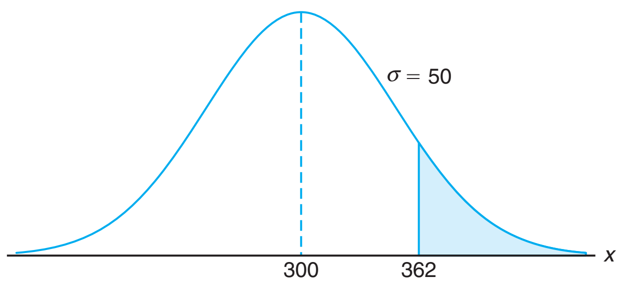

Figure 6.12: Area for Example 6.5.

According to Chebyshev's theorem on page 137, the probability that a random variable assumes a value within 2 standard deviations of the mean is at least 3/4. If the random variable has a normal distribution, the z values corresponding to $x_1 = \mu - 2\sigma$ and $x_2 = \mu + 2\sigma$ are easily computed to be

$$z_1 = \frac{(\mu - 2\sigma) - \mu}{\sigma} = -2 \text{ and } z_2 = \frac{(\mu + 2\sigma) - \mu}{\sigma} = 2.$$

Hence,

$$\begin{aligned} P(\mu - 2\sigma < X < \mu + 2\sigma) &= P(-2 < Z < 2) = P(Z < 2) - P(Z < -2) \\ &= 0.9772 - 0.0228 = 0.9544, \end{aligned}$$

which is a much stronger statement than that given by Chebyshev's theorem.

Using the Normal Curve in Reverse

Sometimes, we are required to find the value of z corresponding to a specified probability that falls between values listed in Table A.3 (see Example 6.6). For convenience, we shall always choose the z value corresponding to the tabular probability that comes closest to the specified probability.

The preceding two examples were solved by going first from a value of x to a z value and then computing the desired area. In Example 6.6, we reverse the process and begin with a known area or probability, find the z value, and then determine x by rearranging the formula

$$z = \frac{x - \mu}{\sigma} \quad \text{to give} \quad x = \sigma z + \mu.$$

Example 6.6: Given a normal distribution with $\mu = 40$ and $\sigma = 6$, find the value of x that has

(a) 45% of the area to the left and

(b) 14% of the area to the right.

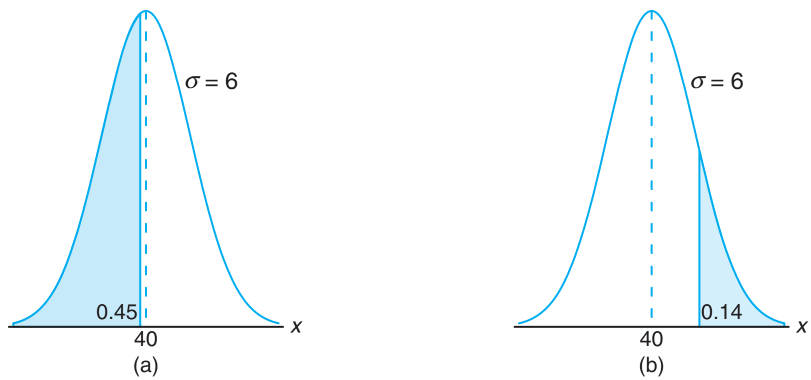

Figure 6.13: Areas for Example 6.6.

Solution: (a) An area of 0.45 to the left of the desired x value is shaded in Figure 6.13(a). We require a z value that leaves an area of 0.45 to the left. From Table A.3 we find $P(Z < -0.13) = 0.45$, so the desired z value is -0.13. Hence,

$$x = (6)(-0.13) + 40 = 39.22.$$

(b) In Figure 6.13(b), we shade an area equal to 0.14 to the right of the desired x value. This time we require a z value that leaves 0.14 of the area to the right and hence an area of 0.86 to the left. Again, from Table A.3, we find $P(Z < 1.08) = 0.86$, so the desired z value is 1.08 and

$$x = (6)(1.08) + 40 = 46.48.$$

6.4 Applications of the Normal Distribution

Some of the many problems for which the normal distribution is applicable are treated in the following examples. The use of the normal curve to approximate binomial probabilities is considered in Section 6.5.

Example 6.7: A certain type of storage battery lasts, on average, 3.0 years with a standard deviation of 0.5 year. Assuming that battery life is normally distributed, find the probability that a given battery will last less than 2.3 years.

***Solution*:** First construct a diagram such as Figure 6.14, showing the given distribution of battery lives and the desired area. To find $P(X < 2.3)$, we need to evaluate the area under the normal curve to the left of 2.3. This is accomplished by finding the area to the left of the corresponding z value. Hence, we find that

$$z = \frac{2.3 - 3}{0.5} = -1.4,$$

and then, using Table A.3, we have

$$P(X < 2.3) = P(Z < -1.4) = 0.0808.$$

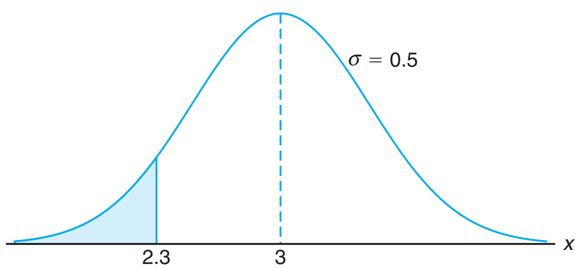

Figure 6.14: Area for Example 6.7.

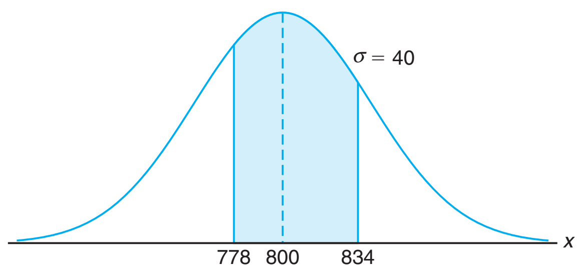

Figure 6.15: Area for Example 6.8.

Example 6.8: An electrical firm manufactures light bulbs that have a life, before burn-out, that is normally distributed with mean equal to 800 hours and a standard deviation of 40 hours. Find the probability that a bulb burns between 778 and 834 hours.

***Solution*:** The distribution of light bulb life is illustrated in Figure 6.15. The z values corresponding to $x_1 = 778$ and $x_2 = 834$ are

$$z_1 = \frac{778 - 800}{40} = -0.55 \text{ and } z_2 = \frac{834 - 800}{40} = 0.85.$$

Hence,

$$\begin{aligned} P(778 < X < 834) &= P(-0.55 < Z < 0.85) = P(Z < 0.85) - P(Z < -0.55) \\ &= 0.8023 - 0.2912 = 0.5111. \end{aligned}$$

Example 6.9: In an industrial process, the diameter of a ball bearing is an important measurement. The buyer sets specifications for the diameter to be 3.0 ± 0.01 cm. The implication is that no part falling outside these specifications will be accepted. It is known that in the process the diameter of a ball bearing has a normal distribution with mean $\mu = 3.0$ and standard deviation $\sigma = 0.005$. On average, how many manufactured ball bearings will be scrapped?

***Solution*:** The distribution of diameters is illustrated by Figure 6.16. The values corresponding to the specification limits are $x_1 = 2.99$ and $x_2 = 3.01$. The corresponding z values are

$$z_1 = \frac{2.99 - 3.0}{0.005} = -2.0 \text{ and } z_2 = \frac{3.01 - 3.0}{0.005} = +2.0.$$

Hence,

$$P(2.99 < X < 3.01) = P(-2.0 < Z < 2.0).$$

From Table A.3, $P(Z < -2.0) = 0.0228$. Due to symmetry of the normal distribution, we find that

$$P(Z < -2.0) + P(Z > 2.0) = 2(0.0228) = 0.0456.$$

As a result, it is anticipated that, on average, 4.56% of manufactured ball bearings will be scrapped.

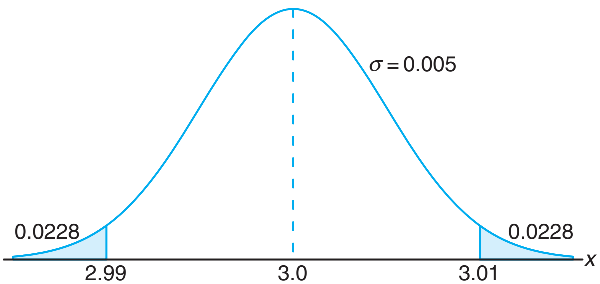

Figure 6.16: Area for Example 6.9.

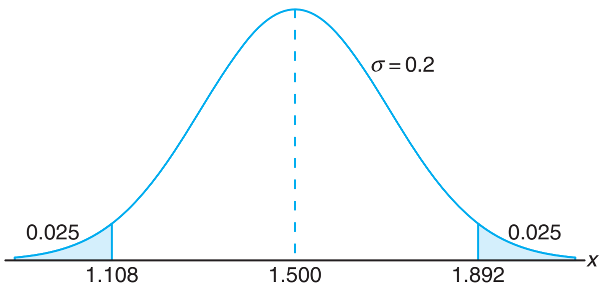

Figure 6.17: Specifications for Example 6.10.

Example 6.10: Gauges are used to reject all components for which a certain dimension is not within the specification $1.50 \pm d$. It is known that this measurement is normally distributed with mean 1.50 and standard deviation 0.2. Determine the value d such that the specifications "cover" 95% of the measurements.

***Solution*:** From Table A.3 we know that

$$P(-1.96 < Z < 1.96) = 0.95.$$

Therefore,

$$1.96 = \frac{(1.50 + d) - 1.50}{0.2},$$

from which we obtain

$$d = (0.2)(1.96) = 0.392.$$

An illustration of the specifications is shown in Figure 6.17.

Example 6.11: A certain machine makes electrical resistors having a mean resistance of 40 ohms and a standard deviation of 2 ohms. Assuming that the resistance follows a normal distribution and can be measured to any degree of accuracy, what percentage of resistors will have a resistance exceeding 43 ohms?

***Solution*:** A percentage is found by multiplying the relative frequency by 100%. Since the relative frequency for an interval is equal to the probability of a value falling in the interval, we must find the area to the right of $x = 43$ in Figure 6.18. This can be done by transforming $x = 43$ to the corresponding z value, obtaining the area to the left of z from Table A.3, and then subtracting this area from 1. We find

$$z = \frac{43 - 40}{2} = 1.5.$$

Therefore,

$$P(X > 43) = P(Z > 1.5) = 1 - P(Z < 1.5) = 1 - 0.9332 = 0.0668.$$

Hence, 6.68% of the resistors will have a resistance exceeding 43 ohms.

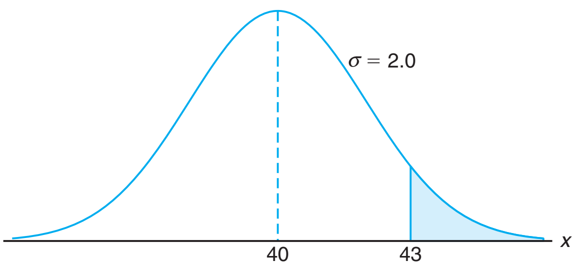

Figure 6.18: Area for Example 6.11.

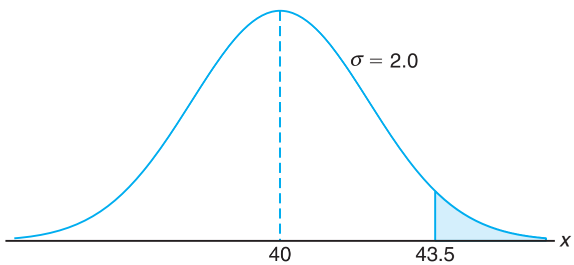

Figure 6.19: Area for Example 6.12.

Example 6.12: Find the percentage of resistances exceeding 43 ohms for Example 6.11 if resistance is measured to the nearest ohm.

***Solution*:** This problem differs from that in Example 6.11 in that we now assign a measurement of 43 ohms to all resistors whose resistances are greater than 42.5 and less than 43.5. We are actually approximating a discrete distribution by means of a continuous normal distribution. The required area is the region shaded to the right of 43.5 in Figure 6.19. We now find that

$$z = \frac{43.5 - 40}{2} = 1.75.$$

Hence,

$$P(X > 43.5) = P(Z > 1.75) = 1 - P(Z < 1.75) = 1 - 0.9599 = 0.0401.$$

Therefore, 4.01% of the resistances exceed 43 ohms when measured to the nearest ohm. The difference 6.68% − 4.01% = 2.67% between this answer and that of Example 6.11 represents all those resistance values greater than 43 and less than 43.5 that are now being recorded as 43 ohms.

Example 6.13: The average grade for an exam is 74, and the standard deviation is 7. If 12% of the class is given As, and the grades are curved to follow a normal distribution, what is the lowest possible A and the highest possible B?

***Solution*:** In this example, we begin with a known area of probability, find the z value, and then determine x from the formula $x = \sigma z + \mu$. An area of 0.12, corresponding to the fraction of students receiving As, is shaded in Figure 6.20. We require a z value that leaves 0.12 of the area to the right and, hence, an area of 0.88 to the left. From Table A.3, $P(Z < 1.18)$ has the closest value to 0.88, so the desired z value is 1.18. Hence,

$$x = (7)(1.18) + 74 = 82.26.$$

Therefore, the lowest A is 83 and the highest B is 82.

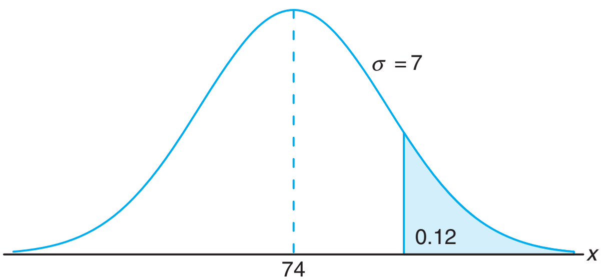

Figure 6.20: Area for Example 6.13.

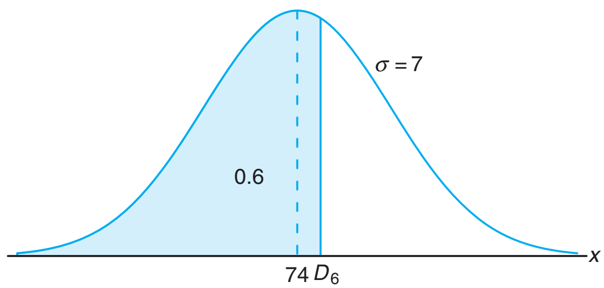

Figure 6.21: Area for Example 6.14.

Example 6.14: Refer to Example 6.13 and find the sixth decile.

***Solution*:** The sixth decile, written D_6, is the x value that leaves 60% of the area to the left, as shown in Figure 6.21. From Table A.3 we find $P(Z < 0.25) \approx 0.6$, so the desired z value is 0.25. Now $x = (7)(0.25) + 74 = 75.75$. Hence, $D_6 = 75.75$. That is, 60% of the grades are 75 or less.

Exercises

6.1 Given a continuous uniform distribution, show that

(a) $\mu = \frac{A+B}{2}$ and

(b) $\sigma^2 = \frac{(B-A)^2}{12}$.

6.2 Suppose X follows a continuous uniform distribution from 1 to 5. Determine the conditional probability $P(X > 2.5 \mid X \leq 4)$.

6.3 The daily amount of coffee, in liters, dispensed by a machine located in an airport lobby is a random variable X having a continuous uniform distribution with $A = 7$ and $B = 10$. Find the probability that on a given day the amount of coffee dispensed by this machine will be

(a) at most 8.8 liters;

(b) more than 7.4 liters but less than 9.5 liters;

(c) at least 8.5 liters.

6.4 A bus arrives every 10 minutes at a bus stop. It is assumed that the waiting time for a particular individual is a random variable with a continuous uniform distribution.

(a) What is the probability that the individual waits more than 7 minutes?

(b) What is the probability that the individual waits between 2 and 7 minutes?

6.5 Given a standard normal distribution, find the area under the curve that lies

(a) to the left of $z = -1.39$;

(b) to the right of $z = 1.96$;

(c) between $z = -2.16$ and $z = -0.65$;

(d) to the left of $z = 1.43$;

(e) to the right of $z = -0.89$;

(f) between $z = -0.48$ and $z = 1.74$.

6.6 Find the value of z if the area under a standard normal curve

(a) to the right of z is 0.3622;

(b) to the left of z is 0.1131;

(c) between 0 and z, with $z > 0$, is 0.4838;

(d) between $-z$ and z, with $z > 0$, is 0.9500.

6.7 Given a standard normal distribution, find the value of k such that

(a) $P(Z > k) = 0.2946$;

(b) $P(Z < k) = 0.0427$;

(c) $P(-0.93 < Z < k) = 0.7235$.

6.8 Given a normal distribution with $\mu = 30$ and $\sigma = 6$, find

(a) the normal curve area to the right of $x = 17$;

(b) the normal curve area to the left of $x = 22$;

(c) the normal curve area between $x = 32$ and $x = 41$;

(d) the value of x that has 80% of the normal curve area to the left;

(e) the two values of x that contain the middle 75% of the normal curve area.

6.9 Given the normally distributed variable X with mean 18 and standard deviation 2.5, find

(a) $P(X < 15)$;

(b) the value of k such that $P(X < k) = 0.2236$;

(c) the value of k such that $P(X > k) = 0.1814$;

(d) $P(17 < X < 21)$.

6.10 According to Chebyshev's theorem, the probability that any random variable assumes a value within 3 standard deviations of the mean is at least 8/9. If it is known that the probability distribution of a random variable X is normal with mean μ and variance σ^2, what is the exact value of $P(\mu - 3\sigma < X < \mu + 3\sigma)$?

6.11 A soft-drink machine is regulated so that it discharges an average of 200 milliliters per cup. If the amount of drink is normally distributed with a standard deviation equal to 15 milliliters,

(a) what fraction of the cups will contain more than 224 milliliters?

(b) what is the probability that a cup contains between 191 and 209 milliliters?

(c) how many cups will probably overflow if 230-milliliter cups are used for the next 1000 drinks?

(d) below what value do we get the smallest 25% of the drinks?

6.12 The loaves of rye bread distributed to local stores by a certain bakery have an average length of 30 centimeters and a standard deviation of 2 centimeters. Assuming that the lengths are normally distributed, what percentage of the loaves are

(a) longer than 31.7 centimeters?

(b) between 29.3 and 33.5 centimeters in length?

(c) shorter than 25.5 centimeters?

6.13 A research scientist reports that mice will live an average of 40 months when their diets are sharply restricted and then enriched with vitamins and proteins. Assuming that the lifetimes of such mice are normally distributed with a standard deviation of 6.3 months, find the probability that a given mouse will live

(a) more than 32 months;

(b) less than 28 months;

(c) between 37 and 49 months.

6.14 The finished inside diameter of a piston ring is normally distributed with a mean of 10 centimeters and a standard deviation of 0.03 centimeter.

(a) What proportion of rings will have inside diameters exceeding 10.075 centimeters?

(b) What is the probability that a piston ring will have an inside diameter between 9.97 and 10.03 centimeters?

(c) Below what value of inside diameter will 15% of the piston rings fall?

6.15 A lawyer commutes daily from his suburban home to his midtown office. The average time for a one-way trip is 24 minutes, with a standard deviation of 3.8 minutes. Assume the distribution of trip times to be normally distributed.

(a) What is the probability that a trip will take at least 1/2 hour?

(b) If the office opens at 9:00 A.M. and the lawyer leaves his house at 8:45 A.M. daily, what percentage of the time is he late for work?

(c) If he leaves the house at 8:35 A.M. and coffee is served at the office from 8:50 A.M. until 9:00 A.M., what is the probability that he misses coffee?

(d) Find the length of time above which we find the slowest 15% of the trips.

(e) Find the probability that 2 of the next 3 trips will take at least 1/2 hour.

6.16 In the November 1990 issue of *Chemical Engineering Progress*, a study discussed the percent purity of oxygen from a certain supplier. Assume that the mean was 99.61 with a standard deviation of 0.08. Assume that the distribution of percent purity was approximately normal.

(a) What percentage of the purity values would you expect to be between 99.5 and 99.7?

(b) What purity value would you expect to exceed exactly 5% of the population?

6.17 The average life of a certain type of small motor is 10 years with a standard deviation of 2 years. The manufacturer replaces free all motors that fail while under guarantee. If she is willing to replace only 3% of the motors that fail, how long a guarantee should be offered? Assume that the lifetime of a motor follows a normal distribution.

6.18 The heights of 1000 students are normally distributed with a mean of 174.5 centimeters and a standard deviation of 6.9 centimeters. Assuming that the heights are recorded to the nearest half-centimeter, how many of these students would you expect to have heights

(a) less than 160.0 centimeters?

(b) between 171.5 and 182.0 centimeters inclusive?

(c) equal to 175.0 centimeters?

(d) greater than or equal to 188.0 centimeters?

6.19 A company pays its employees an average wage of \$15.90 an hour with a standard deviation of \$1.50. If the wages are approximately normally distributed and paid to the nearest cent,

(a) what percentage of the workers receive wages between \$13.75 and \$16.22 an hour inclusive?

(b) the highest 5% of the employee hourly wages is greater than what amount?

6.20 The weights of a large number of miniature poodles are approximately normally distributed with a mean of 8 kilograms and a standard deviation of 0.9 kilogram. If measurements are recorded to the nearest tenth of a kilogram, find the fraction of these poodles with weights

(a) over 9.5 kilograms;

(b) of at most 8.6 kilograms;

(c) between 7.3 and 9.1 kilograms inclusive.

6.21 The tensile strength of a certain metal component is normally distributed with a mean of 10,000 kilograms per square centimeter and a standard deviation of 100 kilograms per square centimeter. Measurements are recorded to the nearest 50 kilograms per square centimeter.

(a) What proportion of these components exceed 10,150 kilograms per square centimeter in tensile strength?

(b) If specifications require that all components have tensile strength between 9800 and 10,200 kilograms per square centimeter inclusive, what proportion of pieces would we expect to scrap?

6.22 If a set of observations is normally distributed, what percent of these differ from the mean by

(a) more than 1.3σ?

(b) less than 0.52σ?

6.23 The IQs of 600 applicants to a certain college are approximately normally distributed with a mean of 115 and a standard deviation of 12. If the college requires an IQ of at least 95, how many of these students will be rejected on this basis of IQ, regardless of their other qualifications? Note that IQs are recorded to the nearest integers.

6.5 Normal Approximation to the Binomial

Probabilities associated with binomial experiments are readily obtainable from the formula $b(x; n, p)$ of the binomial distribution or from Table A.1 when n is small. In addition, binomial probabilities are readily available in many computer software packages. However, it is instructive to learn the relationship between the binomial and the normal distribution. In Section 5.5, we illustrated how the Poisson distribution can be used to approximate binomial probabilities when n is quite large and p is very close to 0 or 1. Both the binomial and the Poisson distributions are discrete. The first application of a continuous probability distribution to approximate probabilities over a discrete sample space was demonstrated in Example 6.12, where the normal curve was used. The normal distribution is often a good approximation to a discrete distribution when the latter takes on a symmetric bell shape. From a theoretical point of view, some distributions converge to the normal as their parameters approach certain limits. The normal distribution is a convenient approximating distribution because the cumulative distribution function is so easily tabled. The binomial distribution is nicely approximated by the normal in practical problems when one works with the cumulative distribution function. We now state a theorem that allows us to use areas under the normal curve to approximate binomial properties when n is sufficiently large.

Theorem 6.3: If X is a binomial random variable with mean $\mu = np$ and variance $\sigma^2 = npq$, then the limiting form of the distribution of

$$Z = \frac{X - np}{\sqrt{npq}},$$

as $n \to \infty$, is the standard normal distribution $n(z; 0, 1)$.

It turns out that the normal distribution with $\mu = np$ and $\sigma^2 = np(1-p)$ not only provides a very accurate approximation to the binomial distribution when n is large and p is not extremely close to 0 or 1 but also provides a fairly good approximation even when n is small and p is reasonably close to 1/2.

To illustrate the normal approximation to the binomial distribution, we first draw the histogram for $b(x; 15, 0.4)$ and then superimpose the particular normal curve having the same mean and variance as the binomial variable X. Hence, we draw a normal curve with

$$\mu = np = (15)(0.4) = 6 \text{ and } \sigma^2 = npq = (15)(0.4)(0.6) = 3.6.$$

The histogram of $b(x; 15, 0.4)$ and the corresponding superimposed normal curve, which is completely determined by its mean and variance, are illustrated in Figure 6.22.

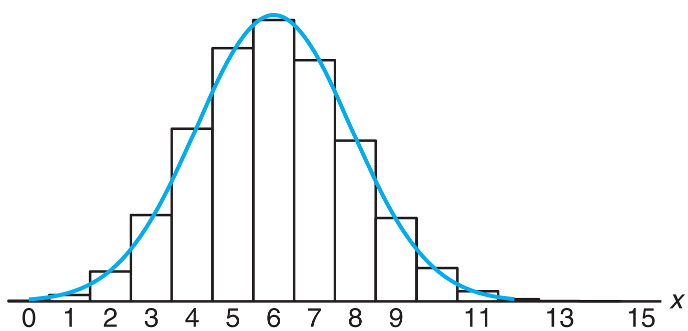

Figure 6.22: Normal approximation of $b(x; 15, 0.4)$.

The exact probability that the binomial random variable X assumes a given value x is equal to the area of the bar whose base is centered at x. For example, the exact probability that X assumes the value 4 is equal to the area of the rectangle with base centered at $x = 4$. Using Table A.1, we find this area to be

$$P(X = 4) = b(4; 15, 0.4) = 0.1268,$$

which is approximately equal to the area of the shaded region under the normal curve between the two ordinates $x_1 = 3.5$ and $x_2 = 4.5$ in Figure 6.23. Converting to z values, we have

$$z_1 = \frac{3.5 - 6}{1.897} = -1.32 \qquad \text{and} \qquad z_2 = \frac{4.5 - 6}{1.897} = -0.79.$$

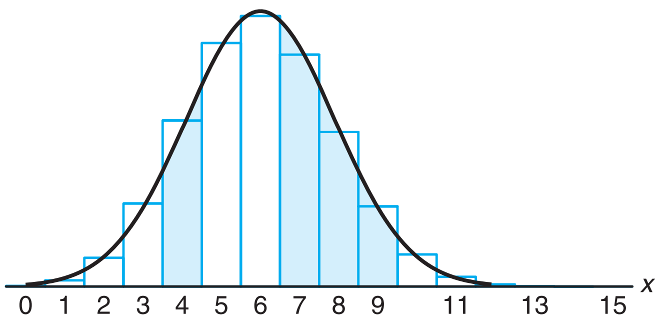

Figure 6.23: Normal approximation of $b(x; 15, 0.4)$ and $\sum_{x=7}^{9} b(x; 15, 0.4)$.

If X is a binomial random variable and Z a standard normal variable, then

$$\begin{aligned} P(X = 4) = b(4; 15, 0.4) &\approx P(-1.32 < Z < -0.79) \\ &= P(Z < -0.79) - P(Z < -1.32) = 0.2148 - 0.0934 = 0.1214. \end{aligned}$$

This agrees very closely with the exact value of 0.1268.

The normal approximation is most useful in calculating binomial sums for large values of n. Referring to Figure 6.23, we might be interested in the probability that X assumes a value from 7 to 9 inclusive. The exact probability is given by

$$\begin{aligned} P(7 \le X \le 9) &= \sum_{x=0}^{9} b(x; 15, 0.4) - \sum_{x=0}^{6} b(x; 15, 0.4) \\ &= 0.9662 - 0.6098 = 0.3564, \end{aligned}$$

which is equal to the sum of the areas of the rectangles with bases centered at $x = 7$, 8, and 9. For the normal approximation, we find the area of the shaded region under the curve between the ordinates $x_1 = 6.5$ and $x_2 = 9.5$ in Figure 6.23. The corresponding z values are

$$z_1 = \frac{6.5 - 6}{1.897} = 0.26 \quad \text{and} \quad z_2 = \frac{9.5 - 6}{1.897} = 1.85.$$

Now,

$$P(7 \leq X \leq 9) \approx P(0.26 < Z < 1.85) = P(Z < 1.85) - P(Z < 0.26)$$
$$= 0.9678 - 0.6026 = 0.3652.$$

Once again, the normal curve approximation provides a value that agrees very closely with the exact value of 0.3564. The degree of accuracy, which depends on how well the curve fits the histogram, will increase as n increases. This is particularly true when p is not very close to $1/2$ and the histogram is no longer symmetric. Figures 6.24 and 6.25 show the histograms for $b(x; 6, 0.2)$ and $b(x; 15, 0.2)$, respectively. It is evident that a normal curve would fit the histogram considerably better when $n = 15$ than when $n = 6$.

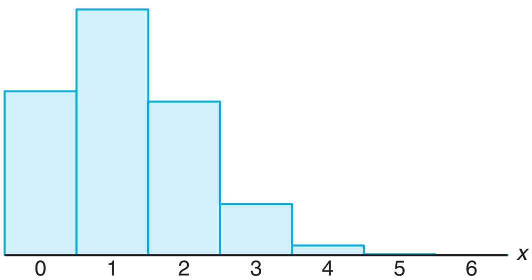

Figure 6.24: Histogram for $b(x; 6, 0.2)$.

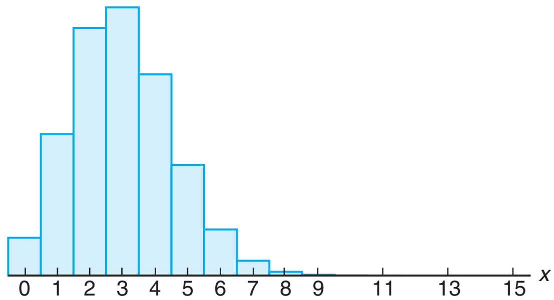

Figure 6.25: Histogram for $b(x; 15, 0.2)$.

In our illustrations of the normal approximation to the binomial, it became apparent that if we seek the area under the normal curve to the left of, say, x, it is more accurate to use $x + 0.5$. This is a correction to accommodate the fact that a discrete distribution is being approximated by a continuous distribution. The correction **+0.5** is called a **continuity correction**. The foregoing discussion leads to the following formal normal approximation to the binomial.

Normal Approximation to the Binomial Distribution

Let X be a binomial random variable with parameters n and p. For large n, X has approximately a normal distribution with $\mu = np$ and $\sigma^2 = npq = np(1-p)$ and

$$\begin{aligned} P(X \leq x) &= \sum_{k=0}^{x} b(k; n, p) \\ &\approx \text{area under normal curve to the left of } x + 0.5 \\ &= P\left(Z \leq \frac{x + 0.5 - np}{\sqrt{npq}}\right), \end{aligned}$$

and the approximation will be good if np and $n(1-p)$ are greater than or equal to 5.

As we indicated earlier, the quality of the approximation is quite good for large n. If p is close to $1/2$, a moderate or small sample size will be sufficient for a reasonable approximation. We offer Table 6.1 as an indication of the quality of the approximation. Both the normal approximation and the true binomial cumulative probabilities are given. Notice that at $p = 0.05$ and $p = 0.10$, the approximation is fairly crude for $n = 10$. However, even for $n = 10$, note the improvement for $p = 0.50$. On the other hand, when p is fixed at $p = 0.05$, note the improvement of the approximation as we go from $n = 20$ to $n = 100$.

Table 6.1: Normal Approximation and True Cumulative Binomial Probabilities

	$p = 0.05,\ n = 10$		$p = 0.10,\ n = 10$		$p = 0.50,\ n = 10$	
r	**Binomial**	**Normal**	**Binomial**	**Normal**	**Binomial**	**Normal**
0	0.5987	0.5000	0.3487	0.2981	0.0010	0.0022
1	0.9139	0.9265	0.7361	0.7019	0.0107	0.0136
2	0.9885	0.9981	0.9298	0.9429	0.0547	0.0571
3	0.9990	1.0000	0.9872	0.9959	0.1719	0.1711
4	1.0000	1.0000	0.9984	0.9999	0.3770	0.3745
5			1.0000	1.0000	0.6230	0.6255
6					0.8281	0.8289
7					0.9453	0.9429
8					0.9893	0.9864
9					0.9990	0.9978
10					1.0000	0.9997
	$p = 0.05$					
	$n = 20$		$n = 50$		$n = 100$	
r	**Binomial**	**Normal**	**Binomial**	**Normal**	**Binomial**	**Normal**
0	0.3585	0.3015	0.0769	0.0968	0.0059	0.0197
1	0.7358	0.6985	0.2794	0.2578	0.0371	0.0537
2	0.9245	0.9382	0.5405	0.5000	0.1183	0.1251
3	0.9841	0.9948	0.7604	0.7422	0.2578	0.2451
4	0.9974	0.9998	0.8964	0.9032	0.4360	0.4090
5	0.9997	1.0000	0.9622	0.9744	0.6160	0.5910
6	1.0000	1.0000	0.9882	0.9953	0.7660	0.7549
7			0.9968	0.9994	0.8720	0.8749
8			0.9992	0.9999	0.9369	0.9463
9			0.9998	1.0000	0.9718	0.9803
10			1.0000	1.0000	0.9885	0.9941

Example 6.15: The probability that a patient recovers from a rare blood disease is 0.4. If 100 people are known to have contracted this disease, what is the probability that fewer than 30 survive?

Solution: Let the binomial variable X represent the number of patients who survive. Since $n = 100$, we should obtain fairly accurate results using the normal-curve approximation with

$$\mu = np = (100)(0.4) = 40 \text{ and } \sigma = \sqrt{npq} = \sqrt{(100)(0.4)(0.6)} = 4.899.$$

To obtain the desired probability, we have to find the area to the left of $x = 29.5$.

The z value corresponding to 29.5 is

$$z = \frac{29.5 - 40}{4.899} = -2.14,$$

and the probability of fewer than 30 of the 100 patients surviving is given by the shaded region in Figure 6.26. Hence,

$$P(X < 30) \approx P(Z < -2.14) = 0.0162.$$

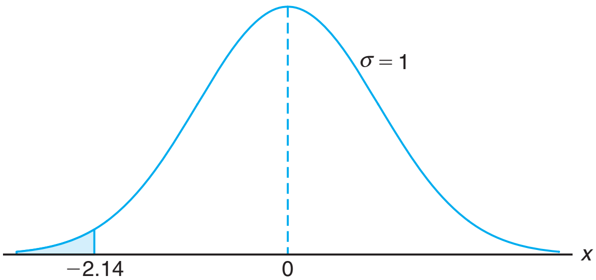

Figure 6.26: Area for Example 6.15.

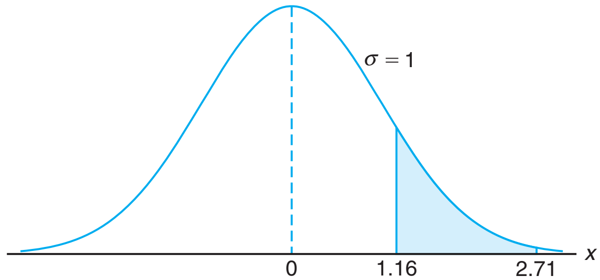

Figure 6.27: Area for Example 6.16.

Example 6.16: A multiple-choice quiz has 200 questions, each with 4 possible answers of which only 1 is correct. What is the probability that sheer guesswork yields from 25 to 30 correct answers for the 80 of the 200 problems about which the student has no knowledge?

***Solution*:** The probability of guessing a correct answer for each of the 80 questions is $p = 1/4$. If X represents the number of correct answers resulting from guesswork, then

$$P(25 \le X \le 30) = \sum_{x=25}^{30} b(x; 80, 1/4).$$

Using the normal curve approximation with

$$\mu = np = (80)\left(\frac{1}{4}\right) = 20$$

and

$$\sigma = \sqrt{npq} = \sqrt{(80)(1/4)(3/4)} = 3.873,$$

we need the area between $x_1 = 24.5$ and $x_2 = 30.5$. The corresponding z values are

$$z_1 = \frac{24.5 - 20}{3.873} = 1.16 \text{ and } z_2 = \frac{30.5 - 20}{3.873} = 2.71.$$

The probability of correctly guessing from 25 to 30 questions is given by the shaded region in Figure 6.27. From Table A.3 we find that

$$\begin{aligned} P(25 \le X \le 30) &= \sum_{x=25}^{30} b(x; 80, 0.25) \approx P(1.16 < Z < 2.71) \\ &= P(Z < 2.71) - P(Z < 1.16) = 0.9966 - 0.8770 = 0.1196. \end{aligned}$$

Exercises

6.24 A coin is tossed 400 times. Use the normal curve approximation to find the probability of obtaining

(a) between 185 and 210 heads inclusive;

(b) exactly 205 heads;

(c) fewer than 176 or more than 227 heads.

6.25 A process for manufacturing an electronic component yields items of which 1% are defective. A quality control plan is to select 100 items from the process, and if none are defective, the process continues. Use the normal approximation to the binomial to find

(a) the probability that the process continues given the sampling plan described;

(b) the probability that the process continues even if the process has gone bad (i.e., if the frequency of defective components has shifted to 5.0% defective).

6.26 A process yields 10% defective items. If 100 items are randomly selected from the process, what is the probability that the number of defectives

(a) exceeds 13?

(b) is less than 8?

6.27 The probability that a patient recovers from a delicate heart operation is 0.9. Of the next 100 patients having this operation, what is the probability that

(a) between 84 and 95 inclusive survive?

(b) fewer than 86 survive?

6.28 Researchers at George Washington University and the National Institutes of Health claim that approximately 75% of people believe "tranquilizers work very well to make a person more calm and relaxed." Of the next 80 people interviewed, what is the probability that

(a) at least 50 are of this opinion?

(b) at most 56 are of this opinion?

6.29 If 20% of the residents in a U.S. city prefer a white telephone over any other color available, what is the probability that among the next 1000 telephones installed in that city

(a) between 170 and 185 inclusive will be white?

(b) at least 210 but not more than 225 will be white?

6.30 A drug manufacturer claims that a certain drug cures a blood disease, on the average, 80% of the time. To check the claim, government testers use the drug on a sample of 100 individuals and decide to accept the claim if 75 or more are cured.

(a) What is the probability that the claim will be rejected when the cure probability is, in fact, 0.8?

(b) What is the probability that the claim will be accepted by the government when the cure probability is as low as 0.7?

6.31 One-sixth of the male freshmen entering a large state school are out-of-state students. If the students are assigned at random to dormitories, 180 to a building, what is the probability that in a given dormitory at least one-fifth of the students are from out of state?

6.32 A pharmaceutical company knows that approximately 5% of its birth-control pills have an ingredient that is below the minimum strength, thus rendering the pill ineffective. What is the probability that fewer than 10 in a sample of 200 pills will be ineffective?

6.33 Statistics released by the National Highway Traffic Safety Administration and the National Safety Council show that on an average weekend night, 1 out of every 10 drivers on the road is drunk. If 400 drivers are randomly checked next Saturday night, what is the probability that the number of drunk drivers will be

(a) less than 32?

(b) more than 49?

(c) at least 35 but less than 47?

6.34 A pair of dice is rolled 180 times. What is the probability that a total of 7 occurs

(a) at least 25 times?

(b) between 33 and 41 times inclusive?

(c) exactly 30 times?

6.35 A company produces component parts for an engine. Parts specifications suggest that 95% of items meet specifications. The parts are shipped to customers in lots of 100.

(a) What is the probability that more than 2 items in a given lot will be defective?

(b) What is the probability that more than 10 items in a lot will be defective?

6.36 A common practice of airline companies is to sell more tickets for a particular flight than there are seats on the plane, because customers who buy tickets do not always show up for the flight. Suppose that the percentage of no-shows at flight time is 2%. For a particular flight with 197 seats, a total of 200 tickets were sold. What is the probability that the airline overbooked this flight?

6.37 The serum cholesterol level X in 14-year-old boys has approximately a normal distribution with mean 170 and standard deviation 30.

(a) Find the probability that the serum cholesterol level of a randomly chosen 14-year-old boy exceeds 230.

(b) In a middle school there are 300 14-year-old boys. Find the probability that at least 8 boys have a serum cholesterol level that exceeds 230.

6.38 A telemarketing company has a special letter-opening machine that opens and removes the contents of an envelope. If the envelope is fed improperly into the machine, the contents of the envelope may not be removed or may be damaged. In this case, the machine is said to have "failed."

(a) If the machine has a probability of failure of 0.01, what is the probability of more than 1 failure occurring in a batch of 20 envelopes?

(b) If the probability of failure of the machine is 0.01 and a batch of 500 envelopes is to be opened, what is the probability that more than 8 failures will occur?

6.6 Gamma and Exponential Distributions

Although the normal distribution can be used to solve many problems in engineering and science, there are still numerous situations that require different types of density functions. Two such density functions, the **gamma** and **exponential distributions**, are discussed in this section.

It turns out that the exponential distribution is a special case of the gamma distribution. Both find a large number of applications. The exponential and gamma distributions play an important role in both queuing theory and reliability problems. Time between arrivals at service facilities and time to failure of component parts and electrical systems often are nicely modeled by the exponential distribution. The relationship between the gamma and the exponential allows the gamma to be used in similar types of problems. More details and illustrations will be supplied later in the section.

The gamma distribution derives its name from the well-known **gamma function**, studied in many areas of mathematics. Before we proceed to the gamma distribution, let us review this function and some of its important properties.

Definition 6.2: The **gamma function** is defined by

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}\, dx, \qquad \text{for } \alpha > 0.$$

The following are a few simple properties of the gamma function.

(a) $\Gamma(n) = (n-1)(n-2)\cdots(1)\Gamma(1)$, for a positive integer n.

To see the proof, integrating by parts with $u = x^{\alpha-1}$ and $dv = e^{-x}\,dx$, we obtain

$$\Gamma(\alpha) = -e^{-x}\, x^{\alpha-1}\Big|_0^\infty + \int_0^\infty e^{-x}(\alpha-1)x^{\alpha-2}\, dx = (\alpha-1)\int_0^\infty x^{\alpha-2}e^{-x}\, dx,$$

for $\alpha > 1$, which yields the recursion formula

$$\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1).$$

The result follows after repeated application of the recursion formula. Using this result, we can easily show the following two properties.

(b) $\Gamma(n) = (n-1)!$ for a positive integer n.

(c) $\Gamma(1) = 1$.

Furthermore, we have the following property of $\Gamma(\alpha)$, which is left for the reader to verify (see Exercise 6.39 on page 206).

(d) $\Gamma(1/2) = \sqrt{\pi}$.

The following is the definition of the **gamma distribution**.

Gamma Distribution

The continuous random variable X has a **gamma distribution**, with parameters α and β, if its density function is given by

$$f(x;\alpha,\beta) = \begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}, & x > 0, \\ 0, & \text{elsewhere}, \end{cases}$$

where $\alpha > 0$ and $\beta > 0$.

Graphs of several gamma distributions are shown in Figure 6.28 for certain specified values of the parameters α and β. The special gamma distribution for which $\alpha = 1$ is called the **exponential distribution**.

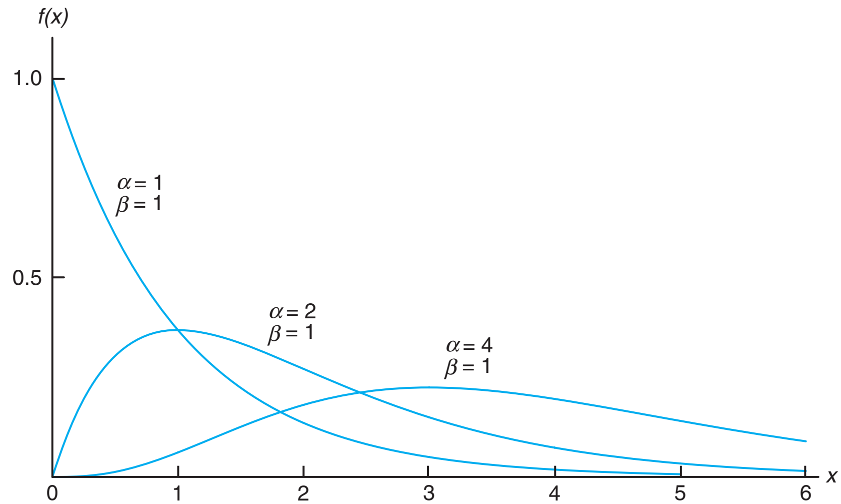

Figure 6.28: Gamma distributions.

Exponential Distribution

The continuous random variable X has an **exponential distribution**, with parameter β, if its density function is given by

$$f(x;\beta) = \begin{cases} \frac{1}{\beta} e^{-x/\beta}, & x > 0, \\ 0, & \text{elsewhere}, \end{cases}$$

where $\beta > 0$.

The following theorem and corollary give the mean and variance of the gamma and exponential distributions.

Theorem 6.4: The mean and variance of the gamma distribution are

$$\mu = \alpha\beta \text{ and } \sigma^2 = \alpha\beta^2.$$

The proof of this theorem is found in Appendix A.26.

Corollary 6.1: The mean and variance of the exponential distribution are

$$\mu = \beta \text{ and } \sigma^2 = \beta^2.$$

Relationship to the Poisson Process

We shall pursue applications of the exponential distribution and then return to the gamma distribution. The most important applications of the exponential distribution are situations where the Poisson process applies (see Section 5.5). The reader should recall that the Poisson process allows for the use of the discrete distribution called the Poisson distribution. Recall that the Poisson distribution is used to compute the probability of specific numbers of "events" during a particular *period of time or span of space*. In many applications, the time period or span of space is the random variable. For example, an industrial engineer may be interested in modeling the time T between arrivals at a congested intersection during rush hour in a large city. An arrival represents the Poisson event.

The relationship between the exponential distribution (often called the negative exponential) and the Poisson process is quite simple. In Chapter 5, the Poisson distribution was developed as a single-parameter distribution with parameter λ, where λ may be interpreted as the mean number of events *per unit "time."* Consider now the *random variable* described by the time required for the first event to occur. Using the Poisson distribution, we find that the probability of no events occurring in the span up to time t is given by

$$p(0; \lambda t) = \frac{e^{-\lambda t}(\lambda t)^0}{0!} = e^{-\lambda t}.$$

We can now make use of the above and let X be the time to the first Poisson event. The probability that the length of time until the first event will exceed x is the same as the probability that no Poisson events will occur in x. The latter, of course, is given by $e^{-\lambda x}$. As a result,

$$P(X > x) = e^{-\lambda x}.$$

Thus, the cumulative distribution function for X is given by

$$P(0 \leq X \leq x) = 1 - e^{-\lambda x}.$$

Now, in order that we may recognize the presence of the exponential distribution, we differentiate the cumulative distribution function above to obtain the density function

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

which is the density function of the exponential distribution with $\lambda = 1/\beta$.

Applications of the Exponential and Gamma Distributions

In the foregoing, we provided the foundation for the application of the exponential distribution in "time to arrival" or time to Poisson event problems. We will illustrate some applications here and then proceed to discuss the role of the gamma distribution in these modeling applications. Notice that the mean of the exponential distribution is the parameter β, the reciprocal of the parameter in the Poisson distribution. The reader should recall that it is often said that the Poisson distribution has no memory, implying that occurrences in successive time periods are independent. The important parameter β is the mean time between events. In reliability theory, where equipment failure often conforms to this Poisson process, β is called **mean time between failures**. Many equipment breakdowns do follow the Poisson process, and thus the exponential distribution does apply. Other applications include survival times in biomedical experiments and computer response time.

In the following example, we show a simple application of the exponential distribution to a problem in reliability. The binomial distribution also plays a role in the solution.

Example 6.17: Suppose that a system contains a certain type of component whose time, in years, to failure is given by T. The random variable T is modeled nicely by the exponential distribution with mean time to failure $\beta = 5$. If 5 of these components are installed in different systems, what is the probability that at least 2 are still functioning at the end of 8 years?

***Solution*:** The probability that a given component is still functioning after 8 years is given by

$$P(T > 8) = \frac{1}{5}\int_8^\infty e^{-t/5}\, dt = e^{-8/5} \approx 0.2.$$

Let X represent the number of components functioning after 8 years. Then using the binomial distribution, we have

$$P(X \geq 2) = \sum_{x=2}^{5} b(x; 5, 0.2) = 1 - \sum_{x=0}^{1} b(x; 5, 0.2) = 1 - 0.7373 = 0.2627.$$

There are exercises and examples in Chapter 3 where the reader has already encountered the exponential distribution. Others involving waiting time and reliability include Example 6.24 and some of the exercises and review exercises at the end of this chapter.

The Memoryless Property and Its Effect on the Exponential Distribution

The types of applications of the exponential distribution in reliability and component or machine lifetime problems are influenced by the **memoryless** (or lack-of-memory) **property** of the exponential distribution. For example, in the case of, say, an electronic component where lifetime has an exponential distribution, the probability that the component lasts, say, t hours, that is, $P(X \geq t)$, is the same as the conditional probability

$$P(X \geq t_0 + t \mid X \geq t_0).$$

So if the component "makes it" to t_0 hours, the probability of lasting an additional t hours is the same as the probability of lasting t hours. There is no "punishment" through wear that may have ensued for lasting the first t_0 hours. Thus, the exponential distribution is more appropriate when the memoryless property is justified. But if the failure of the component is a result of gradual or slow wear (as in mechanical wear), then the exponential does not apply and either the gamma or the Weibull distribution (Section 6.10) may be more appropriate.

The importance of the gamma distribution lies in the fact that it defines a family of which other distributions are special cases. But the gamma itself has important applications in waiting time and reliability theory. Whereas the exponential distribution describes the time until the occurrence of a Poisson event (or the time between Poisson events), the time (or space) occurring until a *specified number of Poisson events occur* is a random variable whose density function is described by the gamma distribution. This specific number of events is the parameter α in the gamma density function. Thus, it becomes easy to understand that when $\alpha = 1$, the special case of the exponential distribution occurs. The gamma density can be developed from its relationship to the Poisson process in much the same manner as we developed the exponential density. The details are left to the reader. The following is a numerical example of the use of the gamma distribution in a waiting-time application.

Example 6.18: Suppose that telephone calls arriving at a particular switchboard follow a Poisson process with an average of 5 calls coming per minute. What is the probability that up to a minute will elapse by the time 2 calls have come in to the switchboard?

***Solution*:** The Poisson process applies, with time until 2 Poisson events following a gamma distribution with $\beta = 1/5$ and $\alpha = 2$. Denote by X the time in minutes that transpires before 2 calls come. The required probability is given by

$$P(X \leq 1) = \int_0^1 \frac{1}{\beta^2} x e^{-x/\beta}\, dx = 25 \int_0^1 x e^{-5x}\, dx = 1 - e^{-5}(1+5) = 0.96.$$

While the origin of the gamma distribution deals in time (or space) until the occurrence of α Poisson events, there are many instances where a gamma distribution works very well even though there is no clear Poisson structure. This is particularly true for **survival time** problems in both engineering and biomedical applications.

Example 6.19: In a biomedical study with rats, a dose-response investigation is used to determine the effect of the dose of a toxicant on their survival time. The toxicant is one that is frequently discharged into the atmosphere from jet fuel. For a certain dose of the toxicant, the study determines that the survival time, in weeks, has a gamma distribution with $\alpha = 5$ and $\beta = 10$. What is the probability that a rat survives no longer than 60 weeks?

Solution: Let the random variable X be the survival time (time to death). The required probability is

$$P(X \le 60) = \frac{1}{\beta^5} \int_0^{60} \frac{x^{\alpha-1} e^{-x/\beta}}{\Gamma(5)}\, dx.$$

The integral above can be solved through the use of the **incomplete gamma function**, which becomes the cumulative distribution function for the gamma distribution. This function is written as

$$F(x; \alpha) = \int_0^x \frac{y^{\alpha-1} e^{-y}}{\Gamma(\alpha)}\, dy.$$

If we let $y = x/\beta$, so $x = \beta y$, we have

$$P(X \le 60) = \int_0^6 \frac{y^4 e^{-y}}{\Gamma(5)}\, dy,$$

which is denoted as $F(6; 5)$ in the table of the incomplete gamma function in Appendix A.23. Note that this allows a quick computation of probabilities for the gamma distribution. Indeed, for this problem, the probability that the rat survives no longer than 60 days is given by

$$P(X \le 60) = F(6; 5) = 0.715.$$

Example 6.20: It is known, from previous data, that the length of time in months between customer complaints about a certain product is a gamma distribution with $\alpha = 2$ and $\beta = 4$. Changes were made to tighten quality control requirements. Following these changes, 20 months passed before the first complaint. Does it appear as if the quality control tightening was effective?

Solution: Let X be the time to the first complaint, which, under conditions prior to the changes, followed a gamma distribution with $\alpha = 2$ and $\beta = 4$. The question centers around how rare $X \ge 20$ is, given that α and β remain at values 2 and 4, respectively. In other words, under the prior conditions is a "time to complaint" as large as 20 months reasonable? Thus, following the solution to Example 6.19,

$$P(X \ge 20) = 1 - \frac{1}{\beta^\alpha} \int_0^{20} \frac{x^{\alpha-1} e^{-x/\beta}}{\Gamma(\alpha)} dx.$$

Again, using $y = x/\beta$, we have

$$P(X \ge 20) = 1 - \int_0^5 \frac{y e^{-y}}{\Gamma(2)}\, dy = 1 - F(5; 2) = 1 - 0.96 = 0.04,$$

where $F(5; 2) = 0.96$ is found from Table A.23.

As a result, we could conclude that the conditions of the gamma distribution with $\alpha = 2$ and $\beta = 4$ are not supported by the data that an observed time to complaint is as large as 20 months. Thus, it is reasonable to conclude that the quality control work was effective.

Example 6.21: Consider Exercise 3.31 on page 94. Based on extensive testing, it is determined that the time Y in years before a major repair is required for a certain washing machine is characterized by the density function

$$f(y) = \begin{cases} \frac{1}{4} e^{-y/4}, & y \ge 0, \\ 0, & \text{elsewhere.} \end{cases}$$

Note that Y is an exponential random variable with $\mu = 4$ years. The machine is considered a bargain if it is unlikely to require a major repair before the sixth year. What is the probability $P(Y > 6)$? What is the probability that a major repair is required in the first year?

Solution: Consider the cumulative distribution function $F(y)$ for the exponential distribution,

$$F(y) = \frac{1}{\beta}\int_0^y e^{-t/\beta}\, dt = 1 - e^{-y/\beta}.$$

Then

$$P(Y > 6) = 1 - F(6) = e^{-3/2} = 0.2231.$$

Thus, the probability that the washing machine will require major repair after year six is 0.223. Of course, it will require repair before year six with probability 0.777. Thus, one might conclude the machine is not really a bargain. The probability that a major repair is necessary in the first year is

$$P(Y < 1) = 1 - e^{-1/4} = 1 - 0.779 = 0.221.$$

6.7 Chi-Squared Distribution

Another very important special case of the gamma distribution is obtained by letting $\alpha = v/2$ and $\beta = 2$, where v is a positive integer. The result is called **the chi-squared distribution**. The distribution has a single parameter, v, called the **degrees of freedom**.

Chi-Squared Distribution

The continuous random variable X has a **chi-squared distribution**, with v **degrees of freedom**, if its density function is given by

$$f(x; v) = \begin{cases} \frac{1}{2^{v/2}\Gamma(v/2)} x^{v/2-1} e^{-x/2}, & x > 0, \\ 0, & \text{elsewhere,} \end{cases}$$

where v is a positive integer.

The chi-squared distribution plays a vital role in statistical inference. It has considerable applications in both methodology and theory. While we do not discuss applications in detail in this chapter, it is important to understand that Chapters 8, 9, and 16 contain important applications. The chi-squared distribution is an important component of statistical hypothesis testing and estimation.

Topics dealing with sampling distributions, analysis of variance, and nonparametric statistics involve extensive use of the chi-squared distribution.

Theorem 6.5: The mean and variance of the chi-squared distribution are

$$\mu = v \text{ and } \sigma^2 = 2v.$$