Example **Determining the mean and variance using the probability density function**

With reference to the example on page 150, find the mean and the variance of the given probability density.

Solution Performing the necessary integrations, using integrations by parts, we get

$$\mu = \int_{-\infty}^{\infty} x f(x)\, dx = \int_{0}^{\infty} x \cdot 2e^{-2x}\, dx = \frac{1}{2}$$

and

$$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx = \int_{0}^{\infty} \left(x - \frac{1}{2}\right)^2 \cdot 2e^{-2x}\, dx = \frac{1}{4}.$$ ■

Exercises

5.1 Verify that the function of the example on page 150 is, in fact, a probability density.

5.2 If the probability density of a random variable is given by

$$f(x) = \begin{cases} kx^3 & 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$$

find the value k and the probability that the random variable takes on a value

(a) between $\frac{1}{4}$ and $\frac{3}{4}$; (b) greater than $\frac{2}{3}$.

5.3 With reference to the preceding exercise, find the corresponding distribution function and use it to determine the probabilities that a random variable having this distribution function will take on a value

(a) greater than 0.8; (b) between 0.2 and 0.4.

5.4 If the probability density of a random variable is given by

$$f(x) = \begin{cases} x & \text{for } 0 < x < 1 \\ 2 - x & \text{for } 1 \le x < 2 \\ 0 & \text{elsewhere} \end{cases}$$

find the probabilities that a random variable having this probability density will take on a value

(a) between 0.2 and 0.8; (b) between 0.6 and 1.2.

5.5 With reference to the preceding exercise, find the corresponding distribution function, and use it to determine the probabilities that a random variable having the distribution function will take on a value

(a) greater than 1.8; (b) between 0.4 and 1.6.

5.6 Given the probability density $f(x) = \dfrac{k}{1+x^2}$ for $-\infty < x < \infty$, find k.

5.7 If the distribution function of a random variable is given by

$$F(x) = \begin{cases} 1 - \dfrac{4}{x^2} & \text{for } x > 2 \\ 0 & \text{for } x \le 2 \end{cases}$$

find the probabilities that this random variable will take on a value

(a) less than 3; (b) between 4 and 5.

5.8 Find the probability density that corresponds to the distribution function of Exercise 5.7. Are there any points at which it is undefined? Also sketch the graphs of the distribution function and the probability density.

5.9 Let the phase error in a tracking device have probability density

$$f(x) = \begin{cases} \cos x & 0 < x < \pi/2 \\ 0 & \text{elsewhere} \end{cases}$$

Find the probability that the phase error is

(a) between 0 and $\pi/4$; (b) greater than $\pi/3$.

5.10 The mileage (in thousands of miles) that car owners get with a certain kind of tire is a random variable having the probability density

$$f(x) = \begin{cases} \dfrac{1}{20}\, e^{-x/20} & \text{for } x > 0 \\ 0 & \text{for } x \le 0 \end{cases}$$

Find the probabilities that one of these tires will last

(a) at most 10,000 miles;
(b) anywhere from 16,000 to 24,000 miles;
(c) at least 30,000 miles.

5.11 In a certain city, the daily consumption of electric power (in millions of kilowatt hours) is a random variable having the probability density

$$f(x) = \begin{cases} \dfrac{1}{9}\, x e^{-x/3} & \text{for } x > 0 \\ 0 & \text{for } x \le 0 \end{cases}$$

If the city's power plant has a daily capacity of 12 million kilowatt-hours, what is the probability that this power supply will be inadequate on any given day?

5.12 Prove that the identity $\sigma^2 = \mu_2' - \mu_2$ holds for any probability density for which these moments exist.

5.13 Find μ and σ^2 for the probability density of Exercise 5.2.

5.14 Find μ and σ^2 for the probability density of Exercise 5.4.

5.15 Find μ and σ for the probability density obtained in Exercise 5.8.

5.16 Find μ and σ for the distribution of the phase error of Exercise 5.9.

5.17 Find μ for the distribution of the mileages of Exercise 5.10.

5.18 Show that μ_2' and, hence, σ^2 do not exist for the probability density of Exercise 5.6.

5.2 The Normal Distribution

Among the special probability densities we shall study in this chapter, the **normal probability density**, usually referred to simply as the **normal distribution**, is by far the most important.[1] It was studied first in the eighteenth century when scientists observed an astonishing degree of regularity in errors of measurement. They found that the patterns (distributions) they observed were closely approximated by a continuous distribution which they referred to as the "normal curve

[1] The words *density* and *distribution* are often used interchangeably in the literature of applied statistics.

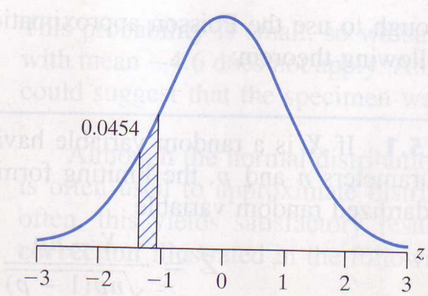

Figure 5.17
Normal approximation to $P(X = 15)$

(b)

$$F\left(\frac{15.5-20}{4}\right) - F\left(\frac{14.5-20}{4}\right) = F(-1.13) - F(-1.38)$$
$$= 0.1292 - 0.0838$$
$$= 0.0454$$

as indicated in Figure 5.17. ■

Had we done the exact binomial calculation on a computer instead of using normal approximation in the preceding example, we would have obtained 0.1285 instead of 0.1292 for part a and 0.0481 instead of 0.0454 for part b. Thus, it can be seen that both approximations are very close.

A good rule of thumb for the normal approximation

Use the normal approximation to the binomial distribution only when np and $n(1-p)$ are both greater than 15.

Exercises

5.19 If a random variable has the standard normal distribution, find the probability that it will take on a value

(a) less than 1.50;
(b) less than −1.20;
(c) greater than 2.16;
(d) greater than −1.75.

5.20 If a random variable has the standard normal distribution, find the probability that it will take on a value

(a) between 0 and 2.7;
(b) between 1.22 and 2.43;
(c) between −1.35 and −0.35;
(d) between −1.70 and 1.35.

5.21 Find z if the probability that a random variable having the standard normal distribution will take on a value

(a) less than z is 0.9911;
(b) greater than z is 0.1093;

(c) greater than z is 0.6443;

(d) less than z is 0.0217;

(e) between $-z$ and z is 0.9298.

5.22 If a random variable has a normal distribution, what are the probabilities that it will take on a value within

(a) 1 standard deviation of the mean;

(b) 2 standard deviations of the mean;

(c) 3 standard deviations of the mean;

(d) 4 standard deviations of the mean?

5.23 Verify that

(a) $z_{0.005} = 2.575$;

(b) $z_{0.025} = 1.96$.

5.24 Given a random variable having the normal distribution with $\mu = 16.2$ and $\sigma^2 = 1.5625$, find the probabilities that it will take on a value

(a) greater than 16.8;

(b) less than 14.9;

(c) between 13.6 and 18.8;

(d) between 16.5 and 16.7.

5.25 The time for a super glue to set can be treated as a random variable having a normal distribution with mean 30 seconds. Find its standard deviation if the probability is 0.20 that it will take on a value greater than 39.2 seconds.

5.26 The time to microwave a bag of popcorn using the automatic setting can be treated as a random variable having a normal distribution standard deviation 10 seconds. If the probability is 0.8212 that the bag will take less than 282.5 seconds to pop, find the probability (a) that it will take longer than 258.3 seconds to pop and (b) that it will take on a value greater than 39.2 seconds.

5.27 The time required to assemble a piece of machinery is a random variable having approximately a normal distribution with $\mu = 12.9$ minutes and $\sigma = 2.0$ minutes. What are the probabilities that the assembly of a piece of machinery of this kind will take

(a) at least 11.5 minutes;

(b) anywhere from 11.0 to 14.8 minutes?

5.28 Find the *quartiles*

$$-z_{0.25} \quad z_{0.50} \quad z_{0.25}$$

of the standard normal distribution.

5.29 In a photographic process, the developing time of prints may be looked upon as a random variable having the normal distribution with a mean of 16.28 seconds and a standard deviation of 0.12 second. Find the probability that it will take

(a) anywhere from 16.00 to 16.50 seconds to develop one of the prints;

(b) at least 16.20 seconds to develop one of the prints;

(c) at most 16.35 seconds to develop one of the prints.

5.30 With reference to the preceding exercise, for which value is the probability 0.95 that it will be exceeded by the time it takes to develop one of the prints?

5.31 Specifications for a certain job call for washers with an inside diameter of 0.300 ± 0.005 inch. If the inside diameters of the washers supplied by a given manufacturer may be looked upon as a random variable having the normal distribution with $\mu = 0.302$ inch and $\sigma = 0.003$ inch, what percentage of these washers will meet specifications?

5.32 With reference to the example on page 158, verify that if the variability of the filling machine is reduced to $\sigma = 0.025$ ounce, this will lower the required average amount of coffee to 4.05 ounces, yet keep 98% of the jars above 4 ounces.

5.33 A stamping machine produces can tops whose diameters are normally distributed with a standard deviation of 0.01 inch. At what "normal" (mean) diameter should the machine be set so that no more than 5% of the can tops produced have diameters exceeding 3 inches?

5.34 Extruded plastic rods are automatically cut into nominal lengths of 6 inches. Actual lengths are normally distributed about a mean of 6 inches and their standard deviation is 0.06 inch.

(a) What proportion of the rods have lengths that are outside the tolerance limits of 5.9 inches to 6.1 inches?

(b) To what value does the standard deviation need to be reduced if 99% of the rods must be within tolerance?

5.35 If a random variable has the binomial distribution with $n = 40$ and $p = 0.60$, use the normal approximation to determine the probabilities that it will take on

(a) the value 14;

(b) a value less than 12.

5.36 A manufacturer knows that, on average, 2% of the electric toasters that he makes will require repairs within 90 days after they are sold. Use the normal approximation to the binomial distribution to determine the probability that among 1,200 of these toasters at least 30 will require repairs within the first 90 days after they are sold.

5.37 The probability that an electronic component will fail in less than 1,000 hours of continuous use is 0.25. Use the normal approximation to find the probability that among 200 such components fewer than 45 will fail in less than 1,000 hours of continuous use.

5.38 A safety engineer feels that 30% of all industrial accidents in her plant are caused by failure of employees to follow instructions. If this figure is correct, find, approximately, the probability that among 84 industrialized accidents in this plant anywhere from 20 to 30 (inclusive) will be due to failure of employees to follow instructions.

5.39 If 62% of all clouds seeded with silver iodide show spectacular growth, what is the probability that among 40 clouds seeded with silver iodide at most 20 will show spectacular growth?

5.40 To illustrate the law of large numbers mentioned on page 125, find the probabilities that the proportion of heads will be anywhere from 0.49 to 0.51 when a balanced coin is flipped

(a) 1,000 times;

(b) 10,000 times.

5.41 Verify the identity $F(-z) = 1 - F(z)$ given on page 154.

5.42 Verify that the parameter μ in the expression for the normal density on page 154, is, in fact, its mean.

5.43 Verify that the parameter σ^2 in the expression for the normal density on page 154 is, in fact, its variance.

5.44 Normal probabilities can be calculated using MINITAB. Let X have a normal distribution with mean 11.3 and standard deviation 5.7. The following steps yield the cumulative probability of 9.31 or smaller, or $P(X \le 9.31)$.

Dialog box:
Calc > **Probability Distribution** > **Normal**
Choose **Cumulative Distribution.** Choose **Input constant** and enter 9.31.
Type 11.3 in **Mean** and 5.7 in **standard deviation.**
Click **OK**.

Output: Normal with mean = 11.3000 and standard deviation = 5.70000

x	P(X <= x)
9.3100	0.3635

For this same normal distribution, find the probability
(a) of 8.493 or smaller,
(b) of 16.074 or smaller.

5.4 Other Probability Densities

In the application of statistics to problems in engineering and physical science, we shall encounter many probability densities other than the normal distribution. Among these are the t, F, and chi-square distributions, the fundamental sampling distributions that we introduce in Chapter 6, and the exponential and Weibull distributions, which we apply to problems of reliability and life testing in Chapter 15.

In the remainder of this chapter we shall discuss five continuous distributions, the **uniform distribution**, the **log-normal distribution**, the **gamma distribution**, the **beta distribution**, and the **Weibull distribution**, for the twofold purpose of giving further examples of probability densities and of laying the foundation for future applications.

5.5 The Uniform Distribution

The **uniform distribution**, with the parameters α and β, has probability density function

Uniform distribution

$$f(x) = \begin{cases} \dfrac{1}{\beta - \alpha} & \text{for } \alpha < x < \beta \\ 0 & \text{elsewhere} \end{cases}$$

whose graph is shown in Figure 5.18. Note that all values of x from α to β are "equally likely" in the sense that the probability that x lies in an interval of width Δx entirely contained in the interval from α to β is equal to $\Delta x/(\beta - \alpha)$, regardless of the exact location of the interval.

To illustrate how a physical situation might give rise to a uniform distribution, suppose that a wheel of a locomotive has the radius r and that x is the location of a point on its circumference measured along the circumference from some reference point 0. When the brakes are applied, some point will

Exercises

5.45 Find the distribution function of a random variable which has the uniform distribution.

5.46 In certain experiments, the error made in determining the solubility of a substance is a random variable having the uniform density with $\alpha = -0.025$ and $\beta = 0.025$. What are the probabilities that such an error will be

(a) between 0.010 and 0.015;

(b) between -0.012 and 0.012?

5.47 From experience Mr. Harris has found that the low bid on a construction job can be regarded as a random variable having the uniform density

$$f(x) = \begin{cases} \dfrac{3}{4C} & \text{for } \dfrac{2C}{3} < x < 2C \\ 0 & \text{elsewhere} \end{cases}$$

where C is his own estimate of the cost of the job. What percentage should Mr. Harris add to his cost estimate when submitting bids to maximize his expected profit?

5.48 Verify the expression given on page 168 for the mean of the log-normal distribution.

5.49 With reference to the example on page 167, find the probability that I_o/I_i will take on a value between 7.0 and 7.5.

5.50 If a random variable has the log-normal distribution with $\alpha = -1$ and $\beta = 2$, find its mean and its standard deviation.

5.51 With reference to the preceding exercise, find the probabilities that the random variable will take on a value

(a) between 3.2 and 8.4;

(b) greater than 5.0.

5.52 If a random variable has the gamma distribution with $\alpha = 2$ and $\beta = 2$, find the mean and the standard deviation of this distribution.

5.53 With reference to Exercise 5.52, find the probability that the random variable will take on a value less than 4.

5.54 In a certain city, the daily consumption of electric power (in millions of kilowatt-hours) can be treated as a random variable having a gamma distribution with $\alpha = 3$ and $\beta = 2$. If the power plant of this city has a daily capacity of 12 million kilowatt-hours, what is the probability that this power supply will be inadequate on any given day?

5.55 With reference to the example on page 168, suppose the expert opinion is in error. Calculate the probability that the supports will survive if

(a) $\mu = 3.0$ and $\sigma^2 = 0.09$;

(b) $\mu = 4.0$ and $\sigma^2 = 0.36$.

5.56 Verify the expression for the variance of the gamma distribution given on page 170.

5.57 Show that when $\alpha > 1$, the graph of the gamma density has a relative maximum at $x = \beta(\alpha - 1)$. What happens when $0 < \alpha < 1$ and when $\alpha = 1$?

5.58 The amount of time that a surveillance camera will run without having to be reset is a random variable having the exponential distribution with $\beta = 50$ days. Find the probabilities that such a camera will

(a) have to be reset in less than 20 days;

(b) not have to be reset in at least 60 days.

5.59 With reference to Exercise 4.97, find the percent of the time that the interval between breakdowns of the computer will be

(a) less than 1 week;

(b) at least 5 weeks.

5.60 With reference to Exercise 4.58, find the probabilities that the time between successive calls arriving at the switchboard of the consulting firm will be

(a) less than $\frac{1}{2}$ minute;

(b) more than 3 minutes.

5.61 Given a Poisson process with on the average α arrivals per unit time, find the probability that there will be no arrivals during a time interval of length t, namely, the probability that the waiting times between successive arrivals will be at least of length t.

5.62 Use the result of Exercise 5.61 to find an expression for the probability density of the waiting time between successive arrivals.

5.63 Verify for $\alpha = 3$ and $\beta = 3$ that the integral of the beta density, from 0 to 1, is equal to 1.

5.64 If the annual proportion of erroneous income tax returns filed with the IRS can be looked upon as a random variable having a beta distribution with $\alpha = 2$ and $\beta = 9$, what is the probability that in any given year there will be fewer than 10% erroneous returns?

5.65 Suppose that the proportion of defectives shipped by a vendor, which varies somewhat from shipment to shipment, may be looked upon as a random variable having the beta distribution with $\alpha = 1$ and $\beta = 4$.

(a) Find the mean of this beta distribution, namely, the average proportion of defectives in a shipment from this vendor.

(b) Find the probability that a shipment from this vendor will contain 25% or more defectives.

5.66 Show that when $\alpha > 1$ and $\beta > 1$, the beta density has a relative maximum at

$$x = \frac{\alpha - 1}{\alpha + \beta - 2}.$$

5.67 With reference to the example on page 174, find the probability that such a battery will not last 100 hours.

5.68 Suppose that the time to failure (in minutes) of certain electronic components subjected to continuous vibrations may be looked upon as a random variable having the Weibull distribution with $\alpha = \frac{1}{5}$ and $\beta = \frac{1}{3}$.

(a) How long can such a component be expected to last?

(b) What is the probability that such a component will fail in less than 5 hours?

5.69 Suppose that the service life (in hours) of a semiconductor is a random variable having the Weibull distribution with $\alpha = 0.025$ and $\beta = 0.500$. What is the probability that such a semiconductor will still be in operating condition after 4,000 hours?

5.70 Verify the formula for the variance of the Weibull distribution given on page 174.

5.10 Joint Distributions—Discrete and Continuous

Discrete Variables

Often, experiments are conducted where two random variables are observed simultaneously in order to determine not only their individual behavior but also the degree of relationship between them.

and the last term equals $-2(\overline{X}-\mu)\sum_{1}^{n}(X_i-\mu) = -2n(\overline{X}-\mu)^2$. Consequently,

$$\sum_{i=1}^{n}(X_i-\overline{X})^2 = \sum_{i=1}^{n}(X_i-\mu)^2 - n(\overline{X}-\mu)^2.$$

Since $E(X_i-\mu)^2 = Var(X_i) = \sigma^2$ and, by the previous example, $E(\overline{X}) = \mu$ so $E(\overline{X}-\mu)^2 = Var(\overline{X}) = \sigma^2/n$, taking expectation term by term and summing

$$E\left[\sum_{i=1}^{n}(X_i-\overline{X})^2\right] = \sum_{i=1}^{n}\sigma^2 - n\frac{\sigma^2}{n} = (n-1)\sigma^2$$

Dividing both sides by $n-1$, we conclude that σ^2 is the expected value of the sample variance. ■

Exercises

5.71 Two scanners are needed for an experiment. Of the five available, two have electronic defects, another one has a defect in the memory, and two are in good working order. Two units are selected at random.

(a) Find the joint probability distribution of X_1 = the number with electronic defects, and X_2 = the number with a defect in memory.

(b) Find the probability of 0 or 1 total defects among the two selected.

(c) Find the marginal probability distribution of X_1.

(d) Find the conditional probability distribution of X_1 given $X_2 = 0$.

5.72 Two random variables are independent and each has a binomial distribution with success probability 0.3 and 2 trials.

(a) Find the joint probability distribution.

(b) Find the probability that the second random variable is greater than the first.

5.73 If two random variables have the joint density

$$f(x_1, x_2) = \begin{cases} x_1x_2 & \text{for } 0 < x_1 < 1,\ 0 < x_2 < 2 \\ 0 & \text{elsewhere} \end{cases}$$

find the probabilities that

(a) both random variables will take on values less than 1;

(b) the sum of the values taken on by the two random variables will be less than 1.

5.74 With reference to the preceding exercise, find the marginal densities of the two random variables.

5.75 With reference to Exercise 5.73, find the joint distribution function of the two random variables, the distribution functions of the individual random variables, and check whether the two random variables are independent.

5.76 If two random variables have the joint density

$$f(x, y) = \begin{cases} \frac{6}{5}(x + y^2) & \text{for } 0 < x < 1,\ 0 < y < 1 \\ 0 & \text{elsewhere} \end{cases}$$

find the probability that $0.2 < X < 0.5$ and $0.4 < Y < 0.6$.

5.77 With reference to the preceding exercise, find the joint distribution function of the two random variables and use it to verify the value obtained for the probability.

5.78 With reference to Exercise 5.76, find both marginal densities and use them to find the probabilities that

(a) $X > 0.8$;
(b) $Y < 0.5$.

5.79 With reference to Exercise 5.76, find

(a) an expression for $f_1(x|y)$ for $0 < y < 1$;
(b) an expression for $f_1(x|0.5)$;
(c) the mean of the conditional density of the first random variable when the second takes on the value 0.5.

5.80 With reference to the conditional probability example on page 182, find expressions for

(a) the conditional density of the first random variable when the second takes on the value $x_2 = 0.25$;
(b) the conditional density of the second random variable when the first takes on the value x_1.

5.81 If three random variables have the joint density

$$f(x, y, z) = \begin{cases} k(x+y)e^{-z} & \text{for } 0 < x < 1,\ 0 < y < 2, z > 0 \\ 0 & \text{elsewhere} \end{cases}$$

find

(a) the value of k;
(b) the probability that $X < Y$ and $Z > 1$.

5.82 With reference to the preceding exercise, check whether

(a) the three random variables are independent;
(b) any two of the three random variables are pairwise independent.

5.83 A pair of random variables has the **circular normal distribution** if their joint density is given by

$$f(x_1, x_2) = \frac{1}{2\pi\sigma^2}\, e^{-[(x_1-\mu_1)^2 + (x_2-\mu_2)^2]/2\sigma^2}$$

for $-\infty < x_1 < \infty$ and $-\infty < x_2 < \infty$.

(a) If $\mu_1 = 2$ and $\mu_2 = -2$, and $\sigma = 10$, use Table 3 to find the probability that $-8 < X_1 < 14$ and $-9 < X_2 < 3$.
(b) If $\mu_1 = \mu_2 = 0$ and $\sigma = 3$, find the probability that (X_1, X_2) is contained in the region between the two circles $x_1^2 + x_2^2 = 9$ and $x_1^2 + x_2^2 = 36$.

5.84 A precision drill positioned over a target point will make an acceptable hole if it is within 5 microns of the target. Using the target as the origin of a rectangular system of coordinates, assume that the coordinates (x, y) of the point of contact are values of a pair of random variables having the circular normal distribution (see Exercise 5.83) with $\mu_1 = \mu_2 = 0$ and $\sigma = 2$. What is the probability that the hole will be acceptable?

5.85 With reference to Exercise 5.73, find the expected value of the random variable whose values are given by $g(x_1, x_2) = x_1 + x_2$.

5.86 With reference to Exercise 5.76, find the expected value of the random variable whose values are given by $g(x, y) = x^2 y$.

5.87 If measurements of the length and the width of a rectangle have the joint density

$$f(x, y) = \begin{cases} \dfrac{1}{ab} & \text{for } L - \dfrac{a}{2} < x < L + \dfrac{a}{2}, \quad W - \dfrac{b}{2} < y < W + \dfrac{b}{2} \\ 0 & \text{elsewhere} \end{cases}$$

find the mean and the variance of the corresponding distribution of the area of the rectangle.

5.88 Establish a relationship between $f_1(x_1|x_2)$, $f_2(x_2|x_1)$, $f_1(x_1)$, and $f_2(x_2)$.

5.89 If X_1 has mean 1 and variance 5 while X_2 has mean -1 and variance 5, and the two are independent, find

(a) $E(X_1 + X_2)$;
(b) $Var(X_1 + X_2)$.

5.90 If X_1 has mean -2 and variance 2 while X_2 has mean 5 and variance 5, and the two are independent, find

(a) $E(X_1 - X_2)$;
(b) $Var(X_1 - X_2)$.

5.91 If X_1 has mean 1 and variance 5 while X_2 has mean -2 and variance 5, and the two are independent, find

(a) $E(X_1 + 2X_2 - 3)$;
(b) $Var(X_1 + 2X_2 - 3)$.

5.92 The time for an older machine to complete a check on a computer chip, X_1, has mean 65 milliseconds and variance 16, while the time for a newer model, X_2, has mean 45 milliseconds and variance 9. Find the expected time savings using the newer model when

(a) checking a single chip;
(b) checking 200 chips.
(c) Find the standard deviations in parts a and b, assuming all of the checking times are independent.

5.93 Let $X_1, X_2, \ldots, X_{20}$ be independent and let each have the same marginal distribution with mean 10 and variance 3. Find

(a) $E(X_1 + X_2 + \cdots + X_{20})$;
(b) $Var(X_1 + X_2 + \cdots + X_{20})$.

5.11 Checking if the Data Are Normal

In many instances, an experimenter needs to check whether a data set appears to be generated by a normally distributed random variable. As indicated in Figure 2.8, the normal distribution can serve to model variation in some quantities. Further, many commonly used statistical procedures, which we describe in later chapters, require that the probability distribution be nearly normal. Consequently, in a great number of applications it is prudent to check the assumption that the data are normal.

Although they involve an element of subjective judgment, graphical procedures are the most helpful for detecting serious departures from normality. Histograms can be checked for lack of symmetry. A single long tail certainly contradicts the assumption of a normal distribution. However, another special the Box-Muller-Marsaglia method is almost universally preferred. It starts with a pair of independent uniform variables (u_1, u_2) and produces two standard normal variables

$$z_1 = \sqrt{-2\ln(u_2)}\,\cos(2\pi u_1)$$

$$z_2 = \sqrt{-2\ln(u_2)}\,\sin(2\pi u_1)$$

where the angle is expressed in radians. Then $x_1 = \mu + \sigma z_1$ and $x_2 = \mu + \sigma z_2$ are treated as two independent observations of normal random variables (see Exercise 5.103). Most statistical packages include a normal random number generator (see Exercise 5.106).

Example **Simulating two values from a normal distribution**

Simulate two observations of a random variable having the normal distribution with $\mu = 50$ and $\sigma = 5$.

Solution A computer generates the two values 0.253 and 0.531 from a uniform distribution. (Alternatively, they could have been obtained by reading three digits at a time from Table 7.) We first calculate the standard normal values

$$z_1 = \sqrt{-2\ln(0.531)}\,\cos(2\pi \cdot 0.253) = -0.021$$

$$z_2 = \sqrt{-2\ln(0.531)}\,\sin(2\pi \cdot 0.253) = 1.125$$

and then the normal values

$$x_1 = 50 + 5z_1 = 50 + 5(-0.021) = 49.895$$

$$x_2 = 50 + 5z_2 = 50 + 5(1.125) = 55.625.$$ ■

Exercises

5.98 Continue the example in the text, that is, find the simulated values of the Weibull random variable corresponding to the values 0.26, 0.77, 0.12 of a uniform random variable.

5.99 Suppose the number of hours it takes a person to learn how to operate a certain machine is a random variable having a normal distribution with $\mu = 5.8$ and $\sigma = 1.2$. Suppose it takes two persons to operate the machine. Simulate the time it takes four pairs of persons to learn how to operate the machine. That is, for each pair, calculate the maximum of the two learning times.

5.100 Suppose that the durability of paint (in years) is a random variable having an exponential distribution (see page 170) with mean $\beta = 2$. Simulating an experiment in which five houses are painted,

(a) find the time of the first failure;

(b) find the time of the fifth failure.

5.101 Verify that

(a) the exponential density $0.3e^{-0.3x}$, $x > 0$ corresponds to the distribution function $F(x) = 1 - e^{-0.3x}$, $x > 0$;

(b) the solution of $u = F(x)$ is given by $x = [-\ln(1-u)]/0.3$.

More generally,

$$E(a_1X_1 + a_2X_2 + b) = a_1E(X_1) + a_2E(X_2) + b$$

and, if X_1 and X_2 are independent

$$Var(a_1X_1 + a_2X_2 + b) = a_1^2\ Var(X_1) + a_2^2\ Var(X_2).$$

Don'ts

1. Never apply the normal approximation to the binomial

$$Z = \frac{X - np}{\sqrt{np(1-p)}}$$

when the expected number of successes (or failures) is too small. That is, when either

$$np \quad \text{or} \quad n(1-p) \text{ is 15 or less.}$$

2. Don't add variances according to

$$Var(X_1 + X_2) = Var(X_1) + Var(X_2)$$

unless the two random variables are independent or have zero covariance.

3. Don't just assume that data come from a normal distribution. When there are at least 20 to 25 observations, it is good practice to construct a normal scores plot to check this assumption.

Review Exercises

5.108 If the probability density of a random variable is given by

$$f(x) = \begin{cases} k(1-x^2) & \text{for } 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$$

find the value of k and the probabilities that a random variable having this probability density will take on a value

(a) between 0.1 and 0.2;

(b) greater than 0.5.

5.109 With reference to the preceding exercise, find the corresponding distribution function and use it to determine the probabilities that a random variable having this distribution function will take on a value

(a) less than 0.3;

(b) between 0.4 and 0.6.

5.110 In certain experiments, the error made in determining the density of a silicon compound is a random variable having the probability density

$$f(x) = \begin{cases} 25 & \text{for } -0.02 < x < 0.02 \\ 0 & \text{elsewhere} \end{cases}$$

Find the probabilities that such an error will be

(a) between -0.03 and 0.04;
(b) between -0.005 and 0.005.

5.111 Find μ and σ^2 for the probability density of Exercise 5.108.

5.112 If a random variable has the standard normal distribution, find the probability that it will take on a value

(a) between 0 and 2.50;
(b) between 1.22 and 2.35;
(c) between -1.33 and -0.33;
(d) between -1.60 and 1.80.

5.113 The burning time of an experimental rocket is a random variable having the normal distribution with $\mu = 4.76$ seconds and $\sigma = 0.04$ second. What is the probability that this kind of rocket will burn

(a) less than 4.66 seconds;
(b) more than 4.80 seconds;
(c) anywhere from 4.70 to 4.82 seconds?

5.114 Verify that

(a) $z_{0.10} = 1.28$;
(b) $z_{0.001} = 3.09$.

5.115 Referring to Exercise 5.28, find the *quartiles* of the normal distribution with $\mu = 102$ and $\sigma = 27$.

5.116 The probability density shown in Figure 5.6 is the log-normal distribution with $\alpha = 8.85$ and $\beta = 1.03$. Find the probability that

(a) the interrequest time is more than 200 microseconds;
(b) the interrequest time is less than 300 microseconds.

5.117 The probability density shown in Figure 5.8 is the exponential distribution

$$f(x) = \begin{cases} 0.25e^{-0.25x} & 0 < x \\ 0 & \text{elsewhere} \end{cases}$$

Find the probability that

(a) the time to observe a particle is more than 200 microseconds;
(b) the time to observe a particle is less than 10 microseconds.

5.118 Referring to the normal scores in Exercise 5.95, construct a normal scores plot of the suspended solids data in Exercise 2.68.

5.119 Referring to the normal scores in Exercise 5.95, construct a normal scores plot of the velocity of light data in Exercise 2.66.

5.120 If n salespeople are employed in a door-to-door selling campaign, the gross sales volume in thousands of dollars may be regarded as a random variable having the gamma distribution with $\alpha = 100\sqrt{n}$ and $\beta = \frac{1}{2}$. If the sales costs are \$5,000 per salesperson, how many salespeople should be employed to maximize the expected profit?

5.121 A mechanical engineer models the bending strength of a support beam in a transmission tower as a random variable having the Weibull distribution with $\alpha = 0.02$ and $\beta = 3.0$. What is the probability that the beam can support a load of 4.5?

5.122 Let the times to breakdown for the processors of a parallel processing machine have joint density

$$f(x, y) = \begin{cases} 0.04e^{-0.2x - 0.2y} & \text{for } x > 0,\ y > 0 \\ 0 & \text{elsewhere} \end{cases}$$

where x is the time for the first processor and y is the time for the second. Find

(a) the marginal distributions and their means;

(b) the expected value of the random variable whose values are given by $g(x, y) = x + y$.

(c) Verify in this example that the mean of a sum is the sum of the means.

5.123 Two random variables are independent and each has a binomial distribution with success probability 0.7 and 2 trials.

(a) Find the joint probability distribution.

(b) Find the probability that the second random variable is greater than the first.

5.124 If X_1 has mean -5 and variance 3 while X_2 has mean 1 and variance 4, and the two are independent, find

(a) $E(3X_1 + 5X_2 + 2)$;

(b) $Var(3X_1 + 5X_2 + 2)$.

5.125 Let $X_1, X_2, \ldots, X_{30}$ be independent and let each have the same marginal distribution with mean -5 and variance 2. Find

(a) $E(X_1 + X_2 + + X_{30})$;

(b) $Var(X_1 + X_2 + \cdots + X_{30})$.

Key Terms

(with page references)