Chapter 3

Additional Topics in Probability

Objective: In this chapter we present some special distributions, joint distributions of several random variables, functions of random variables, and some important limit theorems.

Johann Carl Friedrich Gauss

(*Source: http://tobiasamuel.files.wordpress.com/2008/06/carl_friedrich_gauss.jpg*)

German mathematician and physicist Carl Friedrich Gauss (1777–1855) is sometimes called the "prince of mathematics." He was a child prodigy. At the age of 7, Gauss started elementary school,

Mathematical Statistics with Applications

and his potential was noticed almost immediately. His teachers were amazed when Gauss summed the integers from 1 to 100 instantly. At age 24, Gauss published one of the most brilliant achievements in mathematics, *Disquisitiones Arithmeticae* (1801). In it, Gauss systematized the study of number theory. Gauss applied many of his mathematical insights in the field of astronomy, and by using the method of least squares he successfully predicted the location of the asteroid Ceres in 1801. In 1820 Gauss made important inventions and discoveries in geodesy, the study of the shape and size of the earth. In statistics, he developed the idea of the normal distribution. In the 1830s he developed theories of non-Euclidean geometry and mathematical techniques for studying the physics of fluids. Although Gauss made many contributions to applied science, especially electricity and magnetism, pure mathematics was his first love. It was Gauss who first called mathematics "the queen of the sciences."

3.1 INTRODUCTION

In the previous chapter, we looked at the basic concepts of probability calculations, random variables, and their distributions. There are many special distributions that have useful applications in statistics. It is worth knowing the type of distribution that we can expect under different circumstances, because a better knowledge of the population will result in better inferential results. In the next section, we discuss some of these distributions with some additional distributions presented in Appendix A3. We also briefly deal with joint distributions of random variables and functions of random variables. Limit theorems play an important role in statistics. We will present two limit theorems: the law of large numbers and the Central Limit Theorem.

3.2 SPECIAL DISTRIBUTION FUNCTIONS

Random variables are often classified according to their probability distribution functions. In any analysis of quantitative data, it is a major step to know the form of the underlying probability distributions. There are certain basic probability distributions that are applicable in many diverse contexts and thus repeatedly arise in practice. A great variety of special distributions have been studied over the years. Also, new ones are frequently being added to the literature. It is impossible to give a comprehensive list of distribution functions in this book. There are many books and Web sites that deal with a range of distribution functions. A good list of distributions can be obtained from http://www.causascientia.org/math_stat/Dists/Compendium.pdf. In this section, we will describe some of the commonly used probability distributions. In Appendix A3, we list some more distributions with their mean, variance, and moment-generating functions. First we discuss some discrete probability distributions.

3.2.1 The Binomial Probability Distribution

The simplest distribution is the one with only two possible outcomes. For example, when a coin (not necessarily fair) is tossed, the outcomes are heads or tails, with each outcome occurring with some positive probability. These two possible outcomes may be referred to as "success" if heads occurs and "failure" if tails occurs. Assume that the probability of heads appearing in a single toss is p; then the probability of tails is $1 - p = q$. We define a random variable X associated with this experiment as taking value *1* with probability p if heads occurs and value *0* if tails occurs with probability q. Such a random variable X is said to have a *Bernoulli probability distribution*. That is, X is a Bernoulli random variable if for some p, $0 \leq p \leq 1$, the probability $P(X = 1) = p$ and $P(X = 0) = 1 - p$. The probability function of a Bernoulli random variable X can be expressed as

$$p(x) = P(X = x) = \begin{cases} p^x(1-p)^{1-x}, & x = 0, 1 \\ 0, & \text{otherwise.} \end{cases}$$

Note that this distribution is characterized by the single parameter p. It can be easily verified that the mean and variance of X are $E[X] = p$, $\text{var}(X) = pq$, respectively, and the moment-generating function is $M_X(t) = pe^t + (1 - p)$.

Even when the experimental values are not dichotomous, reclassifying the variable as a Bernoulli variable can be helpful. For example, consider blood pressure measurements. Instead of representing the numerical values of blood pressure, if we reclassify the blood pressure as "high blood pressure" and "low blood pressure," we may be able to avoid dealing with a possible misclassification due to diurnal variation, stress, and so forth, and concentrate on the main issue, which would be: Is the average blood pressure unusually high?

In a succession of Bernoulli trials, one is more interested in the total number of successes (whenever a 1 occurs in a Bernoulli trial, we term it a "success"). The probability of observing exactly k successes in n independent Bernoulli trials yields the binomial probability distribution. In practice, the binomial probability distribution is used when we are concerned with the occurrence of an event, not its magnitude. For example, in a clinical trial, we may be more interested in the number of survivors after a treatment.

Definition 3.2.1 *A* **binomial experiment** *is one that has the following properties: (1) The experiment consists of n identical trials. (2) Each trial results in one of the two outcomes, called a success S and failure F. (3) The probability of success on a single trial is equal to p and remains the same from trial to trial. The probability of failure is* $1 - p = q$. *(4) The outcomes of the trials are independent. (5) The random variable X is the number of successes in n trials.*

Earlier we have seen that the number of ways of obtaining x successes in n trials is given by

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}.$$

Definition 3.2.2 *A random variable X is said to have* **binomial probability distribution** *with parameters* (n, p) *if and only if*

$$\begin{aligned} P(X = x) = p(x) &= \binom{n}{x} p^x q^{n-x} \\ &= \begin{cases} \frac{n!}{x!(n-x)!} p^x q^{n-x}, & x = 0, 1, 2, \ldots, n, \ 0 \leq p \leq 1, \text{ and } q = 1 - p \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

To show the dependence on n and p, denote $p(x)$ by $b(x, n, p)$ and the cumulative probabilities by

$$B(x, n, p) = \sum_{i=0}^{x} b(i, n, p)$$

Binomial probabilities are tabulated in the binomial table.

By the binomial theorem, we have

$$(p+q)^n = \sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x}.$$

Because $(p + q) = 1$, we conclude that $\sum_{i=0}^{x} b(i, n, p) = \sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x} = 1^n = 1$, for all $n \geq 1$ and $0 \leq p \leq 1$. Hence, $p(x)$ is indeed a probability function. The binomial probability distribution is characterized by two parameters, the number of independent trials n and the probability of success p.

Example 3.2.1

It is known that screws produced by a certain machine will be defective with probability 0.01 independently of each other. If we randomly pick 10 screws produced by this machine, what is the probability that at least two screws will be defective?

Solution

Let X be the number of defective screws out of 10. Then X can be considered as a binomial r.v. with parameters $(10, 0.01)$. Hence, using the binomial pf $p(x)$, given in Definition 3.2.2, we obtain

$$\begin{aligned} P(X \geq 2) &= \sum_{x=2}^{10} \binom{10}{x} (0.01)^x (0.99)^{10-x} \\ &= 1 - [P(X = 0) + P(X = 1)] = 0.004. \end{aligned}$$

In Chapter 2, we saw Mendel's law. In biology, the result "gene frequencies and genotype ratios in a randomly breeding population remain constant from generation to generation" is known as the *Hardy–Weinberg law*.

Example 3.2.2

Suppose we know that the frequency of a dominant gene, A, in a population is equal to 0.2. If we randomly select eight members of this population, what is the probability that at least six of them will display the dominant phenotype? Assume that the population is sufficiently large that removing eight individuals will not affect the frequency and that the population is in Hardy–Weinberg equilibrium.

Solution

First of all, note that an individual can have the dominant gene, A, if the person has traits AA, aA, or Aa. Hence, if the gene frequency is 0.2, the probability that an individual is of genotype A is

$$\begin{aligned} P(A) &= P(AA \cup Aa \cup aA) = P(AA) + 2P(Aa) \\ &= (0.2)^2 + 2(0.2)(0.8) = 0.36. \end{aligned}$$

Let X denote the number of individuals out of eight that display the dominant phenotype. Then X is binomial with $n = 8$, and $p = 0.36$. Thus, the probability that at least six of them will display the dominant phenotype is

$$\begin{aligned} P(X \geq 6) &= P(X = 6) + P(X = 7) + P(X = 8) \\ &= \sum_{i=6}^{8} \binom{10}{i} (0.36)^i (0.64)^{10-i} = 0.029259. \end{aligned}$$

For large n, calculation of binomial probabilities is tedious. Many statistical software packages have binomial probability distribution commands. For the purpose of this book, we will use the binomial table that gives the cumulative probabilities $B(x, n, p)$ for $n=2$ through $n=20$ and $p=0.05, 0.10, 0.15, \ldots, 0.90, 0.95$. If we need the probability of a single term, we can use the relation

$$P(X = x) = b(x, n, p) = B(x, n, p) - B(x - 1, n, p).$$

Example 3.2.3

A manufacturer of inkjet printers claim that only 5% of their printers require repairs within the first year. If of a random sample of 18 of the printers, four required repairs within the first year, does this tend to refute or support the manufacturer's claim?

Solution

Let us assume that the manufacturer's claim is correct; that is, the probability that a printer will require repairs within the first year is 0.05. Suppose 18 printers are chosen at random. Let p be the probability that any one of the printers will require repairs within the first year. We now find the probability that at least four of these out of the 18 will require repairs during the first year. Let X represent the number of printers that require repair within the first year. Then X follows the binomial pmf with $p = 0.05, n = 18$. The probability that four or more of the 18 will require repair within the first year is given by

$$P(X \geq 4) = \sum_{x=4}^{18} \binom{18}{x} (0.05)^x (0.95)^{18-x}$$

or, using the binomial table,

$$\sum_{x=4}^{18} b(x, 18, 0.05) = 1 - B(3, 18, 0.05)$$
$$= 1 - 0.9891$$
$$= 0.0109.$$

This value (approximately 1.1%) is very small. We have shown that if the manufacturer's claim is correct, then the chances of observing four or more bad printers out of 18 are very small. But we did observe exactly four bad ones. Therefore we must conclude that the manufacturer's claim cannot be substantiated.

MEAN, VARIANCE, AND MGF OF A BINOMIAL RANDOM VARIABLE

Theorem 3.2.1 *If X is a binomial random variable with parameters n and p, then*

$$E(X) = \mu = np$$
$$Var(X) = \sigma^2 = np(1 - p).$$

Also the moment-generating function

$$M_X(t) = \left[pe^t + (1 - p)\right]^n.$$

Proof. We derive the mean and the variance. The derivation for mgf is given in Example 2.6.5. Using the binomial pmf, $p(x) = (n!/(x!(n-x)!))p^x q^{n-x}$, and the definition of expectation, we have

$$\mu = E(X) = \sum_{x=0}^{n} xp(x) = \sum_{x=0}^{n} x\frac{n!}{x!(n-x)!}p^x(1-p)^{n-x}$$
$$= \sum_{x=1}^{n} \frac{n!}{(x-1)!(n-x)!}p^x(1-p)^{n-x},$$

since the first term in the sum is zero, as $x = 0$.

Let $i = x - 1$. When x varies from 1 through n, $i = (x - 1)$ varies from zero through $(n - 1)$. Hence,

$$\mu = \sum_{i=0}^{n-1} \frac{n!}{i!(n-i-1)!}p^{i+1}(1-p)^{n-i-1}$$
$$= np\sum_{i=0}^{n-1} \frac{(n-1)!}{i!(n-1-i)!}p^i(1-p)^{n-1-i}$$
$$= np,$$

because the last summand is that of a binomial pmf with parameter $(n-1)$ and p, hence, equals 1.

To find the variance, we first calculate $E[X(X-1)]$.

$$E[X(X-1)] = \sum_{x=0}^{n} x(x-1)\frac{n!}{x!(n-x)!}p^x(1-p)^{n-x}$$

$$= \sum_{x=2}^{n} \frac{n!}{(x-2)!(n-x)!}p^x(1-p)^{n-x},$$

because the first two terms are zero. Let $i = x-2$. Then,

$$E[X(X-1)] = \sum_{i=0}^{n-2} \frac{n!}{i!(n-i-2)!}p^{i+2}(1-p)^{n-i-2}$$

$$= n(n-1)p^2 \sum_{i=0}^{n-2} \frac{(n-2)!}{i!(n-2-i)!}p^i(1-p)^n$$

$$= n(n-1)p^2,$$

because the last summand is that of a binomial pf with parameter $(n-2)$ and p thus equals 1.

Note that $E(X(X-1)) = EX^2 - E(X)$, and so we obtain

$$\sigma^2 = Var(X) = E(X^2) - [E(X)]^2$$

$$= E[X(X-1)] + E(X) - [E(X)]^2$$

$$= n(n-1)p^2 + np - (np)^2 = -np^2 + np$$

$$= np(1-p).$$

□

3.2.2 **Poisson Probability Distribution**

The Poisson probability distribution was introduced by the French mathematician Siméon-Denis Poisson in his book published in 1837, which was entitled *Recherches sur la probabilité des jugements en matières criminelles et matière civile* and dealt with the applications of probability theory to lawsuits, criminal trials, and the like. Consider a statistical experiment of which A is an event of interest. A random variable that counts the number of occurrences of A is called a *counting random variable*. The Poisson random variable is an example of a counting random variable. Here we assume that the numbers of occurrences in disjoint intervals are independent and the mean of the number occurrences is constant.

Definition 3.2.3 *A discrete random variable X is said to follow the* **Poisson probability distribution** *with parameter* $\lambda > 0$, *denoted by Poisson*(λ), *if*

$$P(X = x) = f(x, \lambda) = f(x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots$$

The Poisson probability distribution is characterized by the single parameter, λ, which represents the mean of a Poisson probability distribution. Thus, in order to specify the Poisson distribution, we only need to know the mean number of occurrences. This distribution is of fundamental theoretical and practical importance. Rare events are modeled by the Poisson distribution. For example, the Poisson probability distribution has been used in the study of telephone systems. The number of incoming calls into a telephone exchange during a unit time might be modeled by a Poisson variable assuming that the exchange services a large number of customers who call more or less independently. Some other problems where Poisson representation can be used are the number of misprints in a book, radioactivity counts per unit time, the number of plankton (microscopic plant or animal organisms that float in bodies of water) per aliquot of seawater, or count of bacterial colonies per petri plate in a microbiological study. In stem cell research, the Poisson distribution is used to analyze the redundancy of clusters in the stem cell database. A Poisson probability distribution has the unique property that its mean equals its variance.

MEAN, VARIANCE, AND MOMENT–GENERATING FUNCTION OF A POISSON RANDOM VARIABLE

Theorem 3.2.2 *If X is a Poisson random variable with parameter* λ, *then*

$$E(X) = \lambda$$
$$Var(X) = \lambda.$$

Also the moment-generating function is

$$M_X(t) = e^{\lambda(e^t - 1)}.$$

The proof of this result is similar to that we used in Theorem 3.2.1 in this section. One needs to use the Maclaurin's expansion, $e^\lambda = \sum_{i=0}^{\infty}(\lambda^i/i!)$.

Example 3.2.4

Let X be a Poisson random variable with $\lambda = 1/2$. Find

(a) $P(X = 0)$

(b) $P(X \geq 3)$

Solution

(a) *We have*

$$P(X = 0) = p(0) = \frac{e^{-1/2}(1/2)^0}{0!} = e^{-1/2} = 0.60653.$$

(b) *Here we will use complementary event to compute the required probability. That is,*

$$\begin{aligned}P(X \geq 3) &= 1 - P(X \leq 2) = 1 - [p(0) + p(1) + p(2)] \\ &= 1 - \left[e^{-1/2} + \frac{e^{-1/2}(1/2)}{1!} + \frac{e^{-1/2}(1/2)^2}{2!}\right] \\ &= 1 - 0.98561 = 0.01439.\end{aligned}$$

When n is large and p small, binomial probabilities are often approximated by Poisson probabilities. In these situations, where performing the factorial and exponential operations required for direct calculation of binomial probabilities is a lengthy and tedious process and tables are not available, the Poisson approximation is more feasible. The following theorem states this result.

POISSON APPROXIMATION TO THE BINOMIAL PROBABILITY DISTRIBUTION

Theorem 3.2.3 *If X is a binomial r.v. with parameters n and p, then for each value $x = 0, 1, 2, \ldots$ and as $p \to 0, n \to \infty$ with $np = \lambda$ constant,*

$$\lim_{n\to\infty} \binom{n}{x} p^x (1-p)^{n-x} = \frac{e^{-\lambda}\lambda^x}{x!}.$$

The proof of this result is similar to that we used in Theorem 3.2.1. In the present context, the Poisson probability distribution is sometimes referred to as "the distribution of rare events" because of the fact that p is quite small when n is large. Usually, if $p \leq 0.1$ and $n \geq 40$ we could use the Poisson approximation in practice. In general, another rule of thumb is to use Poisson approximation to binomial in the case of $np < 5$.

Example 3.2.5

If the probability that an individual suffers an adverse reaction from a particular drug is known to be 0.001, determine the probability that out of 2000 individuals, (a) exactly three and (b) more than two individuals will suffer an adverse reaction.

Solution

Let Y be the number of individuals who suffer an adverse reaction. Then Y is binomial with $n = 2000$ and $p = 0.001$. Because n is large and p is small, we can use the Poisson approximation with $\lambda = np = 2$.

(a) *The probability that exactly three individuals will suffer an adverse reaction is*

$$P(Y = 3) = \frac{2^3 e^{-2}}{3!} = 0.18.$$

That is, there is approximately an 18% chance that exactly three individuals of 2000 will suffer an adverse reaction.

(b) *The probability that more than two individuals will suffer an adverse reaction is*

$$P(Y > 2) = 1 - P(Y = 0) - P(Y = 1) - P(Y = 2)$$
$$= 1 - 5e^{-2} = 0.323.$$

Similarly, there is approximately a 32.3% chance that more than two individuals will have an adverse reaction.

■

Now we will discuss some continuous distributions. As mentioned earlier, if X is a continuous random variable with pdf $f(x)$, then

$$P(a \le X \le b) = \int_a^b f(x)dx.$$

3.2.3 **Uniform Probability Distribution**

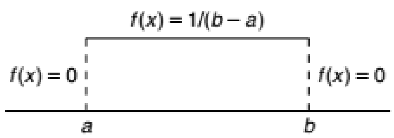
The uniform probability distribution is used to generate random numbers from other distributions and also is useful as a "first guess" if no other information about a random variable X is known, other than that it is between a and b. Also, in real-world problems that have uniform behavior in a given interval, we can characterize the probabilistic behavior of such a phenomenon by the uniform distribution. (See Figure 3.1.)

Definition 3.2.4 *A random variable X is said to have a* **uniform probability distribution** *on (a, b), denoted by $U(a, b)$, if the density function of X is given by*

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & a \le x \le b, \\ 0, & \text{otherwise.} \end{cases}$$

The cumulative distribution function is given by

$$F(x) = \int_{-\infty}^{x} \frac{1}{b-a} dx = \begin{cases} 0, & x < a \\ \dfrac{x-a}{b-a}, & a \le x < b \\ 1, & x \ge b. \end{cases}$$



■ **FIGURE 3.1** Uniform probability density.

Example 3.2.6

If X is a uniformly distributed random variable over $(0, 10)$, calculate the probability that (a) $X < 3$, (b) $X > 6$, and (c) $3 < X < 8$.

Solution

(a)

$$P(X < 3) = \int_0^3 \frac{1}{10} dx = \frac{3}{10}.$$

(b)

$$P(X > 6) = \int_6^{10} \frac{1}{10} dx = \frac{4}{10}.$$

(c)

$$P(3 < X < 8) = \int_3^8 \frac{1}{10} dx = \frac{1}{2}.$$

MEAN, VARIANCE, AND MOMENT–GENERATING FUNCTION OF A UNIFORM RANDOM VARIABLE

Theorem 3.2.4 *If X is a uniformly distributed random variable on (a, b), then*

$$E(X) = \frac{a+b}{2}.$$

and

$$Var(X) = \frac{(b-a)^2}{12}.$$

Also, the moment-generating function is

$$M_X(t) = \begin{cases} \dfrac{e^{tb} - e^{ta}}{t(b-a)}, & t \neq 0 \\ 1, & t = 0. \end{cases}$$

Proof. We will obtain the mean and the variance and leave the derivation of the moment-generating function as an exercise. By definition we have

$$\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} x \frac{1}{b-a} dx \\
&= \int_{a}^{b} x \frac{1}{b-a} dx = \frac{1}{b-a} \left(\frac{x^2}{2} \bigg|_a^b \right) \\
&= \frac{a+b}{2}.
\end{aligned}$$

Also

$$\begin{aligned}
E(X^2) &= \int_{a}^{b} x^2 \frac{1}{b-a} dx = \frac{1}{b-a} \left(\frac{x^3}{3} \bigg|_a^b \right) \\
&= \frac{1}{3} \frac{b^3 - a^3}{b-a} \\
&= \frac{1}{3}(b^2 + ab + a^2) \text{ as } b^3 - a^3 = (b-a)(b^2 + ab + a^2).
\end{aligned}$$

Thus,

$$\begin{aligned}
Var(X) &= E(X^2) - (E(X))^2 \\
&= \frac{1}{3}(b^2 + ab + a^2) - \frac{(a+b)^2}{4} \\
&= \frac{1}{12}(b-a)^2.
\end{aligned}$$

□

Example 3.2.7

The melting point, X, of a certain solid may be assumed to be a continuous random variable that is uniformly distributed between the temperatures 100°C and 120°C. Find the probability that such a solid will melt between 112°C and 115°C.

Solution

The probability density function is given by

$$f(x) = \begin{cases} \frac{1}{20}, & 100 \le x \le 120 \\ 0 & \text{otherwise.} \end{cases}$$

Hence,

$$P(112 \leq X \leq 115) = \int_{112}^{115} \frac{1}{20} dx = \frac{3}{20} = 0.15.$$

Thus, there is a 15% chance of this solid melting between $112°C$ *and* $115°C$.

3.2.4 **Normal Probability Distribution**

The single most important distribution in probability and statistics is the normal probability distribution. The density function of a normal probability distribution is bell shaped and symmetric about the mean. The normal probability distribution was introduced by the French mathematician Abraham de Moivre in 1733. He used it to approximate probabilities associated with binomial random variables when n is large. This was later extended by Laplace to the so-called Central Limit Theorem, which is one of the most important results in probability. Carl Friedrich Gauss in 1809 used the normal distribution to solve the important statistical problem of combining observations. Because Gauss played such a prominent role in determining the usefulness of the normal probability distribution, the normal probability distribution is often called the *Gaussian distribution*. Gauss and Laplace noticed that measurement errors tend to follow a bell-shaped curve, a normal probability distribution. Today, the normal probability distribution arises repeatedly in diverse areas of applications. For example, in biology, it has been observed that the normal probability distribution fits data on the heights and weights of human and animal populations, among others.

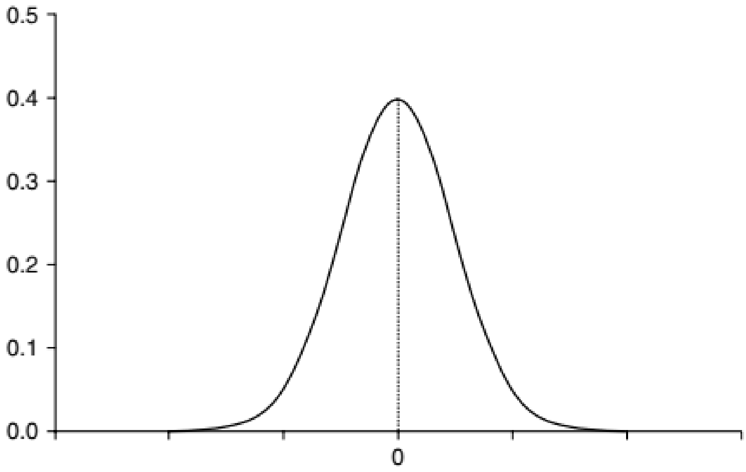
We should also mention here that almost all basic statistical inference is based on the normal probability distribution. The question that often arises is, when do we know that our data follow the normal distribution? To answer this question we have specific statistical procedures that we study in later chapters, but at this point we can obtain some constructive indications of whether the data follows the normal distribution by using descriptive statistics. That is, if the histogram of our data can be capped with a bell-shaped curve (Figure 3.2), if the stem-and-leaf diagram is fairly symmetrical with respect to its center, and/or by invoking the empirical rule "backwards," we can obtain a good indication whether our data follow the normal probability distribution.

Definition 3.2.5 *A random variable X is said to have a* **normal probability distribution** *with parameters* μ *and* σ^2, *if it has a probability density function given by*

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}, \; -\infty < x < \infty, -\infty < \mu < \infty, \sigma > 0.$$

If $\mu = 0$, *and* $\sigma = 1$, *we call it standard normal random variable.*

For any normal random variable with mean μ and variance σ^2, we use the notation $X \sim N(\mu, \sigma^2)$. When a random variable X has a standard normal probability distribution, we will write $X \sim N(0, 1)$ (X is a normal with mean 0 and variance 1). Probabilities for a standard normal probability distribution are given in the normal table.



■ **FIGURE 3.2** Standard normal density function.

MEAN, VARIANCE, AND MGF OF A NORMAL RANDOM VARIABLE

Theorem 3.2.5 *If $X \sim N(\mu, \sigma^2)$, then $E(X) = \mu$ and $Var(X) = \sigma^2$. Also the moment-generating function is*

$$M_X(t) = e^{t\mu + \frac{1}{2}t^2\sigma^2}.$$

If $X \sim N(\mu, \sigma^2)$, then the *z-transform* (or *z-score*) of X, $Z = \frac{X-\mu}{\sigma}$, is an $N(0, 1)$ random variable. This fact will be used in calculating probabilities for normal random variables.

Example 3.2.8

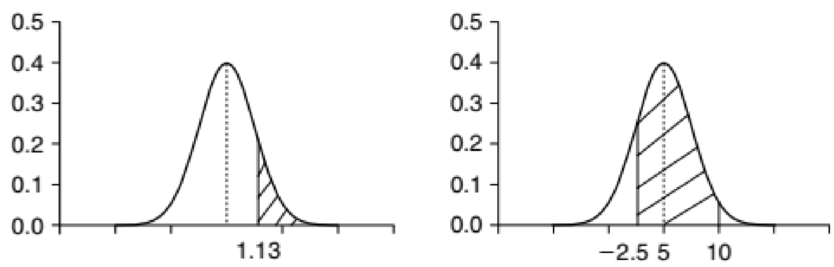
(a) For $X \sim N(0, 1)$, calculate $P(Z \geq 1.13)$.
(b) For $X \sim N(5, 4)$, calculate $P(-2.5 < X < 10)$.

Solution

(a) *Using the normal table,*

$$P(Z \geq 1.13) = 1 - 0.8708 = 0.1292.$$

The shaded part in the graph represents the $P(Z \geq 1.13)$.



(b) *Using the z-transform, we have*

$$\begin{aligned} P(-2.5 < X < 10) &= P\left(\frac{-2.5-5}{2} < Z < \frac{10-5}{2}\right) \\ &= P(-3.75 < Z < 2.5) \\ &= P(-3.75 < Z < 0) + P(0 < Z < 2.5) \\ &= 0.9938. \end{aligned}$$

In the following example, we will show how to find the z values when the probabilities are given.

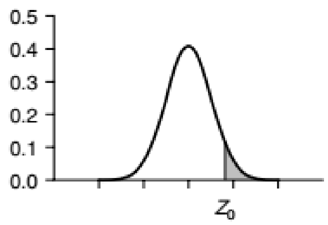
Example 3.2.9

For a standard normal random variable Z, find the value of z_0 such that

(a) $P(Z > z_0) = 0.25$.
(b) $P(Z < z_0) = 0.95$.
(c) $P(Z < z_0) = 0.12$.
(d) $P(Z > z_0) = 0.68$.

Solution

(a) *From the normal table, and using the fact that the shaded area in the figure is 0.25, we obtain* $z_0 \approx 0.675$.
(b) *Because* $P(Z < z_0) = 1 - P(Z \geq z_0) = 0.95 = 0.5 + 0.45$. *This implies,* $P(Z > z_0) = 0.05$. *From the normal table,* $z_0 = 1.645$.



(c) *From the normal table,* $z_0 = -1.175$.
(d) *Using the normal table, we have* $P(Z > z_0) = 0.5 + P(0 < Z < z_0) = 0.68$.
This implies, $P(Z \leq z_0) = 0.32$. *From the normal table,* $z_0 = -0.465$.

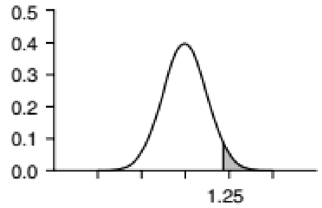
Example 3.2.10

The scores of an examination are assumed to be normally distributed with $\mu = 75$ and $\sigma^2 = 64$. What is the probability that a score chosen at random will be greater than 85?

Solution

Let X be a randomly chosen score from the exam scores. Then, $X \sim N(75, 64)$.

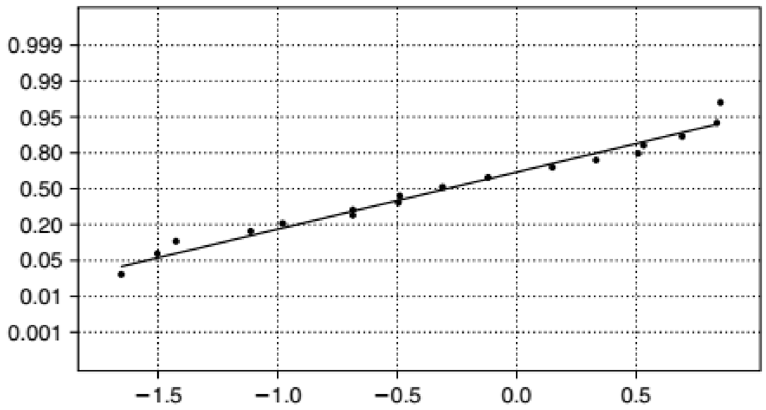
$$P(X > 85) = P\left(\frac{X-75}{8} > \frac{85-75}{8} = 1.25\right)$$
$$= P(Z > 1.25) = 0.1056.$$



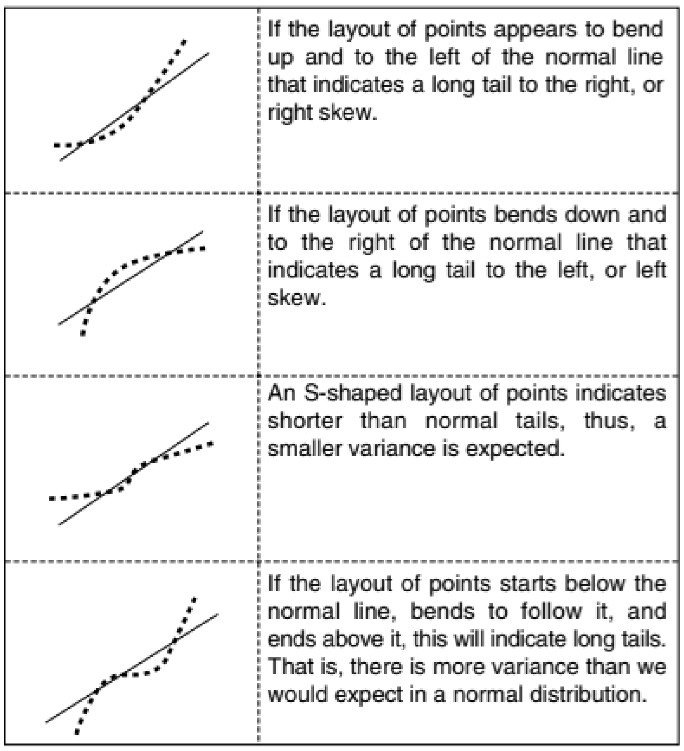
Thus, there is about a 10.56% chance that the score will be greater than 85.

In practice, whenever a large number of small effects are present and *acting additively*, it is reasonable to assume that observations will be normal. When the number of data is small, it is risky to assume a normal distribution without a proper testing. Apart from histogram, box-plot, and stem-and-leaf-displays, one of the most useful tools for assessing normality is a quantile quantile or QQ plot. This is a scatterplot with the quantiles of the scores on the horizontal axis and the expected normal scores on the vertical axis. The expected normal scores are calculated by taking the z-scores of $(r_i - 0.5)/n$, where r_i is the rank ith observation in increasing order. The steps in constructing a QQ plot are as follows: First, we sort the data in an ascending order. If the plot of these scores against the expected normal scores is a straight line, then the data can be considered normal. Any curvature of the points indicates departures from normality. This procedure obtaining a normal plot (QQ plot is similar to normal plot for a normal distribution) is described in Project 4C. Figure 3.3 shows a normal probability plot generated by Minitab.

If plotted points do not fit the line well, but bend away from it in places, the distribution may be nonnormal. The shapes in Figure 3.4 will give some indication of the distribution of the data.



■ **FIGURE 3.3** Normal probability plot.



■ **FIGURE 3.4** Shapes indicating distribution of the data.

Almost all of the statistical software packages include a procedure for obtaining the graph of a normal probability plot that can be used to test the normality of a data. A discussion of how to do this is given in Section 14.4. Errors in the measurements can also act in a *multiplicative* (rather than additive) manner. In that case, the assumption of normality is not justified.

A closely related distribution to normal distribution is the log-normal distribution. A variable might be modeled as log-normal if it can be thought of as the multiplicative effect of many small independent factors. This distribution arises in physical problems when the domain of the variate, X, is greater than zero and its histogram is markedly skewed. If a random variable Y is normally distributed, then $\exp(Y)$ has a *log-normal distribution*. Thus, the natural logarithm of a log-normally distributed variable is normally distributed. That is, if X is a random variable with log-normal distribution, then $\ln(X)$ is normally distributed. Most biological evidence suggests that the growth processes of living tissue proceed by multiplicative, not additive, increments. Thus, the measures of body size should at most follow a log-normal rather than normal distribution. Also, the sizes of plants and animals is approximately log-normal. The log-normal distribution is also useful in modeling of claim sizes in the insurance industry.

The probability density function of a log-normal random variable, X, is given as

$$f(x) = \begin{cases} \dfrac{1}{x\sigma_y\sqrt{2\pi}} e^{-(\ln x - \mu_y)^2/2\sigma_y^2}, & x > 0, \sigma_y > 0, \quad -\infty < \mu_y < \infty \\ 0, & \text{otherwise.} \end{cases}$$

where μ_y and σ_y are the mean and standard deviation of $Y = \ln(X)$. These parameters are related to the parameters of the random variable X as follows:

$$\mu_y = \ln\left(\sqrt{\frac{\mu_x^4}{\mu_x^2 + \sigma_x^2}}\right), \quad \sigma_y = \ln\left(\sqrt{\frac{\mu_x^2 + \sigma_x^2}{\mu_x^2}}\right).$$

We can verify that the expected value X is

$$E(X) = e^{\mu_y + (\sigma_y^2/2)}$$

and the variance is

$$Var(X) = (e^{\sigma_y^2} - 1)e^{2\mu_y + \sigma_y^2}.$$

The question of when the log-normal distribution is applicable in a given physical problem after a certain amount of data has been obtained can be answered by creating a normal probability plot of $\ln(X)$ and testing for normality. Thus, if the natural logarithms of the data show normality, log-normal distribution may be more appropriate.

If X is log-normally distributed with parameters μ_y and σ_y, and $0 < a < b$, then with $Y = \ln(X)$

$$\begin{aligned} P(a \le X \le b) &= P(\ln a \le Y \le \ln b) \\ &= P\left(\frac{\ln a - \mu_y}{\sigma_y} \le \frac{Y - \mu_y}{\sigma_y} \le \frac{\ln b - \mu_y}{\sigma_y}\right) \\ &= P(a' \le Z \le b'), \end{aligned}$$

where $Z \sim N(0, 1)$. This probability can be obtained from the standard normal table.

Example 3.2.11

In an effort to establish a suitable height for the controls of a moving vehicle, information was gathered about X, the amounts by which the heights of the operators vary from 60 inches, which is the minimum height. It was verified that the data that were collected followed the log-normal distribution by normal probability plot of $Y = \ln X$. Assume that $\mu_x = 6\,in.$ and $\sigma_x = 2\,in.$

(a) What percentage of operators would have a height less than 65.5 in.?

(b) If an operator is chosen at random, what is the probability that his or her height will be between 64 and 66 in.?

Solution

(a) *Here,* $X = 65.5 - 60 = 5.5$. *Also,*

$$\mu_y = \ln\left(\sqrt{\frac{\mu_x^4}{\mu_x^2 + \sigma_x^2}}\right) = \ln\sqrt{\frac{6^4}{6^2 + 2^2}} = 1.74,$$

$$\sigma_y = \ln\left(\sqrt{\frac{\mu_x^2+\sigma_x^2}{\mu_x^2}}\right) = \ln\sqrt{\frac{6^2+2^2}{6^2}} = 0.053.$$

Thus,

$$P(X \le 5.5) = P(Y \le \ln 5.5) = P\left(Z \le \frac{(\ln 5.5) - 1.74}{0.053}\right)$$
$$= P(Z \le -0.67) = 0.2514.$$

Hence, about 25.14% of the heights of the operators vary from 60 inches.

(b) *Similar to part (a), we get*

$$P(4 \le X \le 6) = P(\ln 4 \le Y \le \ln 6)$$
$$= P\left(\frac{(\ln 4) - 1.74}{0.053} \le Z \le \frac{(\ln 6) - 1.74}{0.053}\right)$$
$$= P(-6.67 \le Z \le 0.98) = 0.8365.$$

■

3.2.5 Gamma Probability Distribution

The gamma probability distribution has found applications in various fields. For example, in engineering, the gamma probability distribution has been employed in the study of system reliability. We describe the gamma function before we introduce the gamma probability distribution. The *gamma function*, denoted by $\Gamma(a)$, is defined as

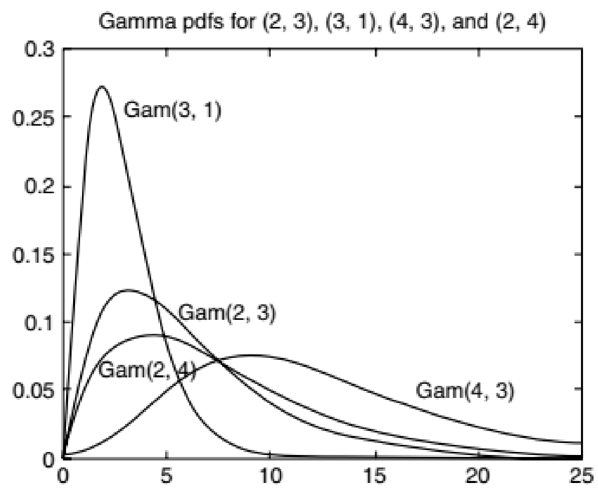
$$\Gamma(a) = \int_0^\infty e^{-x}x^{a-1}dx, a > 0.$$

It can be shown using the integration by parts that for $a > 1$, $\Gamma(a) = (a-1)\Gamma(a-1)$. In particular, if n is a positive integer, $\Gamma(n) = (n-1)!$.

Definition 3.2.6 *A random variable X is said to possess a* **gamma probability distribution** *with parameters $\alpha > 0$ and $\beta > 0$ if it has the pdf given by*

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

The gamma density has two parameters, α and β. We denote this by $Gamma(\alpha, \beta)$. The parameter α is called a *shape parameter*, and β is called a *scale parameter*. Changing α changes the shape of the density, whereas varying β corresponds to changing the units of measurement (such as changing from seconds to minutes). Varying these two parameters will generate different members of the gamma family. If we take α to be a positive integer, we get a special case of gamma probability distribution, known as the *Erlang distribution*. This is used extensively in queuing theory to model waiting times. Figure 3.5 gives an indication of how α and β influence the shape and scale of $f(x)$.



■ **FIGURE 3.5** Gamma pdfs for different degrees of freedom.

MEAN, VARIANCE, AND MGF OF A GAMMA RANDOM VARIABLE

Theorem 3.2.6 *If X is a gamma random variable with parameters $\alpha > 0$ and $\beta > 0$, then*

$$E(X) = \alpha\beta \quad and \quad Var(X) = \alpha\beta^2.$$

Also, the moment-generating function is

$$M_X(t) = \frac{1}{(1-\beta t)^{\alpha}}, \quad t < \frac{1}{\beta}.$$

Example 3.2.12

The daily consumption of aviation fuel in millions of gallons at a certain airport can be treated as a gamma random variable with $\alpha = 3, \beta = 1$.

(a) What is the probability that on a given day the fuel consumption will be less than 1 million gallons?

(b) Suppose the airport can store only 2 million gallons of fuel. What is the probability that the fuel supply will be inadequate on a given day?

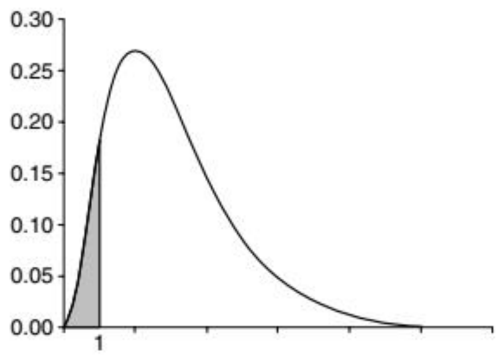
Solution

(a) *Let X be the fuel consumption in millions of gallons on a given day at a certain airport. Then, $X \sim \Gamma(\alpha = 3, \beta = 1)$ and*

$$f(x) = \frac{1}{\Gamma(3)(1^3)} x^{3-1} e^{-x} = \frac{1}{2} x^2 e^{-x}, \quad x > 0.$$

Hence, using integration by parts, we obtain

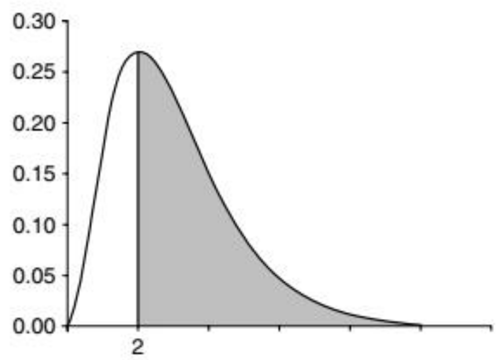
$$P(X < 1) = \frac{1}{2}\int_0^1 x^2 e^{-x} dx = 1 - \frac{5}{2e} = 0.08025.$$



Thus, there is about an 8% chance that on a given day the fuel consumption will be less than 1 million gallons.

(b) *Because the airport can store only 2 million gallons, the fuel supply will be inadequate if the fuel consumption X is greater than 2. Thus,*

$$P(X > 2) = \frac{1}{2}\int_{2}^{\infty} x^2 e^{-x} dx = 0.677.$$



We can conclude that there is about a 67.7% chance that the fuel supply of 2 million gallons will be inadequate on a given day. So, if the model is right, the airport needs to store more than 2 million gallons of fuel.

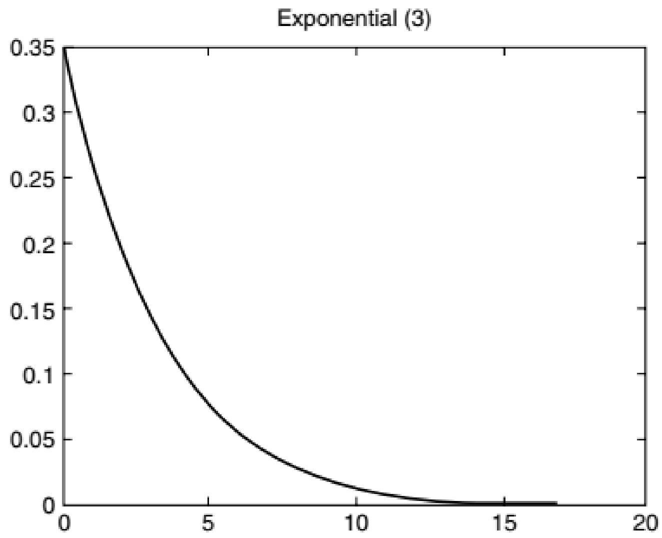
■

We now describe two special cases of gamma probability distribution. In the pdf of the gamma, we let $\alpha = 1$, we get the pdf of an exponential random variable.

Definition 3.2.7 *A random variable X is said to have an* **exponential probability distribution** *with parameter β if the pdf of X is given by*

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

Exponential random variables are often used to model the lifetimes of electronic components such as fuses, for survival analysis, and for reliability analysis, among others. The exponential distribution (Figure 3.6) is also used in developing models of insurance risks.



■ **FIGURE 3.6** Probability density function for exponential r.v.

MEAN, VARIANCE, AND MGF OF AN EXPONENTIAL RANDOM VARIABLE

Theorem 3.2.7 *If X is an exponential random variable with parameters $\beta > 0$, then*

$$E(X) = \beta \quad and \quad Var(X) = \beta^2.$$

Also the moment-generating function is

$$M_X(t) = \frac{1}{(1-\beta t)}, \quad t < \frac{1}{\beta}.$$

Example 3.2.13

The time, in hours, during which an electrical generator is operational is a random variable that follows the exponential distribution with $\beta = 160$. What is the probability that a generator of this type will be operational for

(a) Less than 40 hours?

(b) Between 60 and 160 hours?

(c) More than 200 hours?

Solution

Let X denote the random variable corresponding to time (in hours) during which the generator is operational. Then the density function of X is given by

$$f(x) = \begin{cases} \frac{1}{160} e^{-\left(\frac{x}{160}\right)}, & x \geq 0 \\ 0, & \text{otherwise.} \end{cases}$$

Thus, we have the following:

(a) $P(X \le 40) = \int_0^{40} \frac{1}{160} e^{-(x/160)} dx = 0.22119$. *There is about a 22.1% chance that a generator of this type will be operational for less than 40 hours.*

(b) $P(60 \le X \le 160) = \int_{60}^{160} \frac{1}{160} e^{-(x/160)} dx = 0.3194$. *Hence, there is about a 31.94% chance that a generator of this type will be operational between 60 and 160 hours.*

(c) $P(X > 200) = \int_{200}^{\infty} \frac{1}{160} e^{-(x/160)} dx = 0.2865$. *The chance that the generator will last more than 200 hours is about 28.65%.*

■

Another special case of gamma probability distribution that is useful in statistical inference problems is the chi-square distribution.

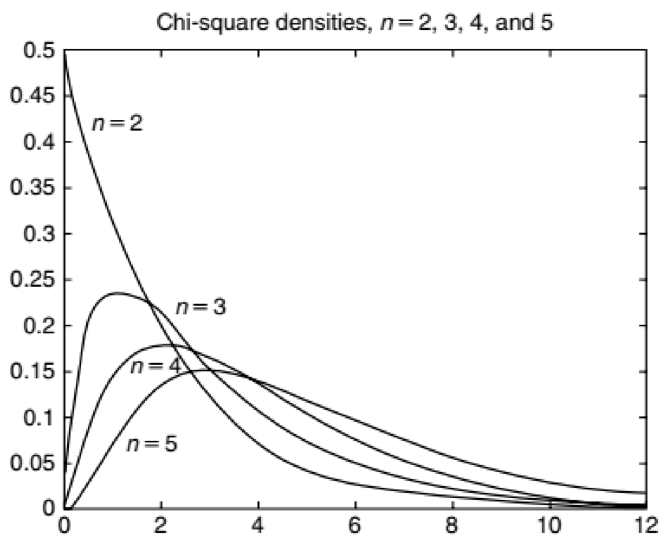
Definition 3.2.8 *Let n be a positive integer. A random variable, X, is said to have a* **chi-square** (χ^2) **distribution** *with n degrees of freedom if and only if X is a gamma random variable with parameters $\alpha = n/2$ and $\beta = 2$. We denote this by $X \sim \chi^2(n)$.*

Hence, the probability density function of a chi-square distribution with n degrees of freedom is given by

$$f(x) = \begin{cases} \dfrac{1}{\Gamma\left(\frac{n}{2}\right) 2^{n/2}} x^{(n/2)-1} e^{-x/2}, & 0 \le x < \infty \\ 0, & \text{otherwise.} \end{cases}$$

Figure 3.7 illustrates the dependence of the chi-square distribution on n.

The mean and variance of a chi-square random variable follow directly from Theorem 3.2.6.



■ **FIGURE 3.7** Chi-square pdfs for different degrees of freedom.

MEAN, VARIANCE, AND MGF OF A CHI-SQUARE RANDOM VARIABLE

Theorem 3.2.8 *If X is a chi-square random variable with n degrees of freedom, then $E(X) = n$ and $Var(X) = 2n$. Also, the moment-generating function is given by*

$$M_X(t) = \frac{1}{(1-2t)^{n/2}}, \quad t < \frac{1}{2}.$$

Another class of distributions that plays a crucial role in Bayesian statistics (see Chapter 11) is the beta distribution. The beta distribution is used as a prior distribution for binomial or geometric proportions. A random variable X is said to have a *beta distribution* with parameters α and β if and only if the density function of X is

$$f(x) = \begin{cases} \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}, & \alpha, \beta > 0; 0 \le x \le 1 \\ 0, & \text{otherwise,} \end{cases}$$

where $B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx$. It can be proved (see Exercise 3.2.31) that $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$, and that $E(X) = \frac{\alpha}{\alpha+\beta}$ and $Var(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.

One of the questions we may have is: "How do we know which distribution to use in a given physical problem?" There is no simple and direct answer to this question. One intuitive way is to construct a histogram from the information at hand; from the shape of this histogram, we decide whether the random variable follows a particular distribution such as gamma distribution. Once we decide that it follows a particular distribution, then the parameters of this distribution, such as α and β, must be statistically estimated. In Chapter 5, we discuss how to estimate these parameters. Then a goodness-of-fit test can be performed to see whether the distribution model seems to be the right one.

EXERCISES 3.2

3.2.1. A fair coin is tossed 10 times. Let X denote the number of heads obtained. Find the following.
(a) $P(X = 7)$
(b) $P(X \le 7)$
(c) $P(X > 0)$
(d) $E(X)$ and $Var(X)$

3.2.2. Let X be a Poisson random variable with $\lambda = 1/3$. Find
(a) $P(X = 0)$
(b) $P(X \ge 4)$.

3.2.3. For a standard normal random variable Z, find the value of z_0 such that
(a) $P(Z > z_0) = 0.05$
(b) $P(Z < z_0) = 0.88$
(c) $P(Z < z_0) = 0.10$
(d) $P(Z > z_0) = 0.95$.

3.2.4. Let $X \sim N(12, 5)$. Find the value of x_0 such that
(a) $P(X > x_0) = 0.05$
(b) $P(X < x_0) = 0.98$
(c) $P(X < x_0) = 0.20$
(d) $P(X > x_0) = 0.90$.

3.2.5. Let $X \sim N(10, 25)$. Compute
(a) $P(X \leq 20)$
(b) $P(X > 5)$
(c) $P(12 \leq X \leq 15)$
(d) $P(|X - 12| \leq 15)$.

3.2.6. A quarterback on a football team has a pass completion rate of 0.62. If, in a given game, he attempts 16 passes, what is the probability that he will complete
(a) 12 passes?
(b) More than half of his passes?
(c) Interpret your result.
(d) Out of the 16 passes, what is the expected number of completions?

3.2.7. A consulting group believes that 70% of the people in a certain county are satisfied with their health coverage. Assuming that this is true, find the probability that in a random sample of 15 people from the county:
(a) Exactly 10 are satisfied with their health coverage, and interpret.
(b) Not more than 10 are satisfied with their health coverage, and interpret.
(c) What is the expected number of people out of 15 that are satisfied with their health coverage?

3.2.8. A man fires at a target six times; the probability of his hitting it each time is independent of other tries and is 0.40.
(a) What is the probability that he will hit at least once?
(b) How many times must he fire at the target so that the probability of hitting it at least once is greater than 0.77?
(c) Interpret your findings.

3.2.9. A certain electronics company produces a particular type of vacuum tube. It has been observed that, on the average, three tubes of 100 are defective. The company packs the tubes in boxes of 400. What is the probability that a certain box of 400 tubes will contain
(a) r defective tubes?
(b) At least k defective tubes?
(c) At most one defective tube?
(d) Interpret your answers to (a), (b), and (c).

3.2.10. Suppose that, on average, in every two pages of a book there is one typographical error, and that the number of typographical errors on a single page of the book is a Poisson r.v. with $\lambda = 1/2$. What is the probability of at least one error on a certain page of the book? Interpret your result.

3.2.11. Show that the probabilities assigned by Poisson probability distribution satisfy the requirements that $0 \le p(x) \le 1$ for all x and $\sum_x p(x) = 1$.

3.2.12. In determining the range of an acoustic source using the triangulation method, the time at which the spherical wave front arrives at a receiving sensor must be measured accurately. Measurement errors in these times can be modeled as possessing uniform probability distribution from -0.05 to 0.05 microseconds. What is the probability that a particular arrival time measurement will be in error by less than 0.01 microsecond? What does your answer mean?

3.2.13. The hardness of a piece of ceramic is proportional to the firing time. Assume that a rating system has been devised to rate the hardness of a ceramic piece and that this measure of hardness is a random variable that is distributed uniformly between 0 and 10. If a hardness in [5,9] is desirable for kitchenware, what is the probability that a piece chosen at random will be suitable for kitchen use?

3.2.14. A receiver receives a string of 0s and 1s transmitted from a certain source. The receiver used a majority rule. That is, if the receiver acquires five symbols, of which three or more are 1s, it decides that a 1 was transmitted. The receiver is correct only 85% of the time. What is $P(W)$, the probability of a wrong decision if the probabilities of receiving 0s and 1s are equally likely? What can you conclude from your result?

3.2.15. The efficiency X of a certain electrical component may be assumed to be a random variable that is distributed uniformly between 0 and 100 units. What is the probability that X is:
(a) Between 60 and 80 units?
(b) Greater than 90 units?
(c) Interpret (a) and (b).

3.2.16. The *reliability function* of a system or a piece of equipment at time t is defined by

$$R(t) = P(T \ge t) = 1 - F(t)$$

where T, the failure time, is a random variable with a known distribution. A certain vacuum tube has been observed to fail uniformly over the interval $[t_1, t_2]$.
(a) Determine the reliability of such a tube at time t, $t_1 \le t \le t_2$.
(b) If $180 \le t \le 220$, what is the reliability of such a tube at 200 hours?
(c) The *failure* or *hazard rate function* $\rho(t)$ is defined by

$$\rho(t) = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{R(t)} = \frac{-\frac{dR(t)}{dt}}{R(t)}.$$

Calculate the failure rate of this vacuum tube. Interpret your result.

3.2.17. An electrical component was studied in the laboratory, and it was determined that its failure rate was approximately equal to $\frac{1}{\beta} = 0.05$. What is the reliability of such a component at 10 hours?

3.2.18. Suppose that the life length of a mechanical component is normally distributed.
(a) If $\sigma = 3$ and $\mu = 100$, find the reliability of such a system at 105 hours.
(b) What should be the expected life of the component if it has reliability of 0.90 for 120 hours?

3.2.19. A geologist defines granite as a rock containing quartz, feldspar, and small amounts of other minerals, provided that it contains not more than 75% quartz. If all the percentages are equally likely, what proportion of granite samples that the geologist collects during his lifetime will contain from 50% to 65% quartz?

3.2.20. For a normal random variable with pdf,

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

show that $\int_{-\infty}^{\infty} f(x)dx = 1$. [Hint: use polar coordinates.]

3.2.21. A professor in a large statistics class has a grading policy such that only the 15% of the students with the highest scores will receive the grade A. The mean score for this class is 72 with a standard deviation of 6. Assuming that all the grades for this class follow a normal probability distribution, what is the minimum score that a student in this class has to get to receive an A grade?

3.2.22. The scores, X, of an examination may be assumed to be normally distributed with $\mu = 70$ and $\sigma^2 = 49$. What is the probability that:
(a) A score chosen at random will be between 80 and 85?
(b) A score will be greater than 75?
(c) A score will be less than 90?
(d) Interpret the meaning of (a), (b), and (c).

3.2.23. Suppose that the diameters of golf balls manufactured by a certain company are normally distributed with $\mu = 1.96$ in. and $\sigma = 0.04$ in. A golf ball will be considered defective if its diameter is less than 1.90 in. or greater than 2.02 in. What is the percentage of defective balls manufactured by the company? What did the answer indicate?

3.2.24. Suppose that the arterial diastolic blood pressure readings in a population follow a normal probability distribution with mean 80 mm Hg and standard deviation 6.2 mm Hg. Suppose it is recommended that a physician be consulted if an individual has an arterial diastolic blood pressure reading of 90 mm Hg or more. If an individual is randomly picked from this population, what is the probability that this individual needs to consult a physician? Discuss the meaning of your result.

3.2.25. In a certain pediatric population, systolic blood pressure is normally distributed with mean 115 mm Hg and standard deviation 10 mm Hg. Find the probability that a randomly selected child from this population will have:
(a) A systolic pressure greater than 125 mm Hg.
(b) A systolic pressure less than 95 mm Hg.

(c) A systolic pressure below which 95% of this population lies.
(d) Interpret (a), (b), and (c).

3.2.26. A physical fitness test was given to a large number of college freshmen. In part of the test, each student was asked to run as far as he or she could in 10 minutes. The distance each student ran in miles was recorded and can be considered to be a random variable, say X. The data showed that the random variable X followed the log-normal distribution with $\mu_y = 0.35$ and $\sigma_y = 0.5$, where $Y = \ln X$. A student is considered physically fit if he or she is able to run 1.5 miles in the time allowed. What percentage of the college freshmen would be considered physically fit if we consider only this part of the test?

3.2.27. An experimenter is designing an experiment to test tetanus toxoid in guinea pigs. The survival of the animal following the dose of the toxoid is a random phenomenon. Past experience has shown that the random variable that describes such a situation follows the log-normal distribution with $\mu_y = 0$ and $\sigma_y = 0.65$. As a requirement of good design the experimenter must choose doses at which the probability of surviving is 0.20, 0.50, and 0.80. What three doses should he choose?

3.2.28. Show that $\Gamma(1) = 1$ and for $a > 1$, $\Gamma(a) = (a-1)\Gamma(a-1)$.

3.2.29. (a) Find the moment-generating function for a gamma probability distribution with parameter $\alpha > 0$ and $\beta > 0$. [Hint: In the integral representation of $E(e^{tX})$, change the variable t to $u = (1-\beta t)x/\beta$, with $(1-\beta t) > 0$.]
(b) Using the mgf of a gamma probability distribution, find $E(X)$ and $Var(X)$.

3.2.30. Let X be an exponential random variable. Show that, for numbers $a > 0$ and $b > 0$,

$$P(X > a + b \mid X > a) = P(X > b).$$

(This property of the exponential distribution is called the memoryless property of the distribution.)

3.2.31. A random variable X is said to have a *beta distribution* with parameters α and β if and only if the density function of X is

$$f(x) = \begin{cases} \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}, & \alpha, \beta > 0; 0 \le x \le 1 \\ 0, & \text{otherwise} \end{cases}$$

where $B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx$.
(a) Show that $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$.
(b) Show that $E(X) = \frac{\alpha}{\alpha+\beta}$ and $Var(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.

3.2.32. The daily proportion of major automobile accidents across the United States can be treated as a random variable having a beta distribution with $\alpha = 6$ and $\beta = 4$. Find the probability that, on a certain day, the percentage of major accidents is less than 80% but greater than 60%. Interpret your answer.

3.2.33. Suppose that network breakdowns occur randomly and independently of each other on an average rate of three per month.

(a) What is the probability that there will be just one network breakdown during December? Interpret.

(b) What is the probability that there will be at least four network breakdowns during December? Interpret.

(c) What is the probability that there will be at most seven network breakdowns during December? Interpret.

3.2.34. Let X be a random variable denoting the number of events occurring in the time interval $(0, t]$. Show that X has a gamma probability distribution with parameters n and λ.

3.2.35. In order to etch an aluminum tray successfully, the pH of the acid solution used must be between 1 and 4. This acid solution is made by mixing a fixed quantity of etching compound in powder form with a given volume of water. The actual pH of the solution obtained by this method is affected by the potency of the etching compound, by slight variations in the volume of water used, and perhaps by the pH of the water. Thus, the pH of the solution varies. Assume that the random variable that describes the random phenomenon is gamma distributed with $\alpha = 2$ and $\beta = 1$.

(a) What is the probability that an acid solution made by the foregoing procedure will satisfactorily etch a tray?

(b) What would the answer to part (a) be if $\alpha = 1$ and $\beta = 2$?

3.3 JOINT PROBABILITY DISTRIBUTIONS

We have thus far confined ourselves to studying one-dimensional or univariate random variables and their properties. In many practical situations, we are required to deal with several, not necessarily independent random variables. For example, we might be interested in a study involving the weights and heights (W, H) of a certain group of persons. In this situation, we need the two random variables (W, H), and it is likely that these two are related. Then it becomes important to study the joint effect of these random variables, which will lead to finding the joint probability distributions. In this section, we confine our studies to two random variables and their joint distributions, which are called *bivariate distributions*. We consider the random variables to be either both discrete or both continuous. We now define joint distribution of two random variables.

Definition 3.3.1 (a) *Let X and Y be random variables. If both X and Y are discrete, then*

$$f(x, y) = P(X = x, Y = y)$$

is called the **joint probability function** *(joint pmf) of X and Y.*

(b) *If both X and Y are continuous then $f(x, y)$ is called the* **joint probability density function** *(joint pdf) of X and Y if and only if*

$$P(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \int_c^d f(x, y) dx dy.$$

Example 3.3.1

A probability class contains 10 African American, 8 Hispanic American, and 15 white students. If 12 students are randomly selected from this class, and if X = number of black students, and Y = number of white students, find the joint probability function of the bivariate random variable (X, Y).

Solution

There are a total of 33 students. The number of ways in which x African American, and y white students can be picked (which means, the remaining $12-(x+y)$ students are Hispanic American) can be obtained using the multiplication principle as

$$\binom{10}{x}\binom{15}{x}\binom{8}{12-x-y}.$$

The number of ways to pick 12 students from 33 students is $\binom{33}{12}$. Hence, the joint probability function is

$$P(X=x, Y=y) = \frac{\binom{10}{x}\binom{15}{y}\binom{8}{12-x-y}}{\binom{33}{12}}$$

where $0 \le x \le 10$, $0 \le y \le 12$, and $4 \le x+y \le 12$. The last constraint is needed because there are only eight Hispanic Americans, so the combined minimum number of whites and African Americans should be at least 4.

We follow the notation: $\sum_{x,y}$ to denote $\sum_x \sum_y$. The joint distribution of two random variables has to satisfy the following conditions.

Theorem 3.3.1 *If X and Y are two random variables with joint probability function $f(x, y)$, then*

1. *$f(x, y) \ge 0$ for all x and y.*
2. *If X and Y are discrete, then $\sum_{x,y} f(x, y) = 1$,*

where the sum is over all values (x, y) that are assigned nonzero probabilities. If X and Y are continuous, then

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y) = 1.$$

Given the joint probability distribution (pdf or pmf), the probability distribution function of a component random variable can be obtained through the marginals.

Definition 3.3.2 *The* **marginal pmf** *of X denoted by $f_X(x)$ (or $f(x)$, when there is no confusion) is defined by*

$$f_X(x) = \begin{cases} \int_{-\infty}^{\infty} f(x, y)dy, & \text{if } X \text{ and } Y \text{ are continuous,} \\ \sum_{\text{all } y} f(x, y), & \text{if } X \text{ and } Y \text{ are discrete.} \end{cases}$$

Similarly, the **marginal pdf** *of Y is defined by*

$$f_Y(y) = \begin{cases} \int_{-\infty}^{\infty} f(x, y)dx, & \text{if } X \text{ and } Y \text{ are continuous,} \\ \sum_{\text{all } x} f(x, y), & \text{if } X \text{ and } Y \text{ are discrete.} \end{cases}$$

Note that

$$P(a \le X \le b) = \begin{cases} \int_{a}^{b} f_X(x)dx, & \text{if } X \text{ and } Y \text{ are continuous,} \\ \sum f_X(x), & \text{if } X \text{ and } Y \text{ are discrete,} \end{cases}$$

where summation is over all values of X from a to b.

Example 3.3.2

Find the marginal probability density function of the random variables X and Y, if their joint probability function is given by Table 3.1.

Table 3.1

	y				
x	−2	0	1	4	Sum
−1	0.2	0.1	0.0	0.2	0.5
3	0.1	0.2	0.1	0.0	0.4
5	0.1	0.0	0.0	0.0	0.1
Sum	0.4	0.3	0.1	0.2	1.0

Find the marginal densities of X and Y.

Solution

By definition, the marginal pdfs of X are given by the column sums (summands over y for fixed x), and the marginal pdfs of Y are obtained by the row sums. Hence,

x_i	−1	3	5	otherwise
$f_X(x_i)$	0.5	0.4	0.1	0

y_j	−2	0	1	4	otherwise
$f_Y(y_j)$	0.4	0.3	0.1	0.2	0

Using the joint probability distribution and the marginals, we can now introduce the conditional probability distribution function.

Definition 3.3.3 *The* **conditional probability distribution** *of the random variable X given Y is given by*

$$f(x|y) = f(x|Y = y)$$
$$= \begin{cases} \dfrac{f(x,y)}{f_Y(y)}, & \text{if } X \text{ and } Y \text{ are continuous, } f_Y(y) \neq 0, \\ \dfrac{P(X = x, Y = y)}{f_Y(y)}, & \text{if } X \text{ and } Y \text{ are discrete.} \end{cases}$$

We note that both the marginal probability densities of X and Y as well as the conditional pdf must satisfy the two important conditions of a pdf.

We know that two events A and B are independent if $P(A \cap B) = P(A)P(B)$. It is usually more convenient to establish independence through the probability functions. Hence, we define independence for bivariate probability distribution as follows.

Definition 3.3.4 *Let X and Y have a joint pmf or pdf $f(x, y)$. Then X and Y are* **independent** *if and only if*

$$f(x,y) = f_X(x) f_Y(y), \quad \text{for all } x \text{ and } y.$$

That is, for independent random variables, the joint pdf is the product of the marginals.

Example 3.3.3

Let

$$f(x,y) = \begin{cases} 3x, & 0 \le y \le x \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

(a) Find $P\left(X \le \frac{1}{2}, \frac{1}{4} < Y < \frac{3}{4}\right)$.
(b) Find the marginals $f_X(x)$ and $f_Y(y)$.
(c) Find the conditional $f(x|y)(0 < y < 1)$. Also compute $f\left(x|Y = \frac{1}{2}\right)$.
(d) Are X and Y independent?

Solution

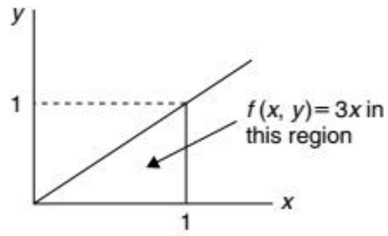
(a) *The domain of the function f(x,y) is given in Figure 3.8. The required probability $P\left(X \le \frac{1}{2}, \frac{1}{4} < Y < \frac{3}{4}\right)$ is the volume over the area of the shaded region as shown by Figure 3.9. That is,*

$$P\left(X \le \tfrac{1}{2}, \tfrac{1}{4} < Y < \tfrac{3}{4}\right) = \int_{1/4}^{1/2} \int_{1/4}^{x} 3x\,dy\,dx$$

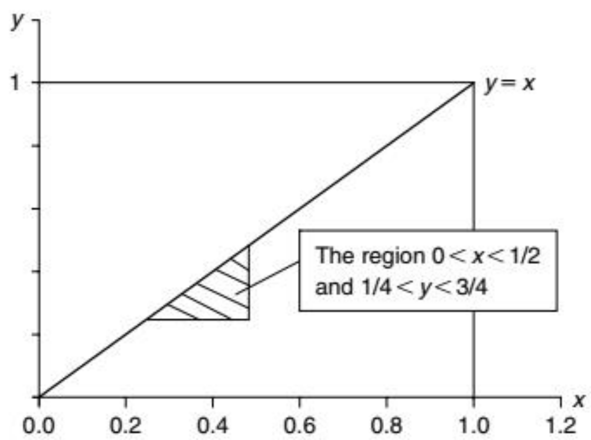
$$= \int_{1/4}^{1/2} 3x\left(x - \frac{1}{4}\right) dx$$

$$= \left(\frac{3x^3}{3} - \frac{3x^2}{8}\right)\Bigg|_{1/4}^{1/2}$$

$$= \frac{5}{128}.$$



■ **FIGURE 3.8** Domain of $f(x, y)$.



■ **FIGURE 3.9** Region of integration.

(b) *To find the marginals, we note that for each x, y varies from 0 to $x(0 < y < x)$. Therefore*

$$f_X(x) = \int_0^x 3x dy = 3x\left(y|_0^x\right) = 3x^2, \quad 0 < x < 1.$$

Similarly, for each y, x varies from y to 1.

$$f_Y(y) = \int_y^1 3x dx = \frac{3x^2}{2}\Bigg|_y^1 = \frac{3}{2} - \frac{3y^2}{2}$$

$$= \frac{3}{2}(1 - y^2), \quad 0 < y < 1.$$

(c) *Using the definition of conditional density*

$$f(x\,|y) = \frac{f(x,y)}{f_Y(y)} = \frac{3x}{\frac{3}{2}(1-y^2)} = \frac{2x}{1-y^2}, \quad y \le x \le 1.$$

From this we have

$$f\left(x\,|y=\tfrac{1}{2}\right) = \frac{2x}{\left(1-\left(\frac{1}{2}\right)^2\right)} = \frac{8}{3}x, \quad \tfrac{1}{2} \le x \le 1.$$

(d) *To check for independence of X and Y*

$$f_X(1)f_Y\left(\tfrac{1}{2}\right) = (3)\left(\tfrac{9}{8}\right) = \tfrac{27}{8} \neq 3 = f\left(1,\tfrac{1}{2}\right).$$

Hence, X and Y are not independent.

■

Recall that in the case of a univariate random variable X, with probability function $f(x)$, we have

$$EX = \begin{cases} \sum\limits_x xf(x), & \text{if } \sum\limits_x |x|f(x) < \infty, \text{ for discrete r.v.} \\ \int xf(x)dx, & \text{if } \int |x|f(x)dx < \infty, \text{ for continuous r.v.} \end{cases}$$

Now we define similar concepts for bivariate distribution.

Definition 3.3.5 *Let $f(x,y)$ be the joint probability function, and let $g(x,y)$ be such that $\sum_{x,y}|g(x,y)|f(x,y) < \infty$ in the discrete case, or $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|g(x,y)|f(x,y)dxdy < \infty$, in the continuous case. Then the* **expected value** *of $g(X,Y)$ is given by*

$$Eg(X,Y) = \begin{cases} \sum\limits_{x,y} g(x,y)f(x,y), & \textit{if } X,Y \textit{ are discrete}, \\ \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} g(x,y)f(x,y)dxdy, & \textit{if } X,Y \textit{ are continuous}. \end{cases}$$

In particular

$$E(X,Y) = \begin{cases} \sum\limits_{x,y} xyf(x,y), & \textit{if } X,Y \textit{ are discrete}, \\ \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} xyf(x,y)dxdy, & \textit{if } X,Y \textit{ are continuous}. \end{cases}$$

The following properties of mathematical expectation are easy to verify.

PROPERTIES OF EXPECTED VALUE

1. $E(aX+bY) = aE(X)+bE(Y)$.
2. If X and Y are independent, then $E(XY) = E(X)E(Y)$. However, the converse is not necessarily true.

Example 3.3.4

Let $f(x, y) = 3x, \quad 0 \le y \le x \le 1.$

(a) Find $E(4X - 3Y)$,

(b) Find $E(XY)$.

Solution

(a) $E(X) = \int x f_X(x)dx$ *and* $E(Y) = \int y f_Y(y)dy$.

Recall that earlier (Example 3.3.3) we have computed $f_X(x) = 3x^2$ $(0 < x < 1)$ *and* $f_Y(y) = \frac{3}{2}(1 - y^2)$, $0 \le y \le 1$. *Using these results, we have*

$$E(X) = \int_0^1 x3x^2 dx = \frac{3}{4},$$

$$E(Y) = \int_0^1 y\frac{3}{2}(1 - y^2)dy = \frac{3}{8}.$$

Hence,

$$E(4X - 3Y) = 3 - \frac{9}{8} = \frac{15}{8}.$$

(b)

$$E(XY) = \int_0^1 \int_0^x xy(3x)dydx = \frac{3}{10}.$$

Conditional expectations are defined in the same way as univariate expectations, except that the conditional density is utilized in place of the unconditional density function.

Definition 3.3.6 *Let X and Y be jointly distributed with pf or pdf $f(x, y)$. Let g be a function of x. Then the* **conditional expectation** *of $g(x)$ given, $Y = y$ is*

$$E(g(X)\,|y) = E(g(X)\,|Y = y)$$
$$= \begin{cases} \sum_{\text{all } x} g(x) f(x\,|y), & if\ X, Y\ are\ discrete, \\ \int g(x) f(x\,|y)dx, & if\ X, Y\ are\ continuous. \end{cases}$$

Note that $E(g(X)\,|y)$ is a function of y. If we let Y range over all of its possible values, the conditional expectation $E(g(X)\,|Y)$ can be thought of as a function of the random variable Y. We will then be able to find the mean and variance of $E(g(X)\,|Y)$, as given in the following result, the proof of which is left as an exercise.

Theorem 3.3.2 *Let X and Y be two random variables. Then*

(a) $E(X) = E[E(X|Y)]$.

(b) $Var(X) = E[Var(X|Y)] + Var[E(X|Y)]$.

Example 3.3.5

Let X and Y be two random variables with joint density function given by

$$f(x,y) = \begin{cases} x^2 + \frac{xy}{3}, & 0 \le x \le 1 \text{ and } 0 \le y \le 2 \\ 0, & \text{otherwise.} \end{cases}$$

Find the conditional expectation, $E\left(X|Y = \frac{1}{2}\right)$.

Solution

First we will find the conditional density, $f(x|y)$. The marginal

$$f_Y(y) = \int_0^1 \left(x^2 + \frac{xy}{3}\right) dx = \frac{1}{3} + \frac{1}{6}y, \quad 0 < y < 2.$$

Therefore,

$$f(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{x^2 + \frac{xy}{3}}{\frac{1}{6}y + \frac{1}{3}}, \quad 0 \le x \le 1.$$

Hence,

$$f\left(x|Y = \frac{1}{2}\right) = \frac{x^2 + \frac{x}{6}}{\frac{1}{12} + \frac{1}{3}} = \frac{12}{5}\left(x^2 + \frac{x}{6}\right).$$

Thus,

$$E\left(X|Y = \frac{1}{2}\right) = \int_0^1 x f(x|y)\, dx$$

$$= \int_0^1 x \frac{12}{5}\left(x^2 + \frac{x}{6}\right) dx = \frac{11}{15} = 0.733.$$

3.3.1 Covariance and Correlation

We will now define the covariance and correlation coefficient of two random variables.

Definition 3.3.7 (i) *The* **covariance** *between two random variables X and Y is defined by*

$$\sigma_{XY} = Cov(X,Y) = E(X - \mu_X)(Y - \mu_Y) = E(XY) - \mu_X \mu_Y,$$

where $\mu_X = E(X)$ and $\mu_Y = E(Y)$.

(ii) *The* **correlation coefficient**, $\rho = \rho(x, y)$ *is defined by*

$$\rho = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}}.$$

Correlation is the measure of the linear relationship between the random variables X and Y. If $Y = aX+b(a \neq 0)$, then $\rho(x, y) = 1$. If dependence on X and Y needs to be specified, we will use the notation, ρ_{XY}.

From the definition of the covariance of X and Y, we note that if small values of X, for which $(X - \mu_X) < 0$, tend to be associated with small values of Y, for which $(Y - \mu_Y) < 0$, and similarly large values of X with large values of Y, then $Cov(X, Y) \equiv E[(X - \mu_X)(Y - \mu_Y)]$ can be expected to be positive. On the other hand, if small values of X tend to be associated with large values of Y and vice versa so that $(X - \mu_X)$ and $(Y - \mu_Y)$ are of opposite signs, then $Cov(X, Y) < 0$. Thus, covariance can be thought of as a signed measure of the variation of Y relative to X. If X and Y are independent, then it follows from the definition of covariance that $Cov(X, Y) = 0$. The correlation coefficient of X and Y, is a dimensionless quantity that measures the linear relationship between the random variables X and Y.

PROPERTIES OF COVARIANCE AND CORRELATION COEFFICIENT

(a) $-1 \leq \rho \leq 1$.

(b) If X and Y are independent, then $\rho = 0$. The converse is not true.

(c) If $Y = aX + b$, then

$$Cov(X; Y) = \begin{cases} 1, & \text{if } a > 0, \\ -1, & \text{if } a < 0. \end{cases}$$

Note that $Cov(X,X) = Var(X)$.

(d) If $U = a_1X + b_1$ and $V = a_2Y + b_2$, then

(i) $Cov(U,V) = a_1a_2Cov(X,Y)$,

and

(ii) $\rho_{UV} = \begin{cases} \rho_{XY}, & \text{if } a_1a_2 > 0 \\ -\rho_{XY}, & \text{otherwise.} \end{cases}$

(e) $Var(aX + bY) = a^2Var(X) + b^2Var(Y) + 2abCov(X,Y)$.

Example 3.3.6

The joint probability density of the random variables X and Y is given by

$$f(x, y) = \begin{cases} \frac{1}{64}e^{-y/8}, & 0 \leq x \leq y < \infty \\ 0, & \text{otherwise.} \end{cases}$$

Find the covariance of X and Y.

Solution

We can use the formula, $Cov(X, Y) = E(XY) - E(X)E(Y)$. Now using integration by parts (three times) we will get

$$E(XY) = \int_0^\infty \int_0^y (xy) \frac{1}{64} e^{-y/8} dx dy$$

$$= \frac{1}{64} \int_0^\infty y e^{-y/8} \left(\int_0^y x dx \right) dy$$

$$= \frac{1}{128} \int_0^\infty y^3 e^{-y/8} dy = 192.$$

We can also obtain

$$E(X) = \int_0^\infty \int_0^y x \frac{1}{64} e^{-y/8} dx dy = 8$$

and

$$E(Y) = \int_0^\infty \int_0^y y \frac{1}{64} e^{-y/8} dx dy = 16.$$

Thus, $Cov(X, Y) = 192 - (8)(16) = 64$.

Next we will define the moment-generating function for the bivariate distributions.

Definition 3.3.8 *Let X and Y be jointly distributed. Then the joint* **moment-generating function** *is defined by*

$$M_{(X,Y)}(t_1, t_2) = E\left(e^{t_1 X + t_2 Y}\right)$$

$$= \begin{cases} \sum_y \sum_x e^{t_1 x + t_2 y} f(x, y), & \text{if } X \text{ and } Y \text{ are discrete} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{t_1 x + t_2 y} f(x, y) dx dy, & \text{if } X \text{ and } Y \text{ are continuous.} \end{cases}$$

EXERCISES 3.3

3.3.1. An experiment consists of drawing four objects from a container, which holds eight operable, six defective, and 10 semioperable objects. Let X be the number of operable objects drawn and Y the number of defective objects drawn.

(a) Find the joint probability function of the bivariate random variable (X, Y).
(b) Find $P(X = 3, Y = 0)$.
(c) Find $P(X < 3, Y = 1)$.
(d) Give a graphical presentation of (a), (b), and (c).

3.3.2. Let

$$f(x, y) = \begin{cases} \frac{1}{50}(x^2 + 2y), & x = 0, 1, 2, 3 \text{ and } y = x + 3, \\ 0, & \text{otherwise.} \end{cases}$$

Show that $f(x, y)$ satisfies the conditions of a probability density function.

3.3.3. Let

$$f(x, y) = c(1 - x)(1 - y), \quad -1 \le x \le 1, \ -1 \le y \le 1.$$

Find the c that makes $f(x, y)$ the joint probability density function of the random variable (X, Y).

3.3.4. Let

$$f(x, y) = xe^{-xy}, \quad x \ge 0, \ y \ge 1.$$

Is $f(x, y)$ a probability density function? If not, find the proper constant to multiply with $f(x, y)$ so that it will be a probability density.

3.3.5. Find the marginal probability density function of the random variables X and Y, if their joint probability density function is given in Table 3.3.1.

Table 3.3.1

		y		
x	−2	0	1	4
−1	0.3	0.1	0.0	0.2
3	0.0	0.2	0.1	0.0
5	0.1	0.0	0.0	0.0

3.3.6. Find the marginal density functions of the random variables X and Y if their joint probability density function is given by

$$f(x, y) = \begin{cases} \frac{1}{5}(3x - y), & 1 \le x \le 2, \ 1 \le y \le 3, \\ 0, & \text{otherwise.} \end{cases}$$

3.3.7. Determine the conditional probability $P(X = -1 \mid Y = 0)$ for the random variables defined in Problem 3.3.5.

3.3.8. Find k so that $f(x, y) = kxy$, $1 \le x \le y \le 2$ will be a probability density function. Also find (i) $P(X \le \frac{3}{2}, Y \le \frac{3}{2})$, and (ii) $P(X + Y \le \frac{3}{2})$.

3.3.9. The random variables X and Y have a joint density

$$f(x, y) = \begin{cases} \frac{8}{9}xy, & 1 \le x \le y \le 2, \\ 0, & \text{elsewhere.} \end{cases}$$

Find:
(a) The marginal of X.
(b) $P(1.5 < X < 1.75, Y > 1)$.

3.3.10. The joint pdf of X and Y is

$$f(x, y) = \begin{cases} \frac{1}{28}(4x + 2y + 1), & 0 \le x \le 2, 0 \le y \le 2 \\ 0, & \text{elsewhere.} \end{cases}$$

Find (a) $f_X(x)$ and $f_Y(y)$, and (b) $f(y \mid x)$.

3.3.11. Find the joint mgf of the random variables (X, Y) defined in Problem 3.3.9.

3.3.12. The joint density of a random variable (X, Y) is given by

$$f(x, y) = \begin{cases} \frac{x^3 y^3}{16}, & 0 \le x \le 2, \ 0 \le y \le 2 \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Find marginals of X and Y, and (b) find $f(y \mid x)$.

3.3.13. The joint probability function of a discrete random variable (X, Y) is given by

$$f(x, y) = \begin{cases} \left[\frac{6xy}{n(n+1)(2n+1)}\right]^2, & x, y = 1, 2, \ldots, n, \\ 0, & \text{otherwise.} \end{cases}$$

Find (a) $f(x \mid y)$, and (b) $f(y \mid x)$.
[Hint: $\sum_{i=1}^{n} i^2 = (n(n + 1)(2n + 1))/6$.]

3.3.14. Consider bivariate random variables with the density

$$f(x, y) = \binom{n}{x} y^{x+\alpha-1}(1 - y)^{n-x+\beta-1}, \quad \text{for } x = 0, 1, \ldots, n$$

$$\text{and } 0 < y \le 1.$$

Verify that

$$f(x\,|y) \propto \binom{n}{x} y^x (1-y)^{n-x}$$

and

$$f(y|x) \propto y^{x+\alpha-1}(1-y)^{n-x+\beta-1}.$$

3.3.15. The joint density function of the discrete random variable (X, Y) is given in Table 3.3.2.

Table 3.3.2

	y		
x	1	2	3
1	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
2	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{12}$
3	$\frac{1}{12}$	$\frac{1}{12}$	0

(a) Find $E(XY)$.
(b) Find $Cov(X, Y)$.
(c) Find the correlation coefficient $\rho_{X,Y}$.

3.3.16. The joint probability function of the continuous random variable (X, Y) is given by

$$f(x, y) = \begin{cases} \frac{1}{28}(4x + 2y + 1), & 0 \le x < 2,\ 0 \le y < 2, \\ 0, & \text{otherwise.} \end{cases}$$

(a) Find $E(XY)$.
(b) Find $Cov(X, Y)$.
(c) Find the correlation coefficient ρ_{XY}.

3.3.17. Let X and Y be random variables and $U = aX + b$, $V = cY + d$, where a, b, c, d are constants. Show that $\rho_{UV} = \begin{cases} \rho_{XY}, & \text{if } ac > 0 \\ -\rho_{XY}, & \text{otherwise.} \end{cases}$

3.3.18. Let X and Y be two independent random variables, and let $Y = aX + b$, where a and b are constants. Show that (a) $\rho_{XY} = 1$ if $a > 0$, and (b) $\rho_{XY} = -1$ if $a < 0$.

3.3.19. If $|\rho_{XY}| = 1$, then prove that $P(Y = aX + b) = 1$.

3.3.20. Let X and Y be two random variables with joint density function

$$f(x, y) = \begin{cases} 8xy, & 0 \le x \le y \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

(a) Find the conditional expectation, $E\left(X|Y = \frac{3}{4}\right)$.
(b) Find $Cov(X, Y)$.

3.3.21. Let X and Y be two random variables with joint density function

$$f(x, y) = \begin{cases} e^{-y}, & 0 \le x \le y \\ 0, & \text{otherwise.} \end{cases}$$

(a) Find the conditional expectation, $E(X|Y = y)$.
(b) Find $Cov(X, Y)$.
(c) Are X and Y independent? Why?

3.3.22. Let

$$f(x, y) = \frac{c}{(1 + x^2)\sqrt{1 - y^2}}, \quad -\infty < x < \infty, \quad -1 < y < 1.$$

Find the c that makes $f(x, y)$ the probability density function of the random variable (X, Y). Determine whether X and Y are independent.

3.3.23. If the random variables X and Y are independent and have equal variances, what is the coefficient of correlation between the random variables X and $aX + Y$, where a is a constant?

3.4 FUNCTIONS OF RANDOM VARIABLES

In this section we discuss the methods of finding the probability distribution of a function of a random variable X. We are given the distribution of X, and we are required to find the distribution of $g(X)$. There are many physical problems that call for the derivation of the distribution of a function of a random variable. The following is one of the classical examples. The velocity V of a gas molecule (Maxwell–Boltzmann law) behaves as a gamma-distributed random variable. We would like to derive the distribution of $E = mV^2$, the kinetic energy of the gas molecule. Because the value of the velocity is the outcome of a random experiment, so is the value of E. This is a problem of finding the distribution of a function of a random variable $E = g(V)$. We now illustrate various techniques for finding the distribution of $g(X)$ by means of examples.

3.4.1 Method of Distribution Functions

Basically the *method of distribution functions* is as follows. If X is a random variable with pdf $f_X(x)$ and if Y is some function of X, then we can find the cdf $F_Y(y) = P(Y \le y)$ directly by integrating $f_X(x)$ over the region for which $\{Y \le y\}$. Now, by differentiating $F_Y(y)$, we get the probability density function $f_Y(y)$ of Y. In general, if Y is a function of random variables $X_1, \ldots, X_n$, say $g(X_1, \ldots, X_n)$, then we can summarize the method of distribution function as follows.

PROCEDURE TO FIND CDF OF A FUNCTION OF R.V. USING THE METHOD OF DISTRIBUTION FUNCTIONS

1. Find the region $\{Y \le y\}$ in the $(x_1, x_2, \ldots, x_n)$ space, that is find the set of $(x_1, x_2, \ldots, x_n)$ for which $g(x_1, \ldots, x_n) \le y$.
2. Find $F_Y(y) = P(Y \le y)$ by integrating $f(x_1, x_2, \ldots, x_n)$ over the region $\{Y \le y\}$.
3. Find the density function $f_Y(y)$ by differentiating $F_Y(y)$.

Example 3.4.1

Let $X \sim N(0, 1)$. Using the cdf of X, find the pdf of X^2.

Solution

Let $Y = X^2$. Note that the pdf of X is

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad -\infty < x < \infty.$$

Then the cumulative distribution function of Y for a given $y \ge 0$ is

$$\begin{aligned} F(y) &= P(Y \le y) = P(X^2 \le y) \\ &= P(-\sqrt{y} \le X \le \sqrt{y}) \\ &= \int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \\ &= 2\int_{0}^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx, \quad \text{(by the symmetry of } e^{-x^2/2}\text{)}. \end{aligned}$$

Hence, by differentiating $F(y)$, we obtain the probability density function as

$$\begin{aligned} f_Y(y) &= \frac{2}{\sqrt{2\pi}} e^{-y/2} \frac{1}{2\sqrt{y}} \\ &= \begin{cases} \frac{1}{\sqrt{2\pi}} y^{-1/2} e^{-y/2}, & 0 < y < \infty \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

This is a χ^2-distribution with 1 degree of freedom.

The same method can be used for the discrete case.

Example 3.4.2

Suppose that the random variable X has a Poisson probability distribution

$$f(x) = \begin{cases} \frac{e^{-\lambda}\lambda^x}{x!}, & x = 0, 1, 2, \ldots \\ 0, & \text{otherwise.} \end{cases}$$

Find the cumulative distribution function of $Y = aX + b$.

Solution

The cdf of Y is given by

$$\begin{aligned} F(y) &= P(Y \le y) = P(aX + b \le y) \\ &= P\left(X \le \frac{y-b}{a}\right) = \sum_{x=0}^{\left[\frac{y-b}{a}\right]} \frac{e^{-\lambda}\lambda^x}{x!}, \end{aligned}$$

where $[x]$ is the largest integer less than or equal to x. Therefore,

$$F(y) = \begin{cases} 0, & y < b \\ \sum_{x=0}^{\left[\frac{y-b}{a}\right]} \frac{e^{-\lambda}\lambda^x}{x!}, & y \ge b. \end{cases}$$

It should be noted here that the pmf, $f_Y(y)$ of Y, can be found from the equation

$$f_Y(y) = F_Y(y) - F_Y(y-1), \quad \text{for } y = an + b, \quad n = 0, 1, 2, \ldots$$

The multivariate case (in particular, the bivariate case), though more difficult, can be handled similarly.

3.4.2 The pdf of $Y = g(X)$, Where g Is Differentiable and Monotone Increasing or Decreasing

We now consider the distribution of a random variable $Y = g(X)$, where X is a continuous random variable with pdf $f_X(x)$. Assume that g is differentiable and the inverse function g^{-1} of g exists. Let $X = g^{-1}(Y)$. Let $f_X(x)$ be the probability density function of X. Then the density function of Y can be obtained using the method just given. Thus,

$$f_Y(y) = f_X(g^{-1}(y)) \cdot \frac{d}{dy} g^{-1}(y).$$

This is a special case of the transformation method, which is explained later in Subsection 3.4.4.

Example 3.4.3
Let $X \sim N(0, 1)$. Find the pdf of $Y = e^X$.

Solution
Here $g(x) = e^x$, *and hence,* $g^{-1}(y) = \ln(y)$. *Thus,* $\frac{d}{dy}g^{-1}(y) = \frac{1}{y}$.
Also,

$$f_X(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}, \quad -\infty < x < \infty.$$

Therefore, the pdf of Y *is*

$$f_Y(y) = \begin{cases} \frac{1}{y\sqrt{2\pi}}e^{-[\ln(y)]^2/2}, & y > 0, \\ 0, & \text{otherwise.} \end{cases}$$

3.4.3 Probability Integral Transformation

Let X be a continuous random variable, with pdf f and cdf F. Let $Y = F(X)$. Then,

$$\begin{aligned} P(Y \le y) &= P(F(X) \le y) = P(X \le F^{-1}(y)) \\ &= \int_{-\infty}^{F^{-1}(y)} f_X(x)dx = F_X(x)\Big|_{-\infty}^{F^{-1}(y)} = y. \end{aligned}$$

Hence,

$$f(y) = \begin{cases} 1, & 0 < y < 1 \\ 0, & \text{otherwise.} \end{cases}$$

Thus, Y has a $U(0, 1)$ distribution. The transformation $Y = F(X)$ is called a *probability integral transformation*. It is interesting to note that irrespective of the pdf of X, Y is always uniform in $(0, 1)$.

Example 3.4.4
Let X be a normal with mean μ and variance σ^2. Thus,

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{-(x-\mu)/2\sigma^2}, -\infty < x < \infty, -\infty < \mu < \infty, \text{and } \sigma^2 > 0.$$

Let $Y = \int_0^X \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)/2\sigma^2} du$. Then $Y = F(X)$, where F is the cdf of a standard normal random variable. Therefore Y is uniform on $(0, 1)$. That is,

$$f(y) = \begin{cases} 1, & \text{if } 0 < y < 1 \\ 0, & \text{otherwise.} \end{cases}$$

3.4.4 Functions of Several Random Variables: Method of Distribution Functions

We now discuss the distribution of Y, when Y is a function of several random variables, $Y = g(X_1, \dots, X_n)$.

Example 3.4.5

Let $X_1, \dots, X_n$ be continuous iid random variables with pdf $f(x)$ (cdf $F(x)$). Find the pdfs of

$$Y_1 = \min(X_1, \dots, X_n) \quad \text{and} \quad Y_n = \max(X_1, \dots, X_n).$$

Solution

For the random variable Y_1, we have

$$\begin{aligned} 1 - F_{Y_1}(y) &= P(Y_1 > y) \\ &= P(X_1 > y, X_2 > y, \dots, X_n > y) \\ &= P(X_1 > y)P(X_2 > y)\dots P(X_n > y) \\ &\qquad \text{(because of independence)} \\ &= (1 - F(y))^n. \end{aligned}$$

This implies

$$F_{Y_1}(y) = 1 - (1 - F(y))^n$$

and

$$f_{Y_1}(y) = n(1 - F(y))^{n-1} f(y).$$

Consider Y_n. Its cdf is given by

$$F_{Y_n}(y) = P(Y_n \le y) = (F(y))^n.$$

This implies that

$$f_{Y_n}(y) = n(F(y))^{n-1} f(y).$$

3.4.5 Transformation Method

A simple generalization of the method of distribution functions to functions of more than one variable is the *transformation method*. We illustrate the method for bivariate distributions. The method is similar for the multivariate case. Let the joint pdf of (X, Y) be $f(x, y)$. Let $U = g_1(X, Y)$; $V = g_2(X, Y)$. The mapping from (X, Y) to (U, V) is assumed to be one-to-one and onto. Hence, there are functions, h_1 and h_2 such that

$$x = h_1^{-1}(u, v),$$

and

$$y = h_2^{-1}(u, v).$$

Define the Jacobian of the transformation J by

$$J = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial u} \end{vmatrix}.$$

Then the joint pdf of U and V is given by

$$f(u, v) = f(h_1^{-1}(u, v), h_2^{-1}(u, v))\,|J|\,.$$

Example 3.4.6

Let X and Y be independent random variables with common pdf $f(x) = e^{-x}$, $(x > 0)$. Find the joint pdf of $U = X/(X + Y)$, $V = X + Y$.

Solution

We have $U = X/(X + Y) = X/V$. *Hence,* $X = UV$ *and* $Y = V - X = V - UV = V(1 - U)$. *Thus, the Jacobian*

$$J = \begin{vmatrix} v & u \\ -v & 1 - u \end{vmatrix}.$$

Then $|J| = v(1 - u) + uv = v(> 0)$. *Note that* $0 \le u \le 1, 0 < v < \infty$.

$$\begin{aligned} f(u, v) &= f\left(h_1^{-1}(u, v), h_2^{-1}(u, v)\right)|J| \\ &= e^{-uv}e^{-v(1-u)}v \\ &= ve^{-v}, \quad 0 \le u \le 1, 0 < v < \infty. \end{aligned}$$

Suppose we want the marginal $f_V(v)$ and $f_U(v)$, that is,

$$f_V(v) = \int_0^1 ve^{-v}du = ve^{-v}, \quad 0 < v < \infty$$

and

$$f_U(u) = \int_0^\infty ve^{-v}dv = 1, \quad 0 \le u \le 1.$$

Sometimes the expressions for two variables, U and V, may not be given. Only one expression is available. In that case, call the given expression of X and Y as U, and define $V = Y$. Then, we can use the previous method to first find the joint density and then find the marginal to obtain the pdf of U. The following example demonstrates the method.

Example 3.4.7

Let X and Y be independent random variables uniformly distributed on [0, 1]. Find the distribution of $X+Y$.

Solution

Let

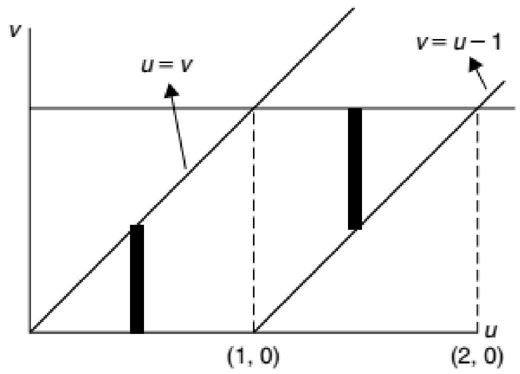
$$\begin{aligned} U &= X + Y, \\ V &= Y, \\ f(x, y) &= 1, \quad 0 \le x \le 1, 0 \le y \le 1, \\ X &= U - V, \\ Y &= V, \\ J &= \begin{vmatrix} 1 & -1 \\ 0 & 1 \end{vmatrix} = 1. \end{aligned}$$

Thus, we have

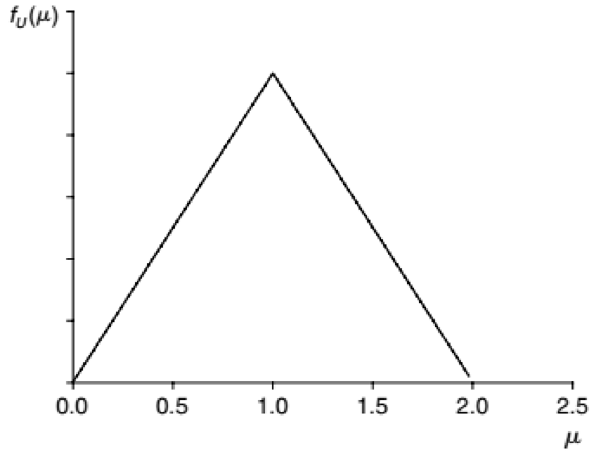
$$f(u, v) = \begin{cases} 1, & 0 \le u - v \le 1, \quad 0 \le v \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

Because V is the variable we introduced, to get the pdf of U, we just need to find the marginal pdf from the joint pdf. From Figure 3.10, the regions of integration are $0 \le u \le 1$, and $0 \le u \le 2$. That is,

$$\begin{aligned} f_U(u) &= \int f(u, v)dv = \int 1dv \\ &= \begin{cases} \int_0^u dv = u, & 0 \le u \le 1 \\ \int_{u-1}^1 dv = 2 - u, & 0 \le u \le 2. \end{cases} \end{aligned}$$



■ **FIGURE 3.10** The regions of integration.



■ **FIGURE 3.11** Graph of $f_U(u)$.

Figure 3.11 shows the graph of $f_U(u)$.

EXERCISES 3.4

3.4.1. Let X be a uniformly distributed random variable over $(0, a)$. Find the pdf of $Y = cX + d$.

3.4.2. The joint pdf of (X, Y) is

$$f(x, y) = \frac{1}{\theta^2} e^{-\frac{x+y}{\theta}}, \quad x, y > 0, \ \theta > 0.$$

Find the pdf of $U = X - Y$.

3.4.3. Let $f(x, y)$ be the probability density function of the continuous random variable (X, Y). If $U = XY$, show that the probability density function of U is given by

$$f_U(u) = \int_{-\infty}^{\infty} f\left(\frac{u}{v}, v\right) \left|\frac{1}{v}\right| dv.$$

3.4.4. The joint pdf of X and Y is

$$f(x, y) = \theta e^{-(x+\theta y)}, \quad \theta > 0, \ x > 0.$$

Find the pdf of XY.

3.4.5. If the joint pdf of (X, Y) is

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{1}{4\sigma_1^2\sigma_2^2}(x^2+y^2)}, \quad -\infty < x < \infty,$$
$$-\infty < y < \infty; \sigma_1, \sigma_2 > 0$$

find the pdf of $X^2 + Y^2$.

3.4.6. Let $X_1, \ldots, X_n$ be independent and identically distributed random variables with pdf $f(x) = (1/\theta)e^{-x/\theta}, x > 0, \theta > 0$. Find the pdf of $\sum_{i=1}^{n} X_i$.

3.4.7. Let $f(x, y)$ be the pdf of the continuous random variable (X, Y). If $U = X + Y$, then show that the probability density function of U is given by

$$f_U(u) = \int_{-\infty}^{\infty} f(u - v, v)dv.$$

3.4.8. Let X be uniformly distributed over $(-2, 2)$ and $Y = X^2$. Find the $Cov(X, Y)$. Are X and Y independent?

3.4.9. Let $X \sim N(\mu, \sigma^2)$. Show that
(a) $Z = \frac{(X-\mu)}{\sigma}$ is $N(0, 1)$.
(b) $U = \frac{(X-\mu)^2}{\sigma^2}$ is $\chi^2(1)$.

3.4.10. Let $X \sim N(\mu, \sigma^2)$. Find the pdf of $Y = e^X$.

3.4.11. The probability density of the velocity, V, of a gas molecule, according to the Maxwell–Boltzmann law, is given by

$$f(v, \beta) = \begin{cases} cv^2 e^{-\beta v^2}, & v > 0, \\ 0, & \text{elsewhere} \end{cases}$$

where c is an appropriate constant and β depends on the mass of the molecule and the absolute temperature. Find the density function of the kinetic energy E, which is given by $E = g(V) = \frac{1}{2}mV^2$.

3.4.12. Let X and Y be two independent random variables, each normally distributed, with parameters (μ_1, σ_1^2), and (μ_2, σ_2^2), respectively. Show that the probability density function of $U = X/Y$ is given by

$$f_U(u) = \frac{\sigma_1\sigma_2}{\pi\left(\sigma_1^2 + \sigma_2^2 u^2\right)}, \quad -\infty < u < \infty.$$

3.4.13. Let

$$f(x, y) = \frac{1}{2\pi\sigma^2} e^{-(1/2\sigma^2)(x^2+y^2)}, \quad -\infty < x, y < \infty$$

be the joint pdf of (X, Y). Let

$$U = \sqrt{X^2 + Y^2} \quad \text{and} \quad V = \tan^{-1}\left(\frac{Y}{X}\right), \quad 0 \le V \le 2\pi.$$

Find the joint pdf of (U, V).

3.4.14. Let the joint pdf of (X, Y) be given by

$$f(x, y) = \begin{cases} \beta^{-2} e^{-\{(x+y)/\beta\}}, & x, y > 0, \beta > 0, \\ 0, & \text{elsewhere.} \end{cases}$$

Let $U = \dfrac{X - Y}{2}$ and $V = Y$. Find the joint pdf of (U, V).

3.4.15. Let X and Y be independent and identically distributed random variables with pdf

$$f(x) = \begin{cases} \frac{1}{2} e^{-x/2}, & x \ge 0, \\ 0, & \text{otherwise.} \end{cases}$$

Find the distribution of $(X - Y)/2$.

3.4.16. If X and Y are independent and chi-square distributed random variables with n_1 and n_2 degrees of freedom, respectively. Obtain the joint distribution of (U, V), where $U = X + Y$ and $V = X/Y$.

3.5 LIMIT THEOREMS

Limit theorems play a very important role in the study of probability theory and in its applications. In Chapter 2, we saw that the frequency interpretation of probability depends on the long-run proportion of times the outcome (event) would occur in repeated experiments. Also, in Section 3.2, we learned that some binomial probabilities can be computed using either the Poisson probability distribution or the normal probability distribution using the limiting arguments. Many random variables that we encounter in nature have distributions close to the normal probability distribution. These modeling simplifications are possible because of various limit theorems. In this section, we discuss the law of large numbers and the Central Limit Theorem.

First we give Chebyshev's theorem, which is a useful result for proving limit theorems. It gives a lower bound for the area under a curve between two points that are on opposite sides of the mean and are equidistant from the mean. The strength of this result lies in the fact that we need not know the distribution of the underlying population, other than its mean and variance. This result was developed by the Russian mathematician Pafnuty Chebyshev (1821–1894).

CHEBYSHEV'S THEOREM

Theorem 3.5.1 *Let the random variable X have a mean μ and standard deviation σ. Then for $K > 0$, a constant,*

$$P(|X-\mu| < K\sigma) \geq 1 - \frac{1}{K^2}.$$

Proof. We will work with the continuous case. By definition of the variance of X,

$$\begin{aligned}
\sigma^2 = E(X-\mu)^2 &= \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx \\
&= \int_{-\infty}^{\mu-K\sigma} (x-\mu)^2 f(x)dx + \int_{\mu-K\sigma}^{\mu+K\sigma} (x-\mu)^2 f(x)dx + \int_{\mu+K\sigma}^{\infty} (x-\mu)^2 f(x)dx \\
&\geq \int_{-\infty}^{\mu-K\sigma} (x-\mu)^2 f(x)dx + \int_{\mu+K\sigma}^{\infty} (x-\mu)^2 f(x)dx.
\end{aligned}$$

Note that $(x-\mu)^2 \geq K^2\sigma^2$ for $x \leq \mu - K\sigma$ or $x \geq \mu + K\sigma$. The equation above can be rewritten as

$$\begin{aligned}
\sigma^2 &\geq K^2\sigma^2 \left[\int_{-\infty}^{\mu-K\sigma} f(x)dx + \int_{\mu+K\sigma}^{\infty} f(x)dx \right] \\
&= K^2\sigma^2 \left[P\{X \leq \mu - K\sigma\} + P\{X \geq \mu + K\sigma\} \right] \\
&= K^2\sigma^2 P\{|X-\mu| \geq K\sigma\}.
\end{aligned}$$

□

This implies that

$$P\{|X-\mu| \geq K\sigma\} \leq \frac{1}{K^2}$$

or

$$P(|X - \mu| < K\sigma) \geq 1 - \frac{1}{K^2}.$$

We can also write Chebyshev's theorem as

$$P\{|X - \mu| \geq \varepsilon\} \leq \frac{E\left[(X-\mu)^2\right]}{\varepsilon^2} = \frac{Var(X)}{\varepsilon^2}, \quad \text{for some } \varepsilon > 0.$$

Equivalently,

$$P\{|X - \mu| \geq K\sigma\} \leq \frac{1}{K^2}.$$

In other words, Chebyshev's inequality states that the probability that a random variable X differs from its mean by at least K standard deviations is less than or equal to $1/K^2 (K \geq 2)$.

In statistics, if we do not have any idea of the population distribution, Chebyshev's theorem is used in the following manner. For any data set (regardless of the shape of the distribution), at least $(1-(1/k^2))100\%$ of observations will lie within $k(\geq 1)$ standard deviations of the mean. For example, at least $(1-(1/2^2))100\% = 75\%$ of the data will fall in the interval $(\bar{x}-2s, \bar{x}+2s)$ and at least 88.9% of the observations will lie within three standard deviations of the mean. If the population distribution is bell shaped, we have a better result than Chebyshev's theorem, namely, the empirical rule that states the following: **(i)** approximately 68% of the observations lie within one standard deviation of the mean; **(ii)** approximately 95% of the observations lie within two standard deviations of the mean; and **(iii)** approximately 99.7% of the observations lie within three standard deviations of the mean.

Example 3.5.1

A random variable X has mean 24 and variance 9. Obtain a bound on the probability that the random variable X assumes values between 16.5 to 31.5.

Solution

From Chebyshev's theorem.

$$P\{\mu - K\sigma < X < \mu + K\sigma\} \geq 1 - \frac{1}{K^2}.$$

Equating $\mu + K\sigma$ *to 31.5 and* $\mu - K\sigma$ *to 16.5 with* $\mu = 24$ *and* $\sigma = \sqrt{9} = 3$, *we obtain* $K = 2.5$. *Hence,*

$$P\{16.5 < X < 31.5\} \geq 1 - \frac{1}{(2.5)^2} = 0.84.$$

Example 3.5.2

Let X be a random variable that represents the systolic blood pressure of the population of 18- to 74-year-old men in the United States. Suppose that X has mean 129 mm Hg and standard deviation 19.8 mm Hg.

(a) Obtain a bound on the probability that the systolic blood pressure of this population will assume values between 89.4 and 168.6 mm Hg.

(b) In addition, assume that the distribution of X is approximately normal. Using the normal table, find $P(89.4 \leq X \leq 168.6)$. Compare this with the empirical rule.

Solution

(a) *Because we are given only the mean and standard deviation, and no distribution is specified, we use Chebyshev's theorem. We have*

$$P\{\mu - K\sigma < X < \mu + K\sigma\} \geq 1 - \frac{1}{K^2}.$$

Equating $\mu + K\sigma$ to 168.6 and $\mu - K\sigma$ to 89.4 with $\mu = 129$ and $\sigma = 19.8$, we obtain $K = 2$. Hence,

$$P\{89.4 \leq X \leq 168.6\} \geq 1 - \frac{1}{(2)^2} = 0.75.$$

(b) *Because X is normally distributed with mean 129 and standard deviation 19.8, using the z-score, we get*

$$\begin{aligned} P(89.4 \leq X \leq 168.6) &= P\left(\frac{89.4 - 129}{19.8} \leq Z \leq \frac{168.6 - 129}{19.8}\right) \\ &= P(-2 \leq Z \leq 2) = 0.9544. \end{aligned}$$

Hence, approximately 95.44% of this population will have systolic blood pressure values between 89.4 and 168.6 mm Hg. This compares well with the 95% value from the empirical rule.

We could use Chebyshev's inequality to prove the following result, which is called the weak law of large numbers. The law of large numbers states that if the sample size n is large, the sample mean rarely deviates from the mean of the distribution of X, which in statistics is called the population mean.

LAW OF LARGE NUMBERS

Theorem 3.5.2 *Let $X_1, \ldots, X_n$ be a set of pairwise independent random variables with $E(X_i) = \mu$, and $var(X_i) = \sigma^2$. Then for any $c > 0$,*

$$P\left\{\mu - c \leq \overline{X} \leq \mu + c\right\} \geq 1 - \frac{\sigma^2}{nc^2}$$

and as $n \to \infty$, the probability approaches 1. Equivalently,

$$P\left(\left|\frac{S_n}{n} - \mu\right| < \varepsilon\right) \to 1$$

as $n \to \infty$.

Proof. Because $X_1, \ldots, X_n$ are iid random variables, we know that $Var(S_n) = n\sigma^2$, and $Var(S_n/n) = \sigma^2/n$. Also, $E(S_n/n) = \mu$. By Chebyshev's theorem, for any $\varepsilon > 0$,

$$P\left(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right) \leq \frac{\sigma^2}{n\varepsilon^2}.$$

Thus, for any fixed ε,

$$P\left(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right) \to 0$$

as $n \to \infty$. Equivalently,

$$P\left(\left|\frac{S_n}{n} - \mu\right| < \varepsilon\right) \to 1$$

as $n \to \infty$. □

Thus, without any knowledge of the probability distribution function of S_n, the (weak) law of large numbers states that the sample mean, $\overline{X} = S_n/n$, will differ from the population mean by less than an arbitrary constant, $\varepsilon > 0$, with probability that tends to 1 as n tends to ∞. Because of this, the law of large numbers is also called the "law of averages." This result basically states that we can start with a random experiment whose outcome cannot be predicted with certainty, and by taking averages, we can obtain an experiment in which the outcome can be predicted with a high degree of accuracy. The law of large numbers in its simplest form for the Bernoulli random variables was introduced by Jacob Bernoulli toward the end of the 16th century. This result in generality was first proved by the Russian mathematician A. Khintchine in 1929. This result is widely used in its applications to insurance, statistics, and the study of heredity.

Example 3.5.3

Let $X_1, \ldots, X_n$ be iid Bernoulli random variables with parameter p. Verify the law of large numbers.

Solution

For Bernoulli random variables we know that $EX_i = p$, and $Var(X_i) = p(1-p)$. Thus, by Chebyshev's theorem,

$$P\left\{p - c \leq \overline{X} \leq p + c\right\} = P\left\{\left|\frac{S_n}{n} - p\right| \leq c\right\} \geq 1 - \frac{\sigma^2}{nc^2}$$
$$= 1 - \frac{p(1-p)}{nc^2} \to 1, \quad \text{as } n \to \infty.$$

This verifies the weak law of large numbers.

Example 3.5.4

Consider n rolls of a balanced die. Let X_i be the outcome of the ith roll, and let $S_n = \sum_{i=1}^{n} X_i$. Show that, for any $\varepsilon > 0$,

$$P\left(\left|\frac{S_n}{n} - \frac{7}{2}\right| \geq \varepsilon\right) \to 0$$

as $n \to \infty$.

Solution

Because the die is balanced, $EX_i = 7/2$. By the law of large numbers, for any $\varepsilon > 0$,

$$P\left(\left|\frac{S_n}{n} - \frac{7}{2}\right| \geq \varepsilon\right) \to 0$$

as $n \to \infty$, or equivalently,

$$P\left(\left|\frac{S_n}{n} - \frac{7}{2}\right| < \varepsilon\right) \to 1$$

as $n \to \infty$.

One of the most important results in probability theory is the Central Limit Theorem. This basically states that the z-transform of the sample mean is asymptotically standard normal. The amazing thing about the Central Limit Theorem is that no matter what the shape of the original distribution is, the (sampling) distribution of the mean approaches a normal probability distribution. We state one version of the Central Limit Theorem. In a restricted case, the proof uses the idea that the moment-generating functions of Z_n converge to the moment-generating function of the standard normal random variable. The general proof is a little bit more involved. Because the proof of the Central Limit Theorem is available in most probability books, we will not give the proof here.

CENTRAL LIMIT THEOREM (CLT)

Theorem 3.5.3 *If $X_1, \ldots, X_n$ is a random sample from an infinite population with mean μ, variance σ^2, and the moment-generating function $M_X(t)$, then the limiting distribution of $Z_n = (\overline{X} - \mu)/(\sigma/\sqrt{n})$ as $n \to \infty$ is the standard normal probability distribution. That is,*

$$\lim_{n\to\infty} P(Z_n \leq z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2} dt.$$

If $S_n = \sum_{i=1}^{n} X_i$, then we can rewrite Z_n as

$$Z_n = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \frac{n\left(\overline{X} - \mu\right)}{n\sigma/\sqrt{n}},$$

$$= \frac{S_n - n\mu}{\sigma\sqrt{n}}, \quad \text{since } n\overline{X} = \sum_{i=1}^{n} X_i.$$

Then the CLT states that $Z_n = (S_n - n\mu)/\sigma\sqrt{n}$ is approximately $N(0, 1)$ for large n.

The Central Limit Theorem basically says that when we repeat an experiment a large number of times, the average (almost always) follows a Gaussian distribution.

Example 3.5.5
$X_1, X_2, \ldots$ are iid random variables such that

$$X_i = \begin{cases} 1, & \text{with probability } p, \\ 0, & \text{with probability } 1 - p. \end{cases}$$

Show that $Z_n = (S_n - np)/\sqrt{npq}$ is approximately normal for large n, where $S_n = \sum_{i=1}^{n} X_i$, and $q = 1 - p$.

Solution
We know that

$$E(X) = p;\ E(X^2) = p;\ Var(X) = p - p^2 = pq.$$

Hence, by the CLT, the limiting distribution of $Z_n = (S_n - np)/\sqrt{npq}$ as $n \to \infty$ is the standard normal probability distribution.

Example 3.5.6
A soft-drink vending machine is set so that the amount of drink dispensed is a random variable with a mean of 8 ounces and a standard deviation of 0.4 ounces. What is the approximate probability that the average of 36 randomly chosen fills exceed 8.1 ounces?

Solution
From the CLT, $((\overline{X} - 8)/(0.4/\sqrt{36})) \sim N(0, 1)$. Hence, from the normal table,

$$P\left\{\overline{X} > 8.1\right\} = P\left\{Z > \frac{8.1 - 8.0}{\frac{0.4}{\sqrt{36}}}\right\}$$
$$= p\{Z > 1.5\} = 0.0668.$$

Example 3.5.7
Numbers in decimal form are often approximated by the closest integers. Suppose n numbers $X_1, \ldots, X_n$ are approximated by their closest integers $J_1, J_2, \ldots, J_n$. Let $U_i = X_i - J_i$. Assume that U_i are uniform on $(-0.5, 0.5)$ and that $U_i's$ are independent.

(a) Show that $\dfrac{\sum_{i=1}^{n} U_i}{\sqrt{n/12}} \sim N(0, 1)$ as $n \to \infty$.

(b) Find $P\left\{\dfrac{-5}{\sqrt{300/12}} \le \dfrac{\sum_{i=1}^{n} U_i}{\sqrt{300/12}} \le \dfrac{5}{\sqrt{300/12}}\right\}$.

(c) Find the value of a such that $P\{-a \leq \sum U_i \leq a\} = 0.95$

(d) For $n = 10^6$, find a such that $P\left\{-a \leq \sum_{i=1}^{106} U_i \leq a\right\} = 0.99.$

Solution

(a) *Because $U_i's$ are uniform in $(-0.5, 0, 5)$, $\sum U_i = 0$, $Var(U_i) = 1/12$. Let, $S_n = \sum_{i=1}^{n} X_i$, and $K_n = \sum_{i=1}^{n} J_i$. Then*

$$P\{|S_n - K_n| \leq a\} = P\left\{-a \leq \sum (X_i - J_i) \leq a\right\}$$
$$= P\left\{-a \leq \sum U_i \leq a\right\}.$$

By the CLT, $\dfrac{\sum_{i=1}^{n} U_i - 0}{\sqrt{n/12}} \sim N(0,1)$ as $n \to \infty$.

(b) *For $n = 300$; $a = 5$. Using the normal table,*

$$P\left\{\frac{-5}{\sqrt{300/12}} \leq \frac{\sum_{i=1}^{n} U_i}{\sqrt{300/12}} \leq \frac{5}{\sqrt{300/12}}\right\} = 0.68.$$

(c) *Now,*

$$0.95 = P\left\{-a \leq \sum U_i \leq a\right\}$$
$$= P\left\{\frac{-a}{\sqrt{300/12}} \leq Z \leq \frac{a}{\sqrt{300/12}}\right\}.$$

From the normal table, we get $\dfrac{a}{\sqrt{300/12}} = 1.96$. *This implies, $a = 9.8$.*

(d) *We have*

$$0.99 = P\left\{-a \leq \sum_{i=1}^{10^6} U_i \leq a\right\}$$
$$= P\left\{\frac{-a}{\sqrt{10^6/12}} \leq Z \leq \frac{a}{\sqrt{10^6/12}}\right\}.$$

Now, using the normal table, we have $a/\sqrt{10^6/12} = 2.58$. Hence, $a = 745$. ■

Example 3.5.8

A casino has a coin, suspected to be biased. Estimate p (probability of heads) such that they can be confident that their estimate (say, $\hat{p}$) is within 0.01 of p (unknown). What is the minimum number of times we need to toss this coin?

Solution

Set

$$X_j = \begin{cases} 1, & \text{if } H \text{ as } j\text{'th toss,} \\ 0, & \text{if } T \text{ as } j\text{'th toss.} \end{cases}$$

Suppose we decided to use $\hat{p} = \frac{\sum X_i}{n}$, *that is,* $\left(\frac{\#Heads}{n}\right)$.

We want $P\{|\overline{X} - p| < 0.01\} = 0.99$.

Because $Y = \sum_{i=1}^{n} X_i \sim Bin(n, p)$, *we have* $EY = np$, $Var(Y) = npq$. *By the CLT,* $(\overline{X} - p)/\sqrt{pq/n} \sim N(0, 1)$. *Now,*

$$\begin{aligned} 0.99 &= P\left\{\frac{-0.01}{\sqrt{pq/n}} < \frac{\overline{X} - p}{\sqrt{pq/n}} < \frac{0.01}{\sqrt{pq/n}}\right\} \\ &= P\left\{\frac{-0.01}{\sqrt{pq/n}} < Z < \frac{0.01}{\sqrt{pq/n}}\right\}. \end{aligned}$$

Using the normal table, $(0.01/\sqrt{pq/n}) = 2.58$, *this implies that* $\sqrt{n} \geq (2.58\sqrt{pq}/0.01)$.

Because the maximum of $pq = 1/4$, *it is sufficient that*

$$\sqrt{n} = \frac{(2.58)(\sqrt{(1/4)})}{0.01} = 129.$$

Hence, $n = (129)^2 = 16{,}641$, *and we should choose the sample size* $n \geq 16{,}641$.

■

The Central Limit Theorem is extremely important in statistics because it says that we can approximate the distribution of certain statistics without much of the knowledge about the underlying distribution of that statistics for a relatively "large" sample size. How large the n should be for this normal approximation to work depends on the distribution of the original distribution. A rule of thumb is that the sample size n must be at least 30. We deal with these issues in Chapter 4.

EXERCISES 3.5

3.5.1. Let X be a random variable with probability density function

$$f(x) = \begin{cases} 630x^4(1-x)^4, & 0 < x < 1, \\ 0, & \text{otherwise.} \end{cases}$$

(a) Obtain the lower bound given by Chebyshev's inequality for $P\{0.2 < X < 0.8\}$.

(b) Compute the exact probability, $P\{0.2 < X < 0.8\}$.

3.5.2. Suppose that the number of cars arriving in 1 hour at a busy intersection is a Poisson probability distribution with $\lambda = 100$. Find, using Chebyshev's inequality, a lower bound for the probability that the number of cars arriving at the intersection in 1 hour is between 70 and 130.

3.5.3. Prove Chebyshev's inequality for the discrete case.

3.5.4. Suppose that the number of cars arriving at a busy intersection in a large city has a Poisson distribution with mean 120. Determine a lower bound for the probability that the number of cars arriving in a given 20-minute period will be between 100 and 140 using Chebyshev's inequality.

3.5.5. Find the smallest value of n in a binomial distribution for which we can assert that

$$P\left(\left|\frac{X_n}{n} - p\right| < 0.1\right) \geq 0.90.$$

3.5.6. How large should the size of a random sample be so that we can be 90% certain that the sample mean $\overline{X}$ will not deviate from the true mean by more than $\sigma/2$?

3.5.7. Let a fair coin be tossed n times and let S_n be the number of heads that turn up. Show that the fraction of heads, S_n/n, will be near to 1/2 for large n. What can we conclude if the coin is not fair?

3.5.8. Suppose that a failure of certain component follows the distribution $f(x) = p^x(1-p)^x$ for $x = 0, 1$, and zero, elsewhere. How many components must one test in order that the sample mean $\overline{X}$ will lie within 0.4 of the true state of nature with probability at least as great as 0.95?

3.5.9. Let $X_1, \ldots, X_n$ be a sequence of mutually independent random variables, with probability distribution

$$P(X_i = \sqrt{i}) = \frac{1}{2} \quad \text{and} \quad P(X_i = -\sqrt{i}) = \frac{1}{2}.$$

Show that this sequence of random variables does not satisfy the conditions of the law of large numbers.

3.5.10. Give a proof of the Central Limit Theorem.

3.5.11. Let $X_1, \ldots, X_n$ be independent discrete random variables identically distributed as

$$f(x_i) = \begin{cases} 0.2, & x_i = 0, 1, 2, 3, 4, \\ 0, & \text{otherwise.} \end{cases}$$

Using CTL, find the approximate value of $P(\overline{X}_{100} > 2)$, where $\overline{X}_{100} = (1/100)\sum_{i=1}^{100} X_i$.

3.5.12. Let $X_1, \ldots, X_n$ be a sequence of independent Poisson-distributed random variables, with parameter λ. Let $S_n = \sum_{i=1}^{n} X_i$. Show that $Z_n = ((S_n - n\lambda)/\sqrt{n\lambda}) \sim N(0, 1)$.

3.5.13. Let $X_1, \ldots, X_n$ be a sequence of independent uniformly-distributed over [0,1) random variables. Let $S_n = \sum_{i=1}^{n} X_i$. Show that $Z_n = ((S_n - n\lambda)/\sqrt{n\lambda}) \sim N(0, 1)$.

3.5.14. Suppose that 2500 customers subscribe to a telephone exchange. There are 80 trunk lines available. Any one customer has the probability of 0.03 of needing a trunk line on a given call. Consider the situation as 2500 trials with probability of "success" $p = 0.03$. What is the approximate probability that the 2500 customers will "tie up" the 80 trunk lines at any given time?

3.5.15. Suppose a group of people have an average IQ of 122 with standard deviation 2. Obtain a bound on the probability that IQ values of this group will be between 104 and 120.

3.5.16. Let X be a random variable that represents the diastolic blood pressure (DBP) of the population of 18- to 74-year-old men in the United States who are not taking any corrective medication. Suppose that X has mean 80.7 mm Hg and standard deviation 9.2.

(a) Obtain a bound on the probability that the DBP of this population will assumes values between 53.1 and 108.3 mm Hg.

(b) In addition, assume that the distribution of X is approximately normal. Using the normal table, find $P(53.1 \leq X \leq 108.3)$. Compare this with the empirical rule.

3.5.17. Color blindness appears in 2% of the people in a certain population. How large must a random sample be in order to be 99% certain that a color-blind person is included in the sample?

3.5.18. A shirt manufacturer knows that, on the average, 2% of his product will not meet quality specifications. Find the greatest number of shirts constituting a lot that will have, with probability 0.95, fewer than five defectives.

3.5.19. A random sample of size 100 is taken from a population with mean 1 and variance 0.04. Find the probability that the sample mean is between 0.99 and 1.

3.5.20. The lifetime X (in hours) of a certain electrical component has the pdf $f(x) = (1/3)e^{-(1/3)x}, x > 0$. If a random sample of 36 is taken from these components, find $P(\overline{X} < 2)$.

3.5.21. A drug manufacturer receives a shipment of 10,000 calibrated "eyedroppers" for administering the Sabin poliovirus vaccine. If the calibration mark is missing on 500 droppers, which are scattered randomly throughout the shipment, what is the probability that, at most, two defective droppers will be detected in a random sample of 125?

3.6 CHAPTER SUMMARY

In this chapter we looked at some special distribution functions that arise in practice. It should be noted that we discussed only a few of the important probability distributions. There many other discrete and continuous distributions that will be useful and appropriate in particular applications. Some of them are given in Appendix A3. A larger list of probability distributions can be found at http://www.causascientia.org/math_stat/Dists/Compendium.pdf, among many other places. For more than one random variable, we learned the joint distributions. We also saw how to find the density and cumulative distribution for the functions of a random variable. Limit theorems are a crucial part of probability theory. We have introduced the Chebyshev's inequality, the law of large numbers, and the Central Limit Theorem for the random variables.

We now list some of the key definitions introduced in this chapter:

- Bernoulli probability distribution
- Binomial experiment
- Poisson probability distribution
- Probability distribution
- Normal (or Gaussian) probability distribution
- Standard normal random variable
- Gamma probability distribution
- Exponential probability distribution
- Chi-square (χ^2) distribution
- Joint probability density function
- Bivariate probability distributions
- Marginal pdf
- Conditional probability distribution
- Independence of two r.v.s
- Expected value of a function of bivariate r.v.s
- Conditional expectation
- Covariance
- Correlation coefficient

In this chapter, we have also learned the following important concepts and procedures:

- Mean, variance, and moment-generating function (mgf) of a binomial random variable
- Mean, variance, and mgf of a Poisson random variable
- Poisson approximation to the binomial probability distribution
- Mean, variance, and mgf of a uniform random variable
- Mean, variance, and mgf of a normal random variable
- Mean, variance, and mgf of a gamma random variable
- Mean, variance, and mgf of an exponential random variable
- Mean, variance, and mgf of a chi-square random variable
- Properties of expected value
- Properties of the covariance and correlation coefficient
- Procedure to find the cdf of a function of r.v. using the method of distribution functions
- The pdf of $Y = g(X)$, where g is differentiable and monotone increasing or decreasing
- The pdf of $Y = g(X)$, using the probability integral transformation
- The transformation method to find the pdf of $Y = g(X_1, \ldots, X_n)$
- Chebyshev's theorem
- Law of large numbers
- Central Limit Theorem (CLT)

3.7 COMPUTER EXAMPLES (OPTIONAL)

3.7.1 Minitab Examples

Minitab contains subroutines that can do pdf and cdf computations. For example, for binomial random variables, the pdf and cdf can be respectively computed using the following comments.

```
MTB > pdf k;
SUBC > binomial n p.
```

and

```
MTB > cdf;
SUBC > binomial n p.
```

Practice: Try the following and see what you get.

```
MTB > pdf 3;
SUBC > binomial 5 0.40.
```

will give

K	$P(X = K)$
3.00	0.2304

and

```
MTB > cdf;
SUBC > binomial 5 0.40.
```

will give

```
BINOMIAL WITH N = 5 P = 0.400000
        K P(X LESS OR = K)
        0 0.0778
        1 0.3370
        2 0.6826
        3 0.9130
        4 0.9898
        5 1.0000
```

Similarly, if we want to calculate the cdf for a normal probability distribution with mean k and standard deviation s, use the following comments.

```
MTB > cdf x;
SUBC > normal k s.
```

will give $P(X \leq x)$.

Practice: Try the following.

```
MTB > cdf 4.20;
SUBC > normal 4 2.
```

We can use the invcdf command to find the inverse cdf. For a given probability p, $P(X \leq x) = F(x) = p$, we can find x for a given distribution. For example, for a normal probability distribution with mean k and standard deviation s, use the following.

```
MTB > invcdf p;
SUBC > normal k s.
```

We can also use the pull-down menus to compute the probabilities. The following example illustrates this for a binomial probability distribution.

Example 3.7.1

A manufacturer of a color printer claims that only 5% of their printers require repairs within the first year. If out of a random sample of 18 of their printers, four required repairs within the first year, does this tend to refute or support the manufacturer's claim? Use Minitab.

Solution

*Type the numbers 1 through 18 in **C1**. Then*

Calc** > **Probability Distributions** > **Binomial. . .** > choose **Cumulative probability** > in **Number of trials**, enter **18** and in **Probability of success**, enter **0.05** > in **Input column:** type **C1** > Click **OK

We will get the following output.

Cumulative Distribution Function

Binomial with n=18 and p=0.0500000

x	$P(X<=x)$
1.00	0.7735
2.00	0.9419
3.00	0.9891
4.00	0.9985
5.00	0.9998
6.00	1.0000
7.00	1.0000
8.00	1.0000
9.00	1.0000
10.00	1.0000
11.00	1.0000
12.00	1.0000
13.00	1.0000
14.00	1.0000
15.00	1.0000
16.00	1.0000
17.00	1.0000
18.00	1.0000

The required probability is $P(X \geq 4) = 1 - P(X \leq 3) = 1 - 0.9891 = 0.0109$.

3.7.2 SPSS Examples

Example 3.7.2

For the data of Example 3.7.1, using SPSS, find $P(X \leq 3)$.

Solution

Enter numbers 1 through 18 in ***C1****. Then use the following.*

Transform > Compute > type in the ***Target Variable: y >*** Use the scroll bar beside the Functions box to find ***CDF.BINOM(q, n, p) >*** Highlight it and use the up button to load it into the ***Numeric Expression:*** box. Set ***q*** to ***3*** (success, the x-value), ***n*** to ***18*** (total trials) and ***p*** to ***0.05*** (probability of success) ***> OK***

In the second column, we will get the y-values as 0.99. Hence, $P(X \leq 3) = 0.99$.

We can use this procedure for many other distributions.

3.7.3 SAS Examples

Sometimes, we can use computer calculations to find out the exact probability of a certain event in lieu of approximations. For example, when n is large in a binomial experiment, we can use normal approximation to calculate the probabilities. The following example shows how to calculate binomial probabilities using SAS codes.

Example 3.7.3

Suppose that a certain drug to treat a disease has a success rate of $p = 0.65$. This drug is given to $n = 500$ patients with the disease.

(a) What is the probability that 335 or fewer show improvement?
(b) What is the probability that more than 320 show improvement?
(c) What is the probability that exactly 300 show improvement?
(d) What is the probability that the number of improvements lies in the interval (300,350)?

Solution

Let $X =$ *number of patients showing improvement. Then* X *is a binomial random variable with parameters* $n = 500$ *and* $p = 0.65$.

(a) *First three lines in the following code are comment lines. In general, it is always helpful to include the comment lines to explain about the program.*

```
/*This program can be used to compute probability*/
/* that a Binomial variable with parameters p*/
/*and n is less than or equal to x*/
data binomial;
     p=0.65;
     n=500;
     x=335;
     y=probbnml(p,n,x);
cards;
proc print;
run;
```

The following is the SAS output from running the foregoing program.

Obs	p	n	x	y
1	0.65	500	335	0.83753

Here $y = 0.83753$ *is the* $P(X \leq 335)$.

(b) *To calculate* $P(X > 320)$, *we can use the following.*

```
data binomial;
     p=0.65;
     n=500;
```

```
     x=320;
     y=probbnml(p,n,x);
     z=1-y;
cards;
proc print;
run;
```

The following is the SAS output from running the foregoing program, where the value of z *is the probability we are looking for.*

Obs	p	n	x	y	z
1	0.65	500	320	0.33516	0.66484

Hence, $P(X > 320) = 0.66484$.

(c) *To find* $P(X = 300)$, *we can use the following.*

```
data binomial;
     p=0.65;
     n= 500;
     x1=300;
     y1=probbnml(p,n,x1);
     x2=299;
     y2=probbnml(p,n,x2);
     z=y1-y2;
cards;
proc print;
run;
```

The following is the SAS output from running the foregoing program, where the value of z *is the probability we are looking for.*

Obs	p	n	x1	y1	x2	y2	z
1	0.65	500	300	0.011327	299	.008864418	.002462253

(d) *To find* $P(300 < X < 350)$, *use the following.*

```
data binomial;
     p=0.65;
     n=500;
     x1=300;
     y1=probbnml(p,n,x1);
     x2=349;
     y2=probbnml(p,n,x2);
```

```
     z=y2-y1;
cards;
proc print;
run;
```

We will get the following output.

Obs	p	n	x1	y1	x2	y2	z
1	0.65	500	300	0.011327	349	0.98982	0.97849

Hence, $P(300 < X < 350) = 0.97849$.

Similar procedures could be used to calculate probabilities for other distributions.

In order to *test for normality* of a given data set using a normal probability plot, we can use PROC UNIVARIATE (see Chapter 1 for explanation) in the following manner. Normal plot is called qqplot in SAS.

```
proc univariate data=K noprint; /*Specify the name of data set as K*/
qqplot standard;
run;
quit;
```

Note that this avoids printing of all the standard output due to the univariate command, and we get only the QQ plot. If we need a straight line in the plot, we can modify the commands as follows.

```
proc univariate data=K noprint; /*Specify the name of data set as B*/
qqplot standard/ normal (mu=m, sigma=s);
run;
quit;
```

PROJECTS FOR CHAPTER 3

3A. Mixture Distribution

In statistical modeling, if the data are contaminated by outliers or if the samples are drawn from a population formed by a mixture of two populations, one could use mixture distributions. Mixture distributions are used frequently in medical applications, such as micro array analysis. Suppose a random variable X has pdf $f_1(x)$ with probability p_1 and pdf $f_2(x)$ with probability p_2, where $p_1 + p_2 = 1$. Then we say that the r.v. X has a *mixture distribution*. This can be thought of as observing a Bernoulli random variable Z that is equal to 1 with probability p_1 and 2 with probability p_2. Thus,

$$X = \begin{cases} X_1 \sim f_1(x), & \text{if } Y = 1, \\ X_2 \sim f_2(x), & \text{if } Y = 2. \end{cases}$$

(a) Show that the pdf of X is given by $f(x) = p_1 f_1(x) + p_2 f_2(x)$.

(b) If (μ_1, σ_1^2) and (μ_2, σ_2^2) are means and variances of $f_1(x)$ and $f_2(x)$, respectively, show that

$$\mu = E(X) = p_1\mu_1 + p_2\mu_2,$$

and

$$\sigma^2 = Var(X) = p_1\sigma_1^2 + p_2\sigma_1^2 + p_1\mu_1^2 + p_2\mu_2^2 - (p_1\mu_1 + p_2\mu_2)^2.$$

3B. Generating Samples from Exponential and Poisson Probability Distribution

(a) Generate a sample from $\frac{1}{\theta}e^{-x/\theta}$ (θ is chosen). Let $Y_1, Y_2, \ldots, Y_n$ be a sample from a $U(0, 1)$ distribution. Let $F(x) = 1-e^{-x/\theta}$ (cdf of exponential). Then $Y = F(x)$ is uniform. $y_j = 1-e^{-x/\theta}$ implies $x_j = -\theta \ln(1 - y_i) = -\theta \ln u_i$, where $u_1, u_2, \ldots, u_n$ is a sample from $U(0, 1)$. Then $X_1, \ldots, X_n$ is a sample from an exponential distribution with parameter θ.

(b) Suppose we want to generate a sample from a Poisson probability distribution with parameter λ. $X_1, \ldots, X_n$ is a sample from an exponential distribution with parameter $1/\lambda$ till $\sum_{i=1}^{n} X_i$ just exceeds 1. Then $y_n(n-1)$ is a sample values form a Poisson probability distribution with parameter λ.

EXERCISE 3B

Let $u_1, u_2, \ldots, u_n$ be a sample from $U(0, 1)$. Show that

(i) $X = -2\sum\limits_{i=1}^{n} \ln(u_i) \sim \chi^2_{2n}$,

(ii) $X = -\beta\sum\limits_{i=1}^{\alpha} \ln(u_i) \sim gamma(\alpha, \beta)$, and

(iii) $X = \dfrac{\sum_{i=1}^{\alpha} \ln(u_i)}{\sum_{i=1}^{\alpha+\beta} \ln(u_i)} \sim Beta(\alpha, \beta)$.

3C. Coupon Collector's Problem

Suppose there are n distinct colors of coupons. Each color of coupon is equally likely to occur. When a complete set of coupons with each color represented is assembled, you win a prize. Let X = # coupons for a complete set. Find **(a)** Distribution of X, **(b)** $E(X)$, and **(c)** $Var(X)$.

3D. Recursive Calculation of Binomial and Poisson Probabilities

A simple way to calculate binomial probabilities is as follows: For a given n and p, evaluate $b(0, n, p)$ and then apply the recursive relationship

$$b(x+1, n, p) = b(x, n, p)\frac{p(n-x)}{(1-p)(x+1)}$$

to obtain other binomial probabilities.

(a) Derive this recursion formula.

(b) For $n = 15$, $p = 0.4$, using the recursive formula, compute all other probabilities starting from $x = 0$.

The following recursive formulas are very useful in calculating successive Poisson probabilities:

$$f(x-1, \lambda) = f(x, \lambda)\frac{x}{\lambda}$$

and

$$f(x+1, \lambda) = \frac{e^{-\lambda}\lambda^{x+1}}{(x+1)!} = f(x, \lambda)\frac{\lambda}{x+1}.$$

For example, if $\lambda = 2.5$, we know that $f(0, 2.5) = e^{-2.5} = 0.08208$. Using this, calculate (c) $f(1, 2.5)$ and $f(2, 2.5)$.