One Random Variable

CHAPTER 4

In Chapter 3 we introduced the notion of a random variable and we developed methods for calculating probabilities and averages for the case where the random variable is discrete. In this chapter we consider the general case where the random variable may be discrete, continuous, or of mixed type. We introduce the cumulative distribution function which is used in the formal definition of a random variable, and which can handle all three types of random variables. We also introduce the probability density function for continuous random variables. The probabilities of events involving a random variable can be expressed as integrals of its probability density function. The expected value of continuous random variables is also introduced and related to our intuitive notion of average. We develop a number of methods for calculating probabilities and averages that are the basic tools in the analysis and design of systems that involve randomness.

4.1 THE CUMULATIVE DISTRIBUTION FUNCTION

The probability mass function of a discrete random variable was defined in terms of events of the form $\{X = b\}$. The cumulative distribution function is an alternative approach which uses events of the form $\{X \le b\}$. The cumulative distribution function has the advantage that it is not limited to discrete random variables and applies to all types of random variables. We begin with a formal definition of a random variable.

> **Definition:** Consider a random experiment with sample space S and event class $\mathcal{F}$. A **random variable** X is a function from the sample space S to R with the property that the set $A_b = \{\zeta : X(\zeta) \le b\}$ is in $\mathcal{F}$ for every b in R.

The definition simply requires that every set A_b have a well defined probability in the underlying random experiment, and this is not a problem in the cases we will consider. Why does the definition use sets of the form $\{\zeta : X(\zeta) \le b\}$ and not $\{\zeta : X(\zeta) = \mathrm{x}_b\}$? We will see that all events of interest in the real line can be expressed in terms of sets of the form $\{\zeta : X(\zeta) \le b\}$.

The **cumulative distribution function** (cdf) of a random variable X is defined as the probability of the event $\{X \le x\}$:

$$F_X(x) = P[X \le x] \qquad \text{for } -\infty < x < +\infty, \tag{4.1}$$

that is, it is the probability that the random variable X takes on a value in the set $(-\infty, x]$. In terms of the underlying sample space, the cdf is the probability of the event $\{\zeta: X(\zeta) \le x\}$. The event $\{X \le x\}$ and its probability vary as x is varied; in other words, $F_X(x)$ is a function of the variable x.

The cdf is simply a convenient way of specifying the probability of all semi-infinite intervals of the real line of the form $(-\infty, b]$. The events of interest when dealing with numbers are intervals of the real line, and their complements, unions, and intersections. We show below that the probabilities of all of these events can be expressed in terms of the cdf.

The cdf has the following interpretation in terms of relative frequency. Suppose that the experiment that yields the outcome ζ, and hence $X(\zeta)$, is performed a large number of times. $F_X(b)$ is then the long-term proportion of times in which $X(\zeta) \le b$.

Before developing the general properties of the cdf, we present examples of the cdfs for three basic types of random variables.

Example 4.1 Three Coin Tosses

Figure 4.1(a) shows the cdf X, the number of heads in three tosses of a fair coin. From Example 3.1 we know that X takes on only the values 0, 1, 2, and 3 with probabilities 1/8, 3/8, 3/8, and 1/8, respectively, so $F_X(x)$ is simply the sum of the probabilities of the outcomes from $\{0, 1, 2, 3\}$ that are less than or equal to x. The resulting cdf is seen to be a nondecreasing staircase function that grows from 0 to 1. The cdf has jumps at the points 0, 1, 2, 3 of magnitudes 1/8, 3/8, 3/8, and 1/8, respectively.

Let us take a closer look at one of these discontinuities, say, in the vicinity of $x = 1$. For δ a small positive number, we have

$$F_X(1 - \delta) = P[X \le 1 - \delta] = P\{0 \text{ heads}\} = \frac{1}{8}$$

so the limit of the cdf as x approaches 1 from the left is 1/8. However,

$$F_X(1) = P[X \le 1] = P[0 \text{ or } 1 \text{ heads}] = \frac{1}{8} + \frac{3}{8} = \frac{1}{2},$$

and furthermore the limit from the right is

$$F_X(1 + \delta) = P[X \le 1 + \delta] = P[0 \text{ or } 1 \text{ heads}] = \frac{1}{2}.$$

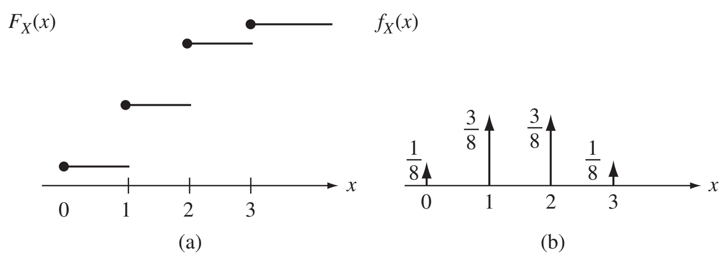


FIGURE 4.1
cdf (a) and pdf (b) of a discrete random variable.

Thus the cdf is continuous from the right and equal to 1/2 at the point $x = 1$. Indeed, we note the magnitude of the jump at the point $x = 1$ is equal to $P[X = 1] = 1/2 - 1/8 = 3/8$. Henceforth we will use dots in the graph to indicate the value of the cdf at the points of discontinuity.

The cdf can be written compactly in terms of the unit step function:

$$u(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \geq 0, \end{cases} \tag{4.2}$$

then

$$F_X(x) = \frac{1}{8}u(x) + \frac{3}{8}u(x-1) + \frac{3}{8}u(x-2) + \frac{1}{8}u(x-3).$$

Example 4.2 Uniform Random Variable in the Unit Interval

Spin an arrow attached to the center of a circular board. Let θ be the final angle of the arrow, where $0 < \theta \leq 2\pi$. The probability that θ falls in a subinterval of $(0, 2\pi]$ is proportional to the length of the subinterval. The random variable X is defined by $X(\theta) = \theta/2\pi$. Find the cdf of X:

As θ increases from 0 to 2π, X increases from 0 to 1. No outcomes θ lead to values $x \leq 0$, so

$$F_X(x) = P[X \leq x] = P[\varnothing] = 0 \qquad \text{for } x < 0.$$

For $0 < x \leq 1$, $\{X \leq x\}$ occurs when $\{\theta \leq 2\pi x\}$ so

$$F_X(x) = P[X \leq x] = P[\{\theta \leq 2\pi x\}] = 2\pi x/2\pi = x \qquad 0 < x \leq 1. \tag{4.3}$$

Finally, for $x > 1$, all outcomes θ lead to $\{X(\theta) \leq 1 < x\}$, therefore:

$$F_X(x) = P[X \leq x] = P[0 < \theta \leq 2\pi] = 1 \qquad \text{for } x > 1.$$

We say that X is a **uniform random variable** in the unit interval. Figure 4.2(a) shows the cdf of the general uniform random variable X. We see that $F_X(x)$ is a nondecreasing continuous function that grows from 0 to 1 as x ranges from its minimum values to its maximum values.

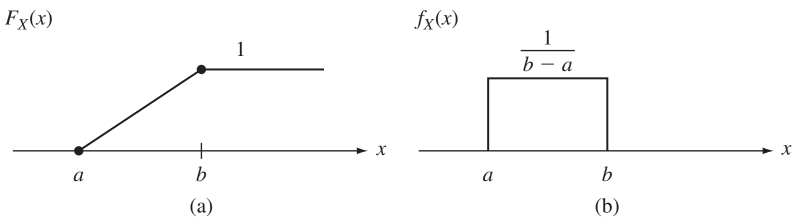


FIGURE 4.2
cdf (a) and pdf (b) of a continuous random variable.

Example 4.3

The waiting time X of a customer at a taxi stand is zero if the customer finds a taxi parked at the stand, and a uniformly distributed random length of time in the interval $[0, 1]$ (in hours) if no taxi is found upon arrival. The probability that a taxi is at the stand when the customer arrives is p. Find the cdf of X.

The cdf is found by applying the theorem on total probability:

$$F_X(x) = P[X \le x] = P[X \le x \,|\, \text{find taxi}]p + P[X \le x \,|\, \text{no taxi}](1 - p).$$

Note that $P[X \le x \,|\, \text{find taxi}] = 1$ when $x \ge 0$ and 0 otherwise. Furthermore $P[X \le x \,|\, \text{no taxi}]$ is given by Eq. (4.3), therefore

$$F_X(x) = \begin{cases} 0 & x < 0 \\ p + (1 - p)x & 0 \le x \le 1 \\ 1 & x > 1. \end{cases}$$

The cdf, shown in Fig. 4.3(a), combines some of the properties of the cdf in Example 4.1 (discontinuity at 0) and the cdf in Example 4.2 (continuity over intervals). Note that $F_X(x)$ can be expressed as the sum of a step function with amplitude p and a continuous function of x.

We are now ready to state the basic properties of the cdf. The axioms of probability and their corollaries imply that the cdf has the following properties:

(i) $0 \le F_X(x) \le 1$.

(ii) $\lim_{x \to \infty} F_X(x) = 1$.

(iii) $\lim_{x \to -\infty} F_X(x) = 0$.

(iv) $F_X(x)$ is a nondecreasing function of x, that is, if $a < b$, then $F_X(a) \le F_X(b)$.

(v) $F_X(x)$ is continuous from the right, that is, for $h > 0$, $F_X(b) = \lim_{h \to 0} F_X(b + h) = F_X(b^+)$.

These five properties confirm that, in general, the cdf is a nondecreasing function that grows from 0 to 1 as x increases from $-\infty$ to ∞. We already observed these properties in Examples 4.1, 4.2, and 4.3. Property (v) implies that at points of discontinuity, the cdf

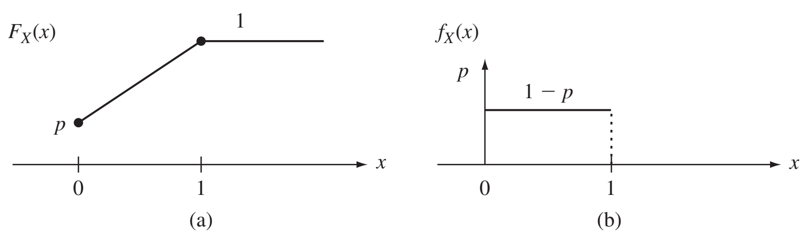


FIGURE 4.3
cdf (a) and pdf (b) of a random variable of mixed type.

is equal to the limit from the right. We observed this property in Examples 4.1 and 4.3. In Example 4.2 the cdf is continuous for all values of x, that is, the cdf is continuous both from the right and from the left for all x.

The cdf has the following properties which allow us to calculate the probability of events involving intervals and single values of X:

(vi) $P[a < X \leq b] = F_X(b) - F_X(a)$.

(vii) $P[X = b] = F_X(b) - F_X(b^-)$.

(viii) $P[X > x] = 1 - F_X(x)$.

Property (vii) states that the probability that $X = b$ is given by the magnitude of the jump of the cdf at the point b. This implies that *if the cdf is continuous at a point b, then* $P[X = b] = 0$. Properties (vi) and (vii) can be combined to compute the probabilities of other types of intervals. For example, since $\{a \leq X \leq b\} = \{X = a\} \cup \{a < X \leq b\}$, then

$$\begin{aligned} P[a \leq X \leq b] &= P[X = a] + P[a < X \leq b] \\ &= F_X(a) - F_X(a^-) + F_X(b) - F_X(a) = F_X(b) - F_X(a^-). \end{aligned} \tag{4.4}$$

If the cdf is continuous at the endpoints of an interval, then the endpoints have zero probability, and therefore they can be included in, or excluded from, the interval without affecting the probability.

Example 4.4

Let X be the number of heads in three tosses of a fair coin. Use the cdf to find the probability of the events $A = \{1 < X \leq 2\}$, $B = \{0.5 \leq X < 2.5\}$, and $C = \{1 \leq X < 2\}$.

From property (vi) and Fig. 4.1 we have

$$P[1 < X \leq 2] = F_X(2) - F_X(1) = 7/8 - 1/2 = 3/8.$$

The cdf is continuous at $x = 0.5$ and $x = 2.5$, so

$$P[0.5 \leq X < 2.5] = F_X(2.5) - F_X(0.5) = 7/8 - 1/8 = 6/8.$$

Since $\{1 \leq X < 2\} \cup \{X = 2\} = \{1 \leq X \leq 2\}$, from Eq. (4.4) we have

$$P\{1 \leq X < 2] + P[X = 2] = F_X(2) - F_X(1^-),$$

and using property (vii) for $P[X = 2]$:

$$\begin{aligned} P\{1 \leq X < 2] &= F_X(2) - F_X(1^-) - \mathrm{P}[X = 2] = F_X(2) - F_X(1^-) - (F_X(2) - F_X(2^-)) \\ &= F_X(2^-) - F_X(1^-) = 4/8 - 1/8 = 3/8. \end{aligned}$$

Example 4.5

Let X be the uniform random variable from Example 4.2. Use the cdf to find the probability of the events $\{-0.5 < X < 0.25\}$, $\{0.3 < X < 0.65\}$, and $\{|X - 0.4| > 0.2\}$.

The cdf of X is continuous at every point so we have:

$$P[-0.5 < X \le 0.25] = F_X(0.25) - F_X(-0.5) = 0.25 - 0 = 0.25,$$

$$P[0.3 < X < 0.65] = F_X(0.65) - F_X(0.3) = 0.65 - 0.3 = 0.35,$$

$$\begin{aligned} P[|X - 0.4| > 0.2] &= P[\{X < 0.2\} \cup \{X > 0.6] = P[X < 0.2] + P[X > 0.6] \\ &= F_X(0.2) + (1 - F_X(0.6)) = 0.2 + 0.4 = 0.6. \end{aligned}$$

We now consider the proof of the properties of the cdf.

- Property (i) follows from the fact that the cdf is a probability and hence must satisfy Axiom I and Corollary 2.
- To obtain property (iv), we note that the event $\{X \le a\}$ is a subset of $\{X \le b\}$, and so it must have smaller or equal probability (Corollary 7).
- To show property (vi), we note that $\{X \le b\}$ can be expressed as the union of mutually exclusive events: $\{X \le a\} \cup \{a < X \le b\} = \{X \le b\}$, and so by Axiom III, $F_X(a) + P[a < X \le b] = F_X(b)$.
- Property (viii) follows from $\{X > x\} = \{X \le x\}^c$ and Corollary 1.

While intuitively clear, properties (ii), (iii), (v), and (vii) require more advanced limiting arguments that are discussed at the end of this section.

4.1.1 The Three Types of Random Variables

The random variables in Examples 4.1, 4.2, and 4.3 are typical of the three most basic types of random variable that we are interested in.

Discrete random variables have a cdf that is a right-continuous, staircase function of x, with jumps at a countable set of points $x_0, x_1, x_2, \ldots$. The random variable in Example 4.1 is a typical example of a discrete random variable. The cdf $F_X(x)$ of a discrete random variable is the sum of the probabilities of the outcomes less than x and can be written as the weighted sum of unit step functions as in Example 4.1:

$$F_X(x) = \sum_{x_k \le x} p_X(x_k) = \sum_k p_X(x_k) u(x - x_k), \tag{4.5}$$

where the pmf $p_X(x_k) = P[X = x_k]$ gives the magnitude of the jumps in the cdf. We see that the pmf can be obtained from the cdf and vice versa.

A **continuous random variable** is defined as a random variable whose cdf $F_X(x)$ is continuous everywhere, and which, in addition, is sufficiently smooth that it can be written as an integral of some nonnegative function $f(x)$:

$$F_X(x) = \int_{-\infty}^{x} f(t)\, dt. \tag{4.6}$$

The random variable discussed in Example 4.2 can be written as an integral of the function shown in Fig. 4.2(b). The continuity of the cdf and property (vii) implies that continuous random variables have $P[X = x] = 0$ for all x. Every possible outcome has probability zero! An immediate consequence is that the pmf cannot be used to characterize the probabilities of X. A comparison of Eqs. (4.5) and (4.6) suggests how we can proceed to characterize continuous random variables. For discrete random variables, (Eq. 4.5), we calculate probabilities as summations of probability masses at discrete points. For continuous random variables, (Eq. 4.6), we calculate probabilities as integrals of "probability densities" over intervals of the real line.

A **random variable of mixed type** is a random variable with a cdf that has jumps on a countable set of points $x_0, x_1, x_2, \ldots$, but that also increases continuously over at least one interval of values of x. The cdf for these random variables has the form

$$F_X(x) = pF_1(x) + (1 - p)F_2(x),$$

where $0 < p < 1$, and $F_1(x)$ is the cdf of a discrete random variable and $F_2(x)$ is the cdf of a continuous random variable. The random variable in Example 4.3 is of mixed type.

Random variables of mixed type can be viewed as being produced by a two-step process: A coin is tossed; if the outcome of the toss is heads, a discrete random variable is generated according to $F_1(x)$; otherwise, a continuous random variable is generated according to $F_2(x)$.

*4.1.2 Fine Point: Limiting properties of cdf

Properties (ii), (iii), (v), and (vii) require the continuity property of the probability function discussed in Section 2.9. For example, for property (ii), we consider the sequence of events $\{X \le n\}$ which increases to include all of the sample space S as n approaches ∞, that is, all outcomes lead to a value of X less than infinity. The continuity property of the probability function (Corollary 8) implies that:

$$\lim_{n\to\infty} F_X(n) = \lim_{n\to\infty} P[X \le n] = P[\lim_{n\to\infty}\{X \le n\}] = P[S] = 1.$$

For property (iii), we take the sequence $\{X \le -n\}$ which decreases to the empty set $\varnothing$, that is, no outcome leads to a value of X less than $-\infty$:

$$\lim_{n\to\infty} F_X(-n) = \lim_{n\to\infty} P[X \le -n] = P[\lim_{n\to\infty}\{X \le -n\}] = P[\varnothing] = 0.$$

For property (v), we take the sequence of events $\{X \le x + 1/n\}$ which decreases to $\{X \le x\}$ from the right:

$$\begin{aligned}\lim_{n\to\infty} F_X(x + 1/n) &= \lim_{n\to\infty} P[X \le x + 1/n] \\ &= P[\lim_{n\to\infty}\{X \le x + 1/n\}] = P[\{X \le x\}] = F_X(x).\end{aligned}$$

Finally, for property (vii), we take the sequence of events, $\{b - 1/n < X \le b\}$ which decreases to $\{b\}$ from the left:

$$\begin{aligned}\lim_{n\to\infty} (F_X(b) - F_X(b - 1/n)) &= \lim_{n\to\infty} P[b - 1/n < X \le b] \\ &= P[\lim_{n\to\infty}\{b - 1/n < X \le b\}] = P[X = b].\end{aligned}$$

4.2 THE PROBABILITY DENSITY FUNCTION

The **probability density function of X** (pdf), if it exists, is defined as the derivative of $F_X(x)$:

$$f_X(x) = \frac{dF_X(x)}{dx}. \tag{4.7}$$

In this section we show that the pdf is an alternative, and more useful, way of specifying the information contained in the cumulative distribution function.

The pdf represents the "density" of probability at the point x in the following sense: The probability that X is in a small interval in the vicinity of x—that is, $\{x < X \le x + h\}$—is

$$\begin{aligned} P[x < X \le x + h] &= F_X(x + h) - F_X(x) \\ &= \frac{F_X(x + h) - F_X(x)}{h} h. \end{aligned} \tag{4.8}$$

If the cdf has a derivative at x, then as h becomes very small,

$$P[x < X \le x + h] \simeq f_X(x)h. \tag{4.9}$$

Thus $f_X(x)$ represents the "density" of probability at the point x in the sense that the probability that X is in a small interval in the vicinity of x is approximately $f_X(x)h$. The derivative of the cdf, when it exists, is positive since the cdf is a nondecreasing function of x, thus

(i) $f_X(x) \ge 0.$ (4.10)

Equations (4.9) and (4.10) provide us with an alternative approach to specifying the probabilities involving the random variable X. We can begin by stating a nonnegative function $f_X(x)$, called the probability density function, which specifies the probabilities of events of the form "X falls in a small interval of width dx about the point x," as shown in Fig. 4.4(a). The probabilities of events involving X are then expressed in terms of the pdf by adding the probabilities of intervals of width dx. As the widths of the intervals approach zero, we obtain an integral in terms of the pdf. For example, the probability of an interval $[a, b]$ is

(ii) $P[a \le X \le b] = \int_a^b f_X(x)\, dx.$ (4.11)

The probability of an interval is therefore the area under $f_X(x)$ in that interval, as shown in Fig. 4.4(b). The probability of any event that consists of the union of disjoint intervals can thus be found by adding the integrals of the pdf over each of the intervals.

The cdf of X can be obtained by integrating the pdf:

(iii) $F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt.$ (4.12)

In Section 4.1, we defined a *continuous random variable* as a random variable X whose cdf was given by Eq. (4.12). Since the probabilities of all events involving X can be written in terms of the cdf, it then follows that these probabilities can be written in

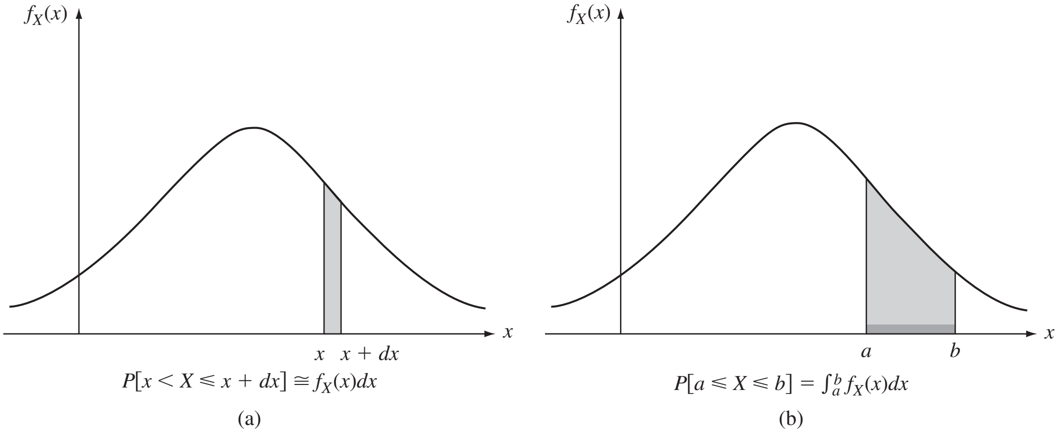


FIGURE 4.4
(a) The probability density function specifies the probability of intervals of infinitesimal width. (b) The probability of an interval $[a, b]$ is the area under the pdf in that interval.

terms of the pdf. Thus *the pdf completely specifies the behavior of continuous random variables*.

By letting x tend to infinity in Eq. (4.12), we obtain a *normalization* condition for pdf's:

$$\textbf{(iv)} \quad 1 = \int_{-\infty}^{+\infty} f_X(t)\, dt. \tag{4.13}$$

The pdf reinforces the intuitive notion of probability as having attributes similar to "physical mass." Thus Eq. (4.11) states that the probability "mass" in an interval is the integral of the "density of probability mass" over the interval. Equation (4.13) states that the total mass available is one unit.

A valid pdf can be formed from any nonnegative, piecewise continuous function $g(x)$ that has a finite integral:

$$\int_{-\infty}^{\infty} g(x)\, dx = c < \infty. \tag{4.14}$$

By letting $f_X(x) = g(x)/c$, we obtain a function that satisfies the normalization condition. Note that the pdf must be defined for all real values of x; if X does not take on values from some region of the real line, we simply set $f_X(x) = 0$ in the region.

Example 4.6 Uniform Random Variable

The pdf of the uniform random variable is given by:

$$f_X(x) = \begin{cases} \dfrac{1}{b - a} & a \le x \le b \\ 0 & x < a \quad \text{and} \quad x > b \end{cases} \tag{4.15a}$$

and is shown in Fig. 4.2(b). The cdf is found from Eq. (4.12):

$$F_X(x) = \begin{cases} 0 & x < a \\ \dfrac{x-a}{b-a} & a \le x \le b \\ 1 & x > b. \end{cases} \tag{4.15b}$$

The cdf is shown in Fig. 4.2(a).

Example 4.7 Exponential Random Variable

The transmission time X of messages in a communication system has an exponential distribution:

$$P[X > x] = e^{-\lambda x} \qquad x > 0.$$

Find the cdf and pdf of X.

The cdf is given by $F_X(x) = 1 - P[X > x]$

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \ge 0. \end{cases} \tag{4.16a}$$

The pdf is obtained by applying Eq. (4.7):

$$f_X(x) = F'_X(x) = \begin{cases} 0 & x < 0 \\ \lambda e^{-\lambda x} & x \ge 0. \end{cases} \tag{4.16b}$$

Example 4.8 Laplacian Random Variable

The pdf of the samples of the amplitude of speech waveforms is found to decay exponentially at a rate α, so the following pdf is proposed:

$$f_X(x) = ce^{-\alpha|x|} \qquad -\infty < x < \infty. \tag{4.17}$$

Find the constant c, and then find the probability $P[|X| < v]$.

We use the normalization condition in (iv) to find c:

$$1 = \int_{-\infty}^{\infty} ce^{-\alpha|x|}\,dx = 2\int_0^{\infty} ce^{-\alpha x}\,dx = \frac{2c}{\alpha}.$$

Therefore $c = \alpha/2$. The probability $P[|X| < v]$ is found by integrating the pdf:

$$P[|X| < v] = \frac{\alpha}{2}\int_{-v}^{v} e^{-\alpha|x|}\,dx = 2\left(\frac{\alpha}{2}\right)\int_0^{v} e^{-\alpha x}\,dx = 1 - e^{-\alpha v}.$$

4.2.1 pdf of Discrete Random Variables

The derivative of the cdf does not exist at points where the cdf is not continuous. Thus the notion of pdf as defined by Eq. (4.7) does not apply to discrete random variables at the points where the cdf is discontinuous. We can generalize the definition of the probability density function by noting the relation between the unit step function and the delta function. The **unit step function** is defined as

$$u(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0. \end{cases} \tag{4.18a}$$

The **delta function** $\delta(t)$ is related to the unit step function by the following equation:

$$u(x) = \int_{-\infty}^{x} \delta(t)\, dt. \tag{4.18b}$$

A translated unit step function is then:

$$u(x - x_0) = \int_{-\infty}^{x-x_0} \delta(t)\, dt = \int_{-\infty}^{x} \delta(t' - x_0)\, dt'. \tag{4.18c}$$

Substituting Eq. (4.18c) into the cdf of a discrete random variables:

$$\begin{aligned} F_X(x) &= \sum_k p_X(x_k) u(x - x_k) = \sum_k p_X(x_k) \int_{-\infty}^{x} \delta(t - x_k)\, dt \\ &= \int_{-\infty}^{x} \sum_k p_X(x_k) \delta(t - x_k)\, dt. \end{aligned} \tag{4.19}$$

This suggests that we define the **pdf for a discrete random variable** by

$$f_X(x) = \frac{d}{dx} F_X(x) = \sum_k p_X(x_k) \delta(x - x_k). \tag{4.20}$$

Thus the generalized definition of pdf places a delta function of weight $P[X = x_k]$ at the points x_k where the cdf is discontinuous.

To provide some intuition on the delta function, consider a narrow rectangular pulse of unit area and width Δ centered at $t = 0$:

$$\pi_\Delta(t) = \begin{cases} 1/\Delta & -\Delta/2 \leq t \leq \Delta/2 \\ 0 & |t| > \Delta. \end{cases}$$

Consider the integral of $\pi_\Delta(t)$:

$$\int_{-\infty}^{x} \pi_\Delta(t)\, dt = \left\{ \begin{array}{ll} \displaystyle\int_{-\infty}^{x} \pi_\Delta(t)\, dt = \int_{-\infty}^{x} 0\, dt = 0 & \text{for } x < -\Delta/2 \\[2ex] \displaystyle\int_{-\infty}^{x} \pi_\Delta(t)\, dt = \int_{-\Delta/2}^{\Delta/2} 1/\Delta\, dt = 1 & \text{for } x > \Delta/2 \end{array} \right\} \to u(x). \tag{4.21}$$

As $\Delta \to 0$, we see that the integral of the narrow pulse approaches the unit step function. For this reason, we visualize the delta function $\delta(t)$ as being zero everywhere except at $x = 0$ where it is unbounded. The above equation does not apply at the value $x = 0$. To maintain the right continuity in Eq. (4.18a), we use the convention:

$$u(0) = 1 = \int_{-\infty}^{0} \delta(t)\, dt.$$

If we replace $\pi_\Delta(t)$ in the above derivation with $g(t)\pi_\Delta(t)$, we obtain the "sifting" property of the delta function:

$$g(0) = \int_{-\infty}^{\infty} g(t)\delta(t)\, dt \quad \text{and} \qquad g(x_0) = \int_{-\infty}^{\infty} g(t)\delta(t - x_0)\, dt. \tag{4.22}$$

The delta function is viewed as sifting through x and picking out the value of g at the point where the delta functions is centered, that is, $g(x_0)$ for the expression on the right.

The pdf for the discrete random variable discussed in Example 4.1 is shown in Fig. 4.1(b). The pdf of a random variable of mixed type will also contain delta functions at the points where its cdf is not continuous. The pdf for the random variable discussed in Example 4.3 is shown in Fig. 4.3(b).

Example 4.9

Let X be the number of heads in three coin tosses as in Example 4.1. Find the pdf of X. Find $P[1 < X \le 2]$ and $P[2 \le X < 3]$ by integrating the pdf.

In Example 4.1 we found that the cdf of X is given by

$$F_X(x) = \frac{1}{8}u(x) + \frac{3}{8}u(x - 1) + \frac{3}{8}u(x - 2) + \frac{1}{8}u(x - 3).$$

It then follows from Eqs. (4.18) and (4.19) that

$$f_X(x) = \frac{1}{8}\delta(x) + \frac{3}{8}\delta(x - 1) + \frac{3}{8}\delta(x - 2) + \frac{1}{8}\delta(x - 3).$$

When delta functions appear in the limits of integration, we must indicate whether the delta functions are to be included in the integration. Thus in $P[1 < X \le 2] = P[X \text{ in } (1, 2]]$, the delta function located at 1 is excluded from the integral and the delta function at 2 is included:

$$P[1 < X \le 2] = \int_{1+}^{2+} f_X(x)\, dx = \frac{3}{8}.$$

Similarly, we have that

$$P[2 \le X < 3] = \int_{2-}^{3-} f_X(x)\, dx = \frac{3}{8}.$$

4.2.2 Conditional cdf's and pdf's

Conditional cdf's can be defined in a straightforward manner using the same approach we used for conditional pmf's. Suppose that event C is given and that $P[C] > 0$. The **conditional cdf of *X* given *C*** is defined by

$$F_X(x|C) = \frac{P[\{X \le x\} \cap C]}{P[C]} \quad \text{if } P[C] > 0. \tag{4.23}$$

It is easy to show that $F_X(x|C)$ satisfies all the properties of a cdf. (See Problem 4.29.) The **conditional pdf of *X* given *C*** is then defined by

$$f_X(x\,|\,C) = \frac{d}{dx}F_X(x|C). \tag{4.24}$$

Example 4.10

The lifetime X of a machine has a continuous cdf $F_X(x)$. Find the conditional cdf and pdf given the event $C = \{X > t\}$ (i.e., "machine is still working at time t").

The conditional cdf is

$$F_X(x|X > t) = P[X \le x\,|\,X > t] = \frac{P[\{X \le x\} \cap \{X > t\}]}{P[X > t]}.$$

The intersection of the two events in the numerator is equal to the empty set when $x < t$ and to $\{t < X \le x\}$ when $x \ge t$. Thus

$$F_X(x|X > t) = \begin{cases} 0 & x \le t \\ \dfrac{F_X(x) - F_X(t)}{1 - F_X(t)} & x > t. \end{cases}$$

The conditional pdf is found by differentiating with respect to x:

$$f_X(x\,|\,X > t) = \frac{f_X(x)}{1 - F_X(t)} \qquad x \ge t.$$

Now suppose that we have a partition of the sample space S into the union of disjoint events $B_1, B_2, \ldots, B_n$. Let $F_X(x|B_i)$ be the conditional cdf of X given event B_i. The theorem on total probability allows us to find the cdf of X in terms of the conditional cdf's:

$$F_X(x) = P[X \le x] = \sum_{i=1}^{n} P[X \le x\,|\,B_i]P[B_i] = \sum_{i=1}^{n} F_X(x\,|\,B_i)P[B_i]. \tag{4.25}$$

The pdf is obtained by differentiation:

$$f_X(x) = \frac{d}{dx}F_X(x) = \sum_{i=1}^{n} f_X(x\,|\,B_i)P[B_i]. \tag{4.26}$$

Example 4.11

A binary transmission system sends a "0" bit by transmitting a $-v$ voltage signal, and a "1" bit by transmitting a $+v$. The received signal is corrupted by Gaussian noise and given by:

$$Y = X + N$$

where X is the transmitted signal, and N is a noise voltage with pdf $f_N(x)$. Assume that $P[\text{"1"}] = p = 1 - P[\text{"0"}]$. Find the pdf of Y.

Let B_0 be the event "0" is transmitted and B_1 be the event "1" is transmitted, then B_0, B_1 form a partition, and

$$\begin{aligned} F_Y(x) &= F_Y(x \mid B_0)[B_0] + F_Y(x \mid B_1)[B_1] \\ &= P[Y \le x \mid X = -v](1 - p) + P[Y \le x \mid X = v]p. \end{aligned}$$

Since $Y = X + N$, the event $\{Y < x|X = v\}$ is equivalent to $\{v + N < x\}$ and $\{N < x - v\}$, and the event $\{Y < x|X = -v\}$ is equivalent to $\{N < x + v\}$. Therefore the conditional cdf's are:

$$F_Y(x|B_0) = P[N \le x + v] = F_N(x + v)$$

and

$$F_Y(x|B_1) = P[N \le x - v] = F_N(x - v).$$

The cdf is:

$$F_Y(x) = F_N(x + v)(1 - p) + F_N(x - v)p.$$

The pdf of N is then:

$$\begin{aligned} f_Y(x) &= \frac{d}{dx} F_Y(x) \\ &= \frac{d}{dx} F_N(x + v)(1 - p) + \frac{d}{dx} F_N(x - v)p \\ &= f_N(x + v)(1 - p) + f_N(x - v)p. \end{aligned}$$

The Gaussian random variable has pdf:

$$f_N(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2} \qquad -\infty < x < \infty.$$

The conditional pdfs are:

$$f_Y(x|B_0) = f_N(x + v) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x+v)^2/2\sigma^2}$$

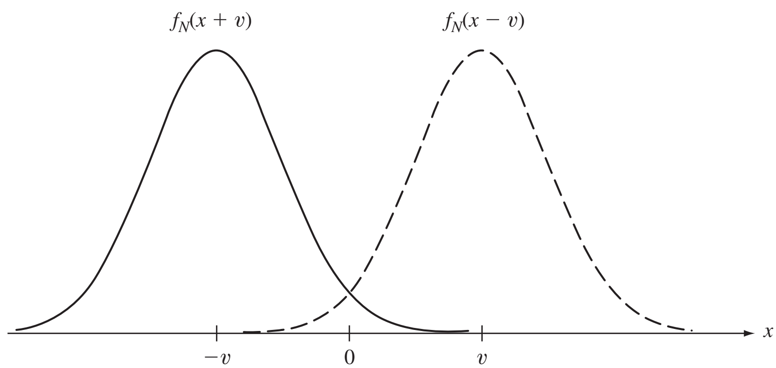


FIGURE 4.5
The conditional pdfs given the input signal

and

$$f_Y(x|B_1) = f_N(x - v) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-v)^2/2\sigma^2}.$$

The pdf of the received signal Y is then:

$$f_Y(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x+v)^2/2\sigma^2}(1 - p) + \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-v)^2/2\sigma^2} p.$$

Figure 4.5 shows the two conditional pdfs. We can see that the transmitted signal X shifts the center of mass of the Gaussian pdf.

4.3 THE EXPECTED VALUE OF X

We discussed the expected value for discrete random variables in Section 3.3, and found that the sample mean of independent observations of a random variable approaches $E[X]$. Suppose we perform a series of such experiments for continuous random variables. Since continuous random variables have $P[X = x] = 0$ for any specific value of x, we divide the real line into small intervals and count the number of times $N_k(n)$ the observations fall in the interval $\{x_k < X < x_k + \Delta\}$. As n becomes large, then the relative frequency $f_k(n) = N_k(n)/n$ will approach $f_X(x_k)\Delta$, the probability of the interval. We calculate the sample mean in terms of the relative frequencies and let $n \to \infty$:

$$\langle X \rangle_n = \sum_k x_k f_k(n) \to \sum_k x_k f_X(x_k)\Delta.$$

The expression on the right-hand side approaches an integral as we decrease Δ.

The **expected value** or **mean** of a random variable X is defined by

$$E[X] = \int_{-\infty}^{+\infty} t f_X(t)\, dt. \tag{4.27}$$

The expected value $E[X]$ is defined if the above integral converges absolutely, that is,

$$E[|X|] = \int_{-\infty}^{+\infty} |t| f_X(t)\, dt \qquad < \infty.$$

If we view $f_X(x)$ as the distribution of mass on the real line, then $E[X]$ represents the center of mass of this distribution.

We already discussed $E[X]$ for discrete random variables in detail, but it is worth noting that the definition in Eq. (4.27) is applicable if we express the pdf of a discrete random variable using delta functions:

$$\begin{aligned} E[X] &= \int_{-\infty}^{+\infty} t \sum_k p_X(x_k)\delta(t - x_k)\, dt \\ &= \sum_k p_X(x_k) \int_{-\infty}^{+\infty} t \sum_k \delta(t - x_k)\, dt \\ &= \sum_k p_X(x_k) x_k. \end{aligned}$$

Example 4.12 Mean of a Uniform Random Variable

The mean for a uniform random variable is given by

$$E[X] = (b - a)^{-1} \int_a^b t\, dt = \frac{a + b}{2},$$

which is exactly the midpoint of the interval $[a, b]$. The results shown in Fig. 3.6 were obtained by repeating experiments in which outcomes were random variables Y and X that had uniform cdf's in the intervals $[-1, 1]$ and $[3, 7]$, respectively. The respective expected values, 0 and 5, correspond to the values about which X and Y tend to vary.

The result in Example 4.12 could have been found immediately by noting that $E[X] = m$ when the pdf is symmetric about a point m. That is, if

$$f_X(m - x) = f_X(m + x) \qquad \text{for all } x,$$

then, assuming that the mean exists,

$$0 = \int_{-\infty}^{+\infty} (m - t) f_X(t)\, dt = m - \int_{-\infty}^{+\infty} t f_X(t)\, dt.$$

The first equality above follows from the symmetry of $f_X(t)$ about $t = m$ and the odd symmetry of $(m - t)$ about the same point. We then have that $E[X] = m$.

Example 4.13 Mean of a Gaussian Random Variable

The pdf of a Gaussian random variable is symmetric about the point $x = m$. Therefore $E[X] = m$.

The following expressions are useful when X is a nonnegative random variable:

$$E[X] = \int_0^{\infty} (1 - F_X(t))\, dt \qquad \text{if } X \text{ continuous and nonnegative} \tag{4.28}$$

and

$$E[X] = \sum_{k=0}^{\infty} P[X > k] \qquad \text{if } X \text{ nonnegative, integer-valued.} \tag{4.29}$$

The derivation of these formulas is discussed in Problem 4.47.

Example 4.14 Mean of Exponential Random Variable

The time X between customer arrivals at a service station has an exponential distribution. Find the mean interarrival time.

Substituting Eq. (4.17) into Eq. (4.27) we obtain

$$E[X] = \int_0^{\infty} t \lambda e^{-\lambda t}\, dt.$$

We evaluate the integral using integration by parts ($\int udv = uv - \int vdu$), with $u = t$ and $dv = \lambda e^{-\lambda t}\, dt$:

$$E[X] = -te^{-\lambda t}\Big|_0^\infty + \int_0^\infty e^{-\lambda t}\, dt$$

$$= \lim_{t\to\infty} te^{-\lambda t} - 0 + \left\{\frac{-e^{-\lambda t}}{\lambda}\right\}_0^\infty$$

$$= \lim_{t\to\infty}\frac{-e^{-\lambda t}}{\lambda} + \frac{1}{\lambda} = \frac{1}{\lambda},$$

where we have used the fact that $e^{-\lambda t}$ and $te^{-\lambda t}$ go to zero as t approaches infinity.

For this example, Eq. (4.28) is much easier to evaluate:

$$E[X] = \int_0^\infty e^{-\lambda t}\, dt = \frac{1}{\lambda}.$$

Recall that λ is the customer arrival rate in *customers per second*. The result that the mean interarrival time $E[X] = 1/\lambda$ *seconds per customer* then makes sense intuitively.

4.3.1 The Expected Value of $Y = g(X)$

Suppose that we are interested in finding the expected value of $Y = g(X)$. As in the case of discrete random variables (Eq. (3.16)), $E[Y]$ can be found directly in terms of the pdf of X:

$$E[Y] = \int_{-\infty}^{\infty} g(x)f_X(x)\, dx. \tag{4.30}$$

To see how Eq. (4.30) comes about, suppose that we divide the y-axis into intervals of length h, we index the intervals with the index k and we let y_k be the value in the center of the kth interval. The expected value of Y is approximated by the following sum:

$$E[Y] \simeq \sum_k y_k f_Y(y_k)h.$$

Suppose that $g(x)$ is strictly increasing, then the kth interval in the y-axis has a unique corresponding equivalent event of width h_k in the x-axis as shown in Fig. 4.6. Let x_k be the value in the kth interval such that $g(x_k) = y_k$, then since $f_Y(y_k)h = f_X(x_k)h_k$,

$$E[Y] \simeq \sum_k g(x_k)f_X(x_k)h_k.$$

By letting h approach zero, we obtain Eq. (4.30). This equation is valid even if $g(x)$ is not strictly increasing.

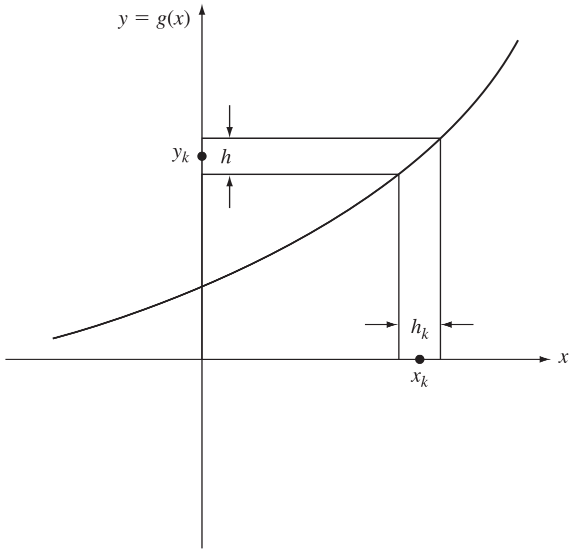


FIGURE 4.6
Two infinitesimal equivalent events.

Example 4.15 Expected Values of a Sinusoid with Random Phase

Let $Y = a\cos(\omega t + \Theta)$ where a, ω, and t are constants, and Θ is a uniform random variable in the interval $(0, 2\pi)$. The random variable Y results from sampling the amplitude of a sinusoid with random phase Θ. Find the expected value of Y and expected value of the power of Y, Y^2.

$$\begin{aligned} E[Y] &= E[a\cos(\omega t + \Theta)] \\ &= \int_0^{2\pi} a\cos(\omega t + \theta)\frac{d\theta}{2\pi} = -a\sin(\omega t + \theta)\Big|_0^{2\pi} \\ &= -a\sin(\omega t + 2\pi) + a\sin(\omega t) = 0. \end{aligned}$$

The average power is

$$\begin{aligned} E[Y^2] &= E[a^2\cos^2(\omega t + \Theta)] = E\left[\frac{a^2}{2} + \frac{a^2}{2}\cos(2\omega t + 2\Theta)\right] \\ &= \frac{a^2}{2} + \frac{a^2}{2}\int_0^{2\pi}\cos(2\omega t + \theta)\frac{d\theta}{2\pi} = \frac{a^2}{2}. \end{aligned}$$

Note that these answers are in agreement with the time averages of sinusoids: the time average ("dc" value) of the sinusoid is zero; the time-average power is $a^2/2$.

Example 4.16 Expected Values of the Indicator Function

Let $g(X) = I_C(X)$ be the indicator function for the event $\{X \text{ in } C\}$, where C is some interval or union of intervals in the real line:

$$g(X) = \begin{cases} 0 & X \text{ not in } C \\ 1 & X \text{ in } C, \end{cases}$$

then

$$E[Y] = \int_{-\infty}^{+\infty} g(X) f_X(x)\, dx = \int_C f_X(x)\, dx = P[X \text{ in } C].$$

Thus the expected value of the indicator of an event is equal to the probability of the event.

It is easy to show that Eqs. (3.17a)–(3.17e) hold for continuous random variables using Eq. (4.30). For example, let c be some constant, then

$$E[c] = \int_{-\infty}^{\infty} c f_X(x)\, dx = c \int_{-\infty}^{\infty} f_X(x)\, dx = c \tag{4.31}$$

and

$$E[cX] = \int_{-\infty}^{\infty} c x f_X(x)\, dx = c \int_{-\infty}^{\infty} x f_X(x)\, dx = cE[X]. \tag{4.32}$$

The expected value of a sum of functions of a random variable is equal to the sum of the expected values of the individual functions:

$$\begin{aligned} E[Y] &= E\left[\sum_{k=1}^{n} g_k(X)\right] \\ &= \int_{-\infty}^{\infty} \sum_{k=1}^{n} g_k(x) f_X(x)\, dx = \sum_{k=1}^{n} \int_{-\infty}^{\infty} g_k(x) f_X(x)\, dx \\ &= \sum_{k=1}^{n} E[g_k(X)]. \end{aligned} \tag{4.33}$$

Example 4.17

Let $Y = g(X) = a_0 + a_1 X + a_2 X^2 + \cdots + a_n X^n$, where a_k are constants, then

$$\begin{aligned} E[Y] &= E[a_0] + E[a_1 X] + \cdots + E[a_n X^n] \\ &= a_0 + a_1 E[X] + a_2 E[X^2] + \cdots + a_n E[X^n], \end{aligned}$$

where we have used Eq. (4.33), and Eqs. (4.31) and (4.32). A special case of this result is that

$$E[X + c] = E[X] + c,$$

that is, *we can shift the mean of a random variable by adding a constant to it.*

4.3.2 Variance of X

The **variance of the random variable** X is defined by

$$\text{VAR}[X] = E[(X - E[X])^2] = E[X^2] - E[X]^2 \tag{4.34}$$

The **standard deviation of the random variable** X is defined by

$$\text{STD}[X] = \text{VAR}[X]^{1/2}. \tag{4.35}$$

Example 4.18 Variance of Uniform Random Variable

Find the variance of the random variable X that is uniformly distributed in the interval $[a, b]$.

Since the mean of X is $(a + b)/2$,

$$\text{VAR}[X] = \frac{1}{b-a}\int_a^b \left(x - \frac{a+b}{2}\right)^2 dx.$$

Let $y = (x - (a + b)/2)$,

$$\text{VAR}[X] = \frac{1}{b-a}\int_{-(b-a)/2}^{(b-a)/2} y^2\,dy = \frac{(b-a)^2}{12}.$$

The random variables in Fig. 3.6 were uniformly distributed in the interval $[-1, 1]$ and [3, 7], respectively. Their variances are then 1/3 and 4/3. The corresponding standard deviations are 0.577 and 1.155.

Example 4.19 Variance of Gaussian Random Variable

Find the variance of a Gaussian random variable.

First multiply the integral of the pdf of X by $\sqrt{2\pi}\,\sigma$ to obtain

$$\int_{-\infty}^{\infty} e^{-(x-m)^2/2\sigma^2}\,dx = \sqrt{2\pi}\,\sigma.$$

Differentiate both sides with respect to σ:

$$\int_{-\infty}^{\infty}\left(\frac{(x-m)^2}{\sigma^3}\right)e^{-(x-m)^2/2\sigma^2}\,dx = \sqrt{2\pi}.$$

By rearranging the above equation, we obtain

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

This result can also be obtained by direct integration. (See Problem 4.46.) Figure 4.7 shows the Gaussian pdf for several values of σ; it is evident that the "width" of the pdf increases with σ.

The following properties were derived in Section 3.3:

$$\text{VAR}[c] = 0 \tag{4.36}$$

$$\text{VAR}[X + c] = \text{VAR}[X] \tag{4.37}$$

$$\text{VAR}[cX] = c^2\,\text{VAR}[X], \tag{4.38}$$

where c is a constant.

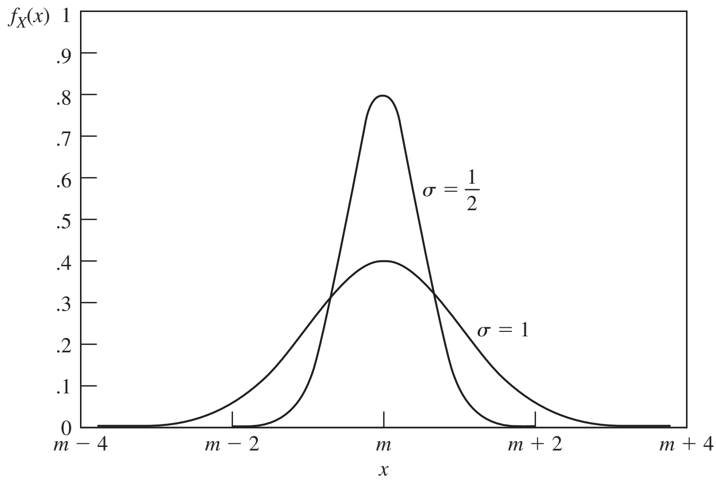


FIGURE 4.7
Probability density function of Gaussian random variable.

The mean and variance are the two most important parameters used in summarizing the pdf of a random variable. Other parameters are occasionally used. For example, the skewness defined by $E[(X - E[X])^3]/\text{STD}[X]^3$ measures the degree of asymmetry about the mean. It is easy to show that if a pdf is symmetric about its mean, then its skewness is zero. The point to note with these parameters of the pdf is that each involves the expected value of a higher power of X. Indeed we show in a later section that, under certain conditions, a pdf is completely specified if the expected values of all the powers of X are known. These expected values are called the moments of X.

The ***n*th moment of the random variable *X*** is defined by

$$E[X^n] = \int_{-\infty}^{\infty} x^n f_X(x)\, dx. \tag{4.39}$$

The mean and variance can be seen to be defined in terms of the first two moments, $E[X]$ and $E[X^2]$.

---

*Example 4.20 Analog-to-Digital Conversion: A Detailed Example

A quantizer is used to convert an analog signal (e.g., speech or audio) into digital form. A quantizer maps a random voltage X into the nearest point $q(X)$ from a set of 2^R representation values as shown in Fig. 4.8(a). The value X is then approximated by $q(X)$, which is identified by an R-bit binary number. In this manner, an "analog" voltage X that can assume a continuum of values is converted into an R-bit number.

The quantizer introduces an error $Z = X - q(X)$ as shown in Fig. 4.8(b). Note that Z is a function of X and that it ranges in value between $-d/2$ and $d/2$, where d is the quantizer step size. Suppose that X has a uniform distribution in the interval $[-x_{\max}, x_{\max}]$, that the quantizer has 2^R levels, and that $2x_{\max} = 2^R d$. It is easy to show that Z is uniformly distributed in the interval $[-d/2, d/2]$ (see Problem 4.93).

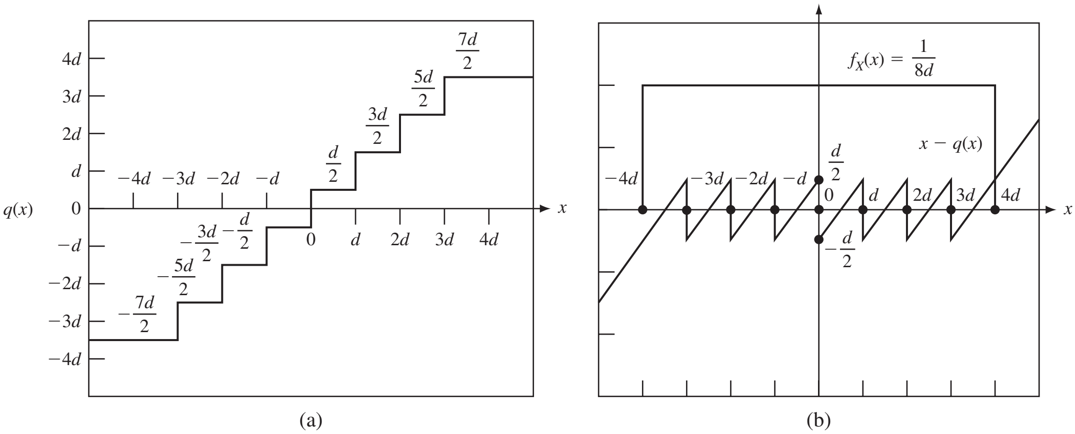


FIGURE 4.8
(a) A uniform quantizer maps the input x into the closest point from the set $\{\pm d/2, \pm 3d/2, \pm 5d/2, \pm 7d/2\}$. (b) The uniform quantizer error for the input x is $x - q(x)$.

Therefore from Example 4.12,

$$E[Z] = \frac{d/2 - d/2}{2} = 0.$$

The error Z thus has mean zero.

By Example 4.18,

$$\text{VAR}[Z] = \frac{(d/2 - (-d/2))^2}{12} = \frac{d^2}{12}.$$

This result is approximately correct for any pdf that is approximately flat over each quantizer interval. This is the case when 2^R is large.

The approximation $q(x)$ can be viewed as a "noisy" version of X since

$$Q(X) = X - Z,$$

where Z is the quantization error Z. The measure of goodness of a quantizer is specified by the SNR ratio, which is defined as the ratio of the variance of the "signal" X to the variance of the distortion or "noise" Z:

$$\text{SNR} = \frac{\text{VAR}[X]}{\text{VAR}[Z]} = \frac{\text{VAR}[X]}{d^2/12}$$
$$= \frac{\text{VAR}[X]}{x_{\max}^2/3} 2^{2R},$$

where we have used the fact that $d = 2x_{\max}/2^R$. When X is nonuniform, the value $x_{\max}$ is selected so that $P[|X| > x_{\max}]$ is small. A typical choice is $x_{\max} = 4\ \text{STD}[X]$. The SNR is then

$$\text{SNR} = \frac{3}{16} 2^{2R}.$$

This important formula is often quoted in decibels:

$$\text{SNR dB} = 10 \log_{10} \text{SNR} = 6R - 7.3 \text{ dB}.$$

The SNR increases by a factor of 4 (6 dB) with each additional bit used to represent X. This makes sense since each additional bit doubles the number of quantizer levels, which in turn reduces the step size by a factor of 2. The variance of the error should then be reduced by the square of this, namely $2^2 = 4$.

4.4 IMPORTANT CONTINUOUS RANDOM VARIABLES

We are always limited to measurements of finite precision, so in effect, every random variable found in practice is a discrete random variable. Nevertheless, there are several compelling reasons for using continuous random variable models. First, in general, continuous random variables are easier to handle analytically. Second, the limiting form of many discrete random variables yields continuous random variables. Finally, there are a number of "families" of continuous random variables that can be used to model a wide variety of situations by adjusting a few parameters. In this section we continue our introduction of important random variables. Table 4.1 lists some of the more important continuous random variables.

4.4.1 The Uniform Random Variable

The uniform random variable arises in situations where all values in an interval of the real line are equally likely to occur. The uniform random variable U in the interval $[a, b]$ has pdf:

$$f_U(x) = \begin{cases} \dfrac{1}{b-a} & a \le x \le b \\ 0 & x < a \quad \text{and} \quad x > b \end{cases} \tag{4.40}$$

and cdf

$$F_U(x) = \begin{cases} 0 & x < a \\ \dfrac{x-a}{b-a} & a \le x \le b \\ 1 & x > b. \end{cases} \tag{4.41}$$

See Figure 4.2. The mean and variance of U are given by:

$$E[U] = \frac{a+b}{2} \quad \text{and} \quad \text{VAR}[X] = \frac{(b-a)^2}{2}. \tag{4.42}$$

The uniform random variable appears in many situations that involve equally likely continuous random variables. Obviously U can only be defined over intervals that are finite in length. We will see in Section 4.9 that the uniform random variable plays a crucial role in generating random variables in computer simulation models.

4.4.2 The Exponential Random Variable

The exponential random variable arises in the modeling of the time between occurrence of events (e.g., the time between customer demands for call connections), and in the modeling of the lifetime of devices and systems. The **exponential random variable** X with parameter λ has pdf

TABLE 4.1 Continuous random variables.

Uniform Random Variable

$S_X = [a, b]$

$f_X(x) = \dfrac{1}{b - a} \qquad a \le x \le b$

$E[X] = \dfrac{a + b}{2} \qquad \text{VAR}[X] = \dfrac{(b - a)^2}{12} \qquad \Phi_X(\omega) = \dfrac{e^{j\omega b} - e^{j\omega a}}{j\omega(b - a)}$

Exponential Random Variable

$S_X = [0, \infty)$

$f_X(x) = \lambda e^{-\lambda x} \qquad x \ge 0 \quad \text{and} \quad \lambda > 0$

$E[X] = \dfrac{1}{\lambda} \qquad \text{VAR}[X] = \dfrac{1}{\lambda^2} \qquad \Phi_X(\omega) = \dfrac{\lambda}{\lambda - j\omega}$

Remarks: The exponential random variable is the only continuous random variable with the memoryless property.

Gaussian (Normal) Random Variable

$S_X = (-\infty, +\infty)$

$f_X(x) = \dfrac{e^{-(x-m)^2/2\sigma^2}}{\sqrt{2\pi}\sigma} \qquad -\infty < x < +\infty \quad \text{and} \quad \sigma > 0$

$E[X] = m \qquad \text{VAR}[X] = \sigma^2 \qquad \Phi_X(\omega) = e^{jm\omega - \sigma^2\omega^2/2}$

Remarks: Under a wide range of conditions X can be used to approximate the sum of a large number of independent random variables.

Gamma Random Variable

$S_X = (0, +\infty)$

$f_X(x) = \dfrac{\lambda(\lambda x)^{\alpha-1}e^{-\lambda x}}{\Gamma(\alpha)} \qquad x > 0 \quad \text{and} \quad \alpha > 0, \lambda > 0$

where $\Gamma(z)$ is the gamma function (Eq. 4.56).

$E[X] = \alpha/\lambda \qquad \text{VAR}[X] = \alpha/\lambda^2 \qquad \Phi_X(\omega) = \dfrac{1}{(1 - j\omega/\lambda)^\alpha}$

Special Cases of Gamma Random Variable

m–1 Erlang Random Variable: $\alpha = m$, a positive integer

$f_X(x) = \dfrac{\lambda e^{-\lambda x}(\lambda x)^{m-2}}{(m - 1)!} \qquad x > 0 \qquad \Phi_X(\omega) = \left(\dfrac{1}{1 - j\omega/\lambda}\right)^m$

Remarks: An m–1 Erlang random variable is obtained by adding m independent exponentially distributed random variables with parameter λ.

Chi-Square Random Variable with k degrees of freedom: $\alpha = k/2$, k a positive integer, and $\lambda = 1/2$

$f_X(x) = \dfrac{x^{(k-2)/2}e^{-x/2}}{2^{k/2}\Gamma(k/2)} \qquad x > 0 \qquad \Phi_X(\omega) = \left(\dfrac{1}{1 - 2j\omega}\right)^{k/2}$

Remarks: The sum of k mutually independent, squared zero-mean, unit-variance Gaussian random variables is a chi-square random variable with k degrees of freedom.

Laplacian Random Variable

$S_X = (-\infty, \infty)$

$f_X(x) = \frac{\alpha}{2} e^{-\alpha|x|} \qquad -\infty < x < +\infty \quad \text{and} \quad \alpha > 0$

$E[X] = 0 \qquad \text{VAR}[X] = 2/\alpha^2 \qquad \Phi_X(\omega) = \frac{\alpha^2}{\omega^2 + \alpha^2}$

Rayleigh Random Variable

$S_X = [0, \infty)$

$f_X(x) = \frac{x}{\alpha^2} e^{-x^2/2\alpha^2} \qquad x \geq 0 \quad \text{and} \quad \alpha > 0$

$E[X] = \alpha\sqrt{\pi/2} \qquad \text{VAR}[X] = (2 - \pi/2)\alpha^2$

Cauchy Random Variable

$S_X = (-\infty, +\infty)$

$f_X(x) = \frac{\alpha/\pi}{x^2 + \alpha^2} \qquad -\infty < x < +\infty \quad \text{and} \quad \alpha > 0$

Mean and variance do not exist. $\qquad \Phi_X(\omega) = e^{-\alpha|\omega|}$

Pareto Random Variable

$S_X = [x_m, \infty) x_m > 0.$

$$f_X(x) = \begin{cases} 0 & x < x_m \\ \alpha \dfrac{x_m^\alpha}{x^{\alpha+1}} & x \geq x_m \end{cases}$$

$E[X] = \frac{\alpha x_m}{\alpha - 1} \quad \text{for } \alpha > 1 \qquad \text{VAR}[X] = \frac{\alpha x_m^2}{(\alpha - 2)(\alpha - 1)^2} \quad \text{for } \alpha > 2$

Remarks: The Pareto random variable is the most prominent example of random variables with "long tails," and can be viewed as a continuous version of the Zipf discrete random variable.

Beta Random Variable

$$f_X(x) = \begin{cases} \dfrac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\,\Gamma(\beta)} x^{\alpha-1}(1 - x)^{\beta-1} & 0 < x < 1 \quad \text{and} \quad \alpha > 0, \beta > 0 \\ 0 & \text{otherwise} \end{cases}$$

$E[X] = \frac{\alpha}{\alpha + \beta} \qquad \text{VAR}[X] = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$

Remarks: The beta random variable is useful for modeling a variety of pdf shapes for random variables that range over finite intervals.

$$f_X(x) = \begin{cases} 0 & x < 0 \\ \lambda e^{-\lambda x} & x \geq 0 \end{cases} \tag{4.43}$$

and cdf

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \geq 0. \end{cases} \tag{4.44}$$

The cdf and pdf of X are shown in Fig. 4.9.

The parameter λ is the rate at which events occur, so in Eq. (4.44) the probability of an event occurring by time x increases at the rate λ increases. Recall from Example 3.31 that the interarrival times between events in a Poisson process (Fig. 3.10) is an exponential random variable.

The mean and variance of X are given by:

$$E[U] = \frac{1}{\lambda} \quad \text{and} \quad VAR[X] = \frac{1}{\lambda^2}. \tag{4.45}$$

In event interarrival situations, λ is in units of events/second and $1/\lambda$ is in units of seconds per event interarrival.

The exponential random variable satisfies the **memoryless property**:

$$P[X > t + h | X > t] = P[X > h]. \tag{4.46}$$

The expression on the left side is the probability of having to wait at least h additional seconds given that one has already been waiting t seconds. The expression on the right side is the probability of waiting at least h seconds when one first begins to wait. Thus the probability of waiting at least an additional h seconds is the same regardless of how long one has already been waiting! We see later in the book that the memoryless property of the exponential random variable makes it the cornerstone for the theory of

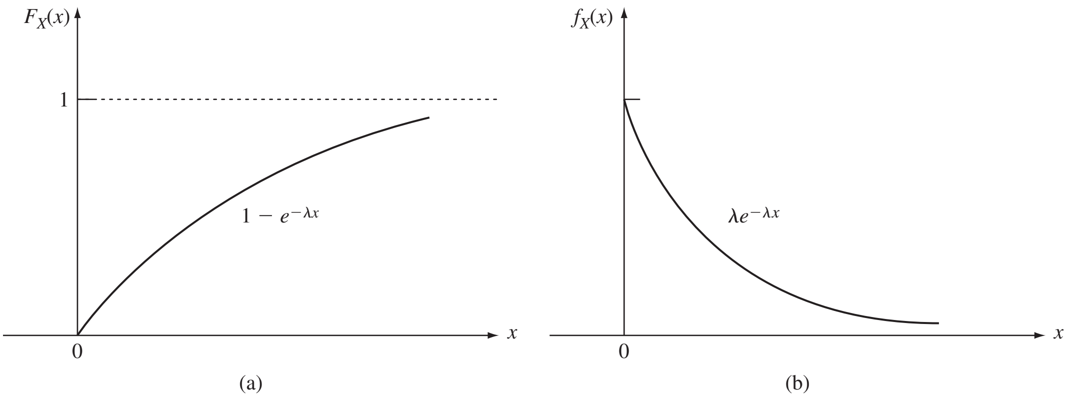


FIGURE 4.9
An example of a continuous random variable—the exponential random variable. Part (a) is the cdf and part (b) is the pdf.

Markov chains, which is used extensively in evaluating the performance of computer systems and communications networks.

We now prove the memoryless property:

$$\begin{aligned} P[X > t + h \mid X > t] &= \frac{P[\{X > t + h\} \cap \{X > t\}]}{P[X > t]} \qquad \text{for } h > 0 \\ &= \frac{P[X > t + h]}{P[X > t]} = \frac{e^{-\lambda(t+h)}}{e^{-\lambda t}} \\ &= e^{-\lambda h} = P[X > h]. \end{aligned}$$

It can be shown that the exponential random variable is the only continuous random variable that satisfies the memoryless property.

Examples 2.13, 2.28, and 2.30 dealt with the exponential random variable.

4.4.3 The Gaussian (Normal) Random Variable

There are many situations in manmade and in natural phenomena where one deals with a random variable X that consists of the sum of a large number of "small" random variables. The exact description of the pdf of X in terms of the component random variables can become quite complex and unwieldy. However, one finds that under very general conditions, as the number of components becomes large, the cdf of X approaches that of the **Gaussian (normal) random variable.**[1] This random variable appears so often in problems involving randomness that it has come to be known as the "normal" random variable.

The pdf for the Gaussian random variable X is given by

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-m)^2/2\sigma^2} \qquad -\infty < x < \infty, \tag{4.47}$$

where m and $\sigma > 0$ are real numbers, which we showed in Examples 4.13 and 4.19 to be the mean and standard deviation of X. Figure 4.7 shows that the Gaussian pdf is a "bell-shaped" curve centered and symmetric about m and whose "width" increases with σ.

The cdf of the Gaussian random variable is given by

$$P[X \leq x] = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} e^{-(x'-m)^2/2\sigma^2}\, dx'. \tag{4.48}$$

The change of variable $t = (x' - m)/\sigma$ results in

$$\begin{aligned} F_X(x) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(x-m)/\sigma} e^{-t^2/2}\, dt \\ &= \Phi\left(\frac{x - m}{\sigma}\right) \end{aligned} \tag{4.49}$$

where $\Phi(x)$ is the cdf of a Gaussian random variable with $m = 0$ and $\sigma = 1$:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt. \tag{4.50}$$

[1]This result, called the central limit theorem, will be discussed in Chapter 7.

Therefore any probability involving an arbitrary Gaussian random variable can be expressed in terms of $\Phi(x)$.

Example 4.21

Show that the Gaussian pdf integrates to one. Consider the square of the integral of the pdf:

$$\left[\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-x^2/2}\,dx\right]^2 = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-x^2/2}\,dx \int_{-\infty}^{\infty} e^{-y^2/2}\,dy$$
$$= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)/2}\,dx\,dy.$$

Let $x = r\cos\theta$ and $y = r\sin\theta$ and carry out the change from Cartesian to polar coordinates, then we obtain:

$$\frac{1}{2\pi}\int_0^{\infty}\int_0^{2\pi} e^{-r^2/2} r\,dr\,d\theta = \int_0^{\infty} re^{-r^2/2}\,dr$$
$$= [-e^{-r^2/2}]_0^{\infty}$$
$$= 1.$$

In electrical engineering it is customary to work with the Q-function, which is defined by

$$Q(x) = 1 - \Phi(x) \tag{4.51}$$

$$= \frac{1}{\sqrt{2\pi}}\int_x^{\infty} e^{-t^2/2}\,dt. \tag{4.52}$$

$Q(x)$ is simply the probability of the "tail" of the pdf. The symmetry of the pdf implies that

$$Q(0) = 1/2 \qquad \text{and} \qquad Q(-x) = 1 - Q(x). \tag{4.53}$$

The integral in Eq. (4.50) does not have a closed-form expression. Traditionally the integrals have been evaluated by looking up tables that list $Q(x)$ or by using approximations that require numerical evaluation [Ross]. The following expression has been found to give good accuracy for $Q(x)$ over the entire range $0 < x < \infty$:

$$Q(x) \simeq \left[\frac{1}{(1-a)x + a\sqrt{x^2 + b}}\right]\frac{1}{\sqrt{2\pi}}e^{-x^2/2}, \tag{4.54}$$

where $a = 1/\pi$ and $b = 2\pi$ [Gallager]. Table 4.2 shows $Q(x)$ and the value given by the above approximation. In some problems, we are interested in finding the value of x for which $Q(x) = 10^{-k}$. Table 4.3 gives these values for $k = 1, \ldots, 10$.

The Gaussian random variable plays a very important role in communication systems, where transmission signals are corrupted by noise voltages resulting from the thermal motion of electrons. It can be shown from physical principles that these voltages will have a Gaussian pdf.

TABLE 4.2 Comparison of $Q(x)$ and approximation given by Eq. (4.54).

x	$Q(x)$	Approximation	x	$Q(x)$	Approximation
0	5.00E-01	5.00E-01	2.7	3.47E-03	3.46E-03
0.1	4.60E-01	4.58E-01	2.8	2.56E-03	2.55E-03
0.2	4.21E-01	4.17E-01	2.9	1.87E-03	1.86E-03
0.3	3.82E-01	3.78E-01	3.0	1.35E-03	1.35E-03
0.4	3.45E-01	3.41E-01	3.1	9.68E-04	9.66E-04
0.5	3.09E-01	3.05E-01	3.2	6.87E-04	6.86E-04
0.6	2.74E-01	2.71E-01	3.3	4.83E-04	4.83E-04
0.7	2.42E-01	2.39E-01	3.4	3.37E-04	3.36E-04
0.8	2.12E-01	2.09E-01	3.5	2.33E-04	2.32E-04
0.9	1.84E-01	1.82E-01	3.6	1.59E-04	1.59E-04
1.0	1.59E-01	1.57E-01	3.7	1.08E-04	1.08E-04
1.1	1.36E-01	1.34E-01	3.8	7.24E-05	7.23E-05
1.2	1.15E-01	1.14E-01	3.9	4.81E-05	4.81E-05
1.3	9.68E-02	9.60E-02	4.0	3.17E-05	3.16E-05
1.4	8.08E-02	8.01E-02	4.5	3.40E-06	3.40E-06
1.5	6.68E-02	6.63E-02	5.0	2.87E-07	2.87E-07
1.6	5.48E-02	5.44E-02	5.5	1.90E-08	1.90E-08
1.7	4.46E-02	4.43E-02	6.0	9.87E-10	9.86E-10
1.8	3.59E-02	3.57E-02	6.5	4.02E-11	4.02E-11
1.9	2.87E-02	2.86E-02	7.0	1.28E-12	1.28E-12
2.0	2.28E-02	2.26E-02	7.5	3.19E-14	3.19E-14
2.1	1.79E-02	1.78E-02	8.0	6.22E-16	6.22E-16
2.2	1.39E-02	1.39E-02	8.5	9.48E-18	9.48E-18
2.3	1.07E-02	1.07E-02	9.0	1.13E-19	1.13E-19
2.4	8.20E-03	8.17E-03	9.5	1.05E-21	1.05E-21
2.5	6.21E-03	6.19E-03	10.0	7.62E-24	7.62E-24
2.6	4.66E-03	4.65E-03			

Example 4.22

A communication system accepts a positive voltage V as input and outputs a voltage $Y = \alpha V + N$, where $\alpha = 10^{-2}$ and N is a Gaussian random variable with parameters $m = 0$ and $\sigma = 2$. Find the value of V that gives $P[Y < 0] = 10^{-6}$.

The probability $P[Y < 0]$ is written in terms of N as follows:

$$\begin{aligned} P[Y < 0] &= P[\alpha V + N < 0] \\ &= P[N < -\alpha V] = \Phi\left(\frac{-\alpha V}{\sigma}\right) = Q\left(\frac{\alpha V}{\sigma}\right) = 10^{-6}. \end{aligned}$$

From Table 4.3 we see that the argument of the Q-function should be $\alpha V/\sigma = 4.753$. Thus $V = (4.753)\sigma/\alpha = 950.6$.

TABLE 4.3 $Q(x) = 10^{-k}$

k	$x = Q^{-1}(10^{-k})$
1	1.2815
2	2.3263
3	3.0902
4	3.7190
5	4.2649
6	4.7535
7	5.1993
8	5.6120
9	5.9978
10	6.3613

4.4.4 The Gamma Random Variable

The gamma random variable is a versatile random variable that appears in many applications. For example, it is used to model the time required to service customers in queueing systems, the lifetime of devices and systems in reliability studies, and the defect clustering behavior in VLSI chips.

The pdf of the **gamma random variable** has two parameters, $\alpha > 0$ and $\lambda > 0$, and is given by

$$f_X(x) = \frac{\lambda(\lambda x)^{\alpha-1}e^{-\lambda x}}{\Gamma(\alpha)} \qquad 0 < x < \infty, \tag{4.55}$$

where $\Gamma(z)$ is the gamma function, which is defined by the integral

$$\Gamma(z) = \int_0^\infty x^{z-1}e^{-x}\,dx \qquad z > 0. \tag{4.56}$$

The gamma function has the following properties:

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi},$$

$$\Gamma(z+1) = z\Gamma(z) \qquad \text{for } z > 0, \text{ and}$$

$$\Gamma(m+1) = m! \qquad \text{for } m \text{ a nonnegative integer.}$$

The versatility of the gamma random variable is due to the richness of the gamma function $\Gamma(z)$. The pdf of the gamma random variable can assume a variety of shapes as shown in Fig. 4.10. By varying the parameters α and λ it is possible to fit the gamma pdf to many types of experimental data. In addition, many random variables are special cases of the gamma random variable. The exponential random variable is obtained by letting $\alpha = 1$. By letting $\lambda = 1/2$ and $\alpha = k/2$, where k is a positive integer, we obtain the **chi-square random variable**, which appears in certain statistical problems. The ***m*-Erlang random variable** is obtained when $\alpha = m$, a positive integer. The m-Erlang random variable is used in the system reliability models and in queueing systems models. Both of these random variables are discussed in later examples.

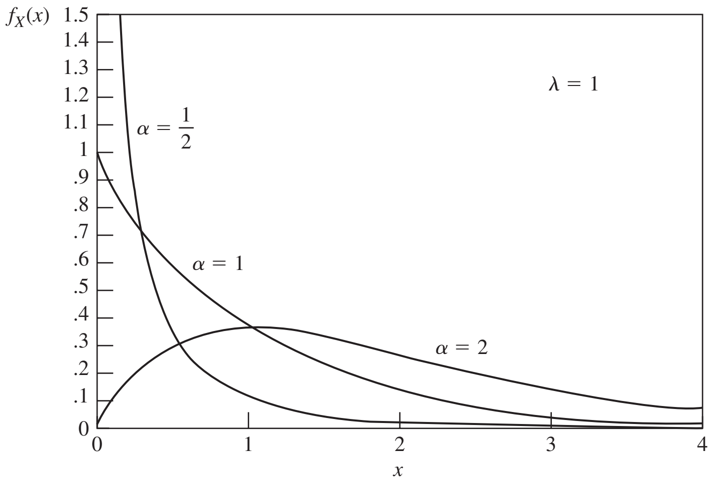


FIGURE 4.10
Probability density function of gamma random variable.

Example 4.23

Show that the pdf of a gamma random variable integrates to one.

The integral of the pdf is

$$\int_0^{\infty} f_X(x)\,dx = \int_0^{\infty} \frac{\lambda(\lambda x)^{\alpha-1}e^{-\lambda x}}{\Gamma(\alpha)}\,dx$$

$$= \frac{\lambda^{\alpha}}{\Gamma(\alpha)}\int_0^{\infty} x^{\alpha-1}e^{-\lambda x}\,dx.$$

Let $y = \lambda x$, then $dx = dy/\lambda$ and the integral becomes

$$\frac{\lambda^{\alpha}}{\Gamma(\alpha)\lambda^{\alpha}}\int_0^{\infty} y^{\alpha-1}e^{-y}\,dy = 1,$$

where we used the fact that the integral equals $\Gamma(\alpha)$.

In general, the cdf of the gamma random variable does not have a closed-form expression. We will show that the special case of the m-Erlang random variable does have a closed-form expression for the cdf by using its close interrelation with the exponential and Poisson random variables. The cdf can also be obtained by integration of the pdf (see Problem 4.74).

Consider once again the limiting procedure that was used to derive the Poisson random variable. Suppose that we observe the time S_m that elapses until the occurrence of the mth event. The times $X_1, X_2, \ldots, X_m$ between events are exponential random variables, so we must have

$$S_m = X_1 + X_2 + \cdots + X_m.$$

We will show that S_m is an m-Erlang random variable. To find the cdf of S_m, let $N(t)$ be the Poisson random variable for the number of events in t seconds. Note that the mth event occurs before time t—that is, $S_m \le t$—if and only if m or more events occur in t seconds, namely $N(t) \ge m$. The reasoning goes as follows. If the mth event has occurred before time t, then it follows that m or more events will occur in time t. On the other hand, if m or more events occur in time t, then it follows that the mth event occurred by time t. Thus

$$F_{S_m}(t) = P[S_m \le t] = P[N(t) \ge m] \tag{4.57}$$

$$= 1 - \sum_{k=0}^{m-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \tag{4.58}$$

where we have used the result of Example 3.31. If we take the derivative of the above cdf, we finally obtain the pdf of the m-Erlang random variable. Thus we have shown that S_m is an m-Erlang random variable.

Example 4.24

A factory has two spares of a critical system component that has an average lifetime of $1/\lambda = 1$ month. Find the probability that the three components (the operating one and the two spares) will last more than 6 months. Assume the component lifetimes are exponential random variables.

The remaining lifetime of the component in service is an exponential random variable with rate λ by the memoryless property. Thus, the total lifetime X of the three components is the sum of three exponential random variables with parameter $\lambda = 1$. Thus X has a 3-Erlang distribution with $\lambda = 1$. From Eq. (4.58) the probability that X is greater than 6 is

$$P[X > 6] = 1 - P[X \le 6]$$

$$= \sum_{k=0}^{2} \frac{6^k}{k!} e^{-6} = .06197.$$

4.4.5 The Beta Random Variable

The beta random variable X assumes values over a closed interval and has pdf:

$$f_X(x) = cx^{a-1}(1 - x)^{b-1} \qquad \text{for } 0 < x < 1 \tag{4.59}$$

where the normalization constant is the reciprocal of the beta function

$$\frac{1}{c} = B(a, b) = \int_0^1 x^{a-1}(1 - x)^{b-1}\,dx$$

and where the beta function is related to the gamma function by the following expression:

$$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a + b)}.$$

When $a = b = 1$, we have the uniform random variable. Other choices of a and b give pdfs over finite intervals that can differ markedly from the uniform. See Problem 4.75. If

$a = b > 1$, then the pdf is symmetric about $x = 1/2$ and is concentrated about $x = 1/2$ as well. When $a = b < 1$, then the pdf is symmetric but the density is concentrated at the edges of the interval. When $a < b$ (or $a > b$) the pdf is skewed to the right (or left).

The mean and variance are given by:

$$E[X] = \frac{a}{a + b} \quad \text{and} \quad \text{VAR}[X] = \frac{ab}{(a + b)^2(a + b + 1)}. \tag{4.60}$$

The versatility of the pdf of the beta random variable makes it useful to model a variety of behaviors for random variables that range over finite intervals. For example, in a Bernoulli trial experiment, the probability of success p could itself be a random variable. The beta pdf is frequently used to model p.

4.4.6 The Cauchy Random Variable

The Cauchy random variable X assumes values over the entire real line and has pdf:

$$f_X(x) = \frac{1/\pi}{1 + x^2}. \tag{4.61}$$

It is easy to verify that this pdf integrates to 1. However, X does not have any moments since the associated integrals do not converge. The Cauchy random variable arises as the tangent of a uniform random variable in the unit interval.

4.4.7 The Pareto Random Variable

The Pareto random variable arises in the study of the distribution of wealth where it has been found to model the tendency for a small portion of the population to own a large portion of the wealth. Recently the Pareto distribution has been found to capture the behavior of many quantities of interest in the study of Internet behavior, e.g., sizes of files, packet delays, audio and video title preferences, session times in peer-to-peer networks, etc. The Pareto random variable can be viewed as a continuous version of the Zipf discrete random variable.

The Pareto random variable X takes on values in the range $x > x_m$, where x_m is a positive real number. X has complementary cdf with shape parameter $\alpha > 0$ given by:

$$P[X > x] = \begin{cases} 1 & x < x_m \\ \dfrac{x_m^\alpha}{x^\alpha} & x \geq x_m. \end{cases} \tag{4.62}$$

The tail of X decays algebraically with x which is rather slower in comparison to the exponential and Gaussian random variables. The Pareto random variable is the most prominent example of random variables with "long tails."

The cdf and pdf of X are:

$$F_X(x) = \begin{cases} 0 & x < x_m \\ 1 - \dfrac{x_m^\alpha}{x^\alpha} & x \geq x_m. \end{cases} \tag{4.63}$$

Because of its long tail, the cdf of X approaches 1 rather slowly as x increases.

$$f_X(x) = \begin{cases} 0 & x < x_m \\ \alpha \dfrac{x_m^\alpha}{x^{\alpha+1}} & x \geq x_m. \end{cases} \tag{4.64}$$

Example 4.25 Mean and Variance of Pareto Random Variable

Find the mean and variance of the Pareto random variable.

$$E[X] = \int_{x_m}^{\infty} t\alpha \frac{x_m^\alpha}{t^{\alpha+1}} dt = \int_{x_m}^{\infty} \alpha \frac{x_m^\alpha}{t^\alpha} dt = \frac{\alpha}{\alpha - 1} \frac{x_m^\alpha}{x_m^{\alpha-1}} = \frac{\alpha x_m}{\alpha - 1} \qquad \text{for } \alpha > 1 \tag{4.65}$$

where the integral is defined for $\alpha > 1$, and

$$E[X^2] = \int_{x_m}^{\infty} t^2\alpha \frac{x_m^\alpha}{t^{\alpha+1}} dt = \int_{x_m}^{\infty} \alpha \frac{x_m^\alpha}{t^{\alpha-1}} dt = \frac{\alpha}{\alpha - 2} \frac{x_m^\alpha}{x_m^{\alpha-2}} = \frac{\alpha x_m^2}{\alpha - 2} \qquad \text{for } \alpha > 2$$

where the second moment is defined for $\alpha > 2$.

The variance of X is then:

$$\text{VAR}[X] = \frac{\alpha x_m^2}{\alpha - 2} - \left(\frac{\alpha x_m^2}{\alpha - 1}\right)^2 = \frac{\alpha x_m^2}{(\alpha - 2)(\alpha - 1)^2} \qquad \text{for } \alpha > 2. \tag{4.66}$$

4.5 FUNCTIONS OF A RANDOM VARIABLE

Let X be a random variable and let $g(x)$ be a real-valued function defined on the real line. Define $Y = g(X)$, that is, Y is determined by evaluating the function $g(x)$ at the value assumed by the random variable X. Then Y is also a random variable. The probabilities with which Y takes on various values depend on the function $g(x)$ as well as the cumulative distribution function of X. In this section we consider the problem of finding the cdf and pdf of Y.

Example 4.26

Let the function $h(x) = (x)^+$ be defined as follows:

$$(x)^+ = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } x \geq 0. \end{cases}$$

For example, let X be the number of active speakers in a group of N speakers, and let Y be the number of active speakers in excess of M, then $Y = (X - M)^+$. In another example, let X be a voltage input to a halfwave rectifier, then $Y = (X)^+$ is the output.

Example 4.27

Let the function $q(x)$ be defined as shown in Fig. 4.8(a), where the set of points on the real line are mapped into the nearest representation point from the set $S_Y = \{-3.5d, -2.5d, -1.5d, -0.5d, 0.5d, 1.5d, 2.5d, 3.5d\}$. Thus, for example, all the points in the interval $(0, d)$ are mapped into the point $d/2$. The function $q(x)$ represents an eight-level uniform quantizer.

Example 4.28

Consider the linear function $c(x) = ax + b$, where a and b are constants. This function arises in many situations. For example, $c(x)$ could be the cost associated with the quantity x, with the constant a being the cost per unit of x, and b being a fixed cost component. In a signal processing context, $c(x) = ax$ could be the amplified version (if $a > 1$) or attenuated version (if $a < 1$) of the voltage x.

The probability of an event C involving Y is equal to the probability of the equivalent event B of values of X such that $g(X)$ is in C:

$$P[Y \text{ in } C] = P[g(X) \text{ in } C] = P[X \text{ in } B].$$

Three types of equivalent events are useful in determining the cdf and pdf of $Y = g(X)$: (1) The event $\{g(X) = y_k\}$ is used to determine the magnitude of the jump at a point y_k where the cdf of Y is known to have a discontinuity; (2) the event $\{g(X) \leq y\}$ is used to find the cdf of Y directly; and (3) the event $\{y < g(X) \leq y + h\}$ is useful in determining the pdf of Y. We will demonstrate the use of these three methods in a series of examples.

The next two examples demonstrate how the pmf is computed in cases where $Y = g(X)$ is discrete. In the first example, X is discrete. In the second example, X is continuous.

Example 4.29

Let X be the number of active speakers in a group of N independent speakers. Let p be the probability that a speaker is active. In Example 2.39 it was shown that X has a binomial distribution with parameters N and p. Suppose that a voice transmission system can transmit up to M voice signals at a time, and that when X exceeds M, $X - M$ randomly selected signals are discarded. Let Y be the number of signals discarded, then

$$Y = (X - M)^+.$$

Y takes on values from the set $S_Y = \{0, 1, \dots, N - M\}$. Y will equal zero whenever X is less than or equal to M, and Y will equal $k > 0$ when X is equal to $M + k$. Therefore

$$P[Y = 0] = P[X \text{ in } \{0, 1, \dots, M\}] = \sum_{j=0}^{M} p_j$$

and

$$P[Y = k] = P[X = M + k] = p_{M+k} \qquad 0 < k \leq N - M,$$

where p_j is the pmf of X.

Example 4.30

Let X be a sample voltage of a speech waveform, and suppose that X has a uniform distribution in the interval $[-4d, 4d]$. Let $Y = q(X)$, where the quantizer input-output characteristic is as shown in Fig. 4.10. Find the pmf for Y.

The event $\{Y = q\}$ for q in S_Y is equivalent to the event $\{X \text{ in } I_q\}$, where I_q is an interval of points mapped into the representation point q. The pmf of Y is therefore found by evaluating

$$P[Y = q] = \int_{I_q} f_X(t)\, dt.$$

It is easy to see that the representation point has an interval of length d mapped into it. Thus the eight possible outputs are equiprobable, that is, $P[Y = q] = 1/8$ for q in S_Y.

In Example 4.30, each constant section of the function $q(X)$ produces a delta function in the pdf of Y. In general, if the function $g(X)$ is constant during certain intervals and if the pdf of X is nonzero in these intervals, then the pdf of Y will contain delta functions. Y will then be either discrete or of mixed type.

The cdf of Y is defined as the probability of the event $\{Y \le y\}$. In principle, it can always be obtained by finding the probability of the equivalent event $\{g(X) \le y\}$ as shown in the next examples.

Example 4.31 A Linear Function

Let the random variable Y be defined by

$$Y = aX + b,$$

where a is a nonzero constant. Suppose that X has cdf $F_X(x)$, then find $F_Y(y)$.

The event $\{Y \le y\}$ occurs when $A = \{aX + b \le y\}$ occurs. If $a > 0$, then $A = \{X \le (y - b)/a\}$ (see Fig. 4.11), and thus

$$F_Y(y) = P\left[X \le \frac{y-b}{a}\right] = F_X\left(\frac{y-b}{a}\right) \qquad a > 0.$$

On the other hand, if $a < 0$, then $A = \{X \ge (y - b)/a\}$, and

$$F_Y(y) = P\left[X \ge \frac{y-b}{a}\right] = 1 - F_X\left(\frac{y-b}{a}\right) \qquad a < 0.$$

We can obtain the pdf of Y by differentiating with respect to y. To do this we need to use the chain rule for derivatives:

$$\frac{dF}{dy} = \frac{dF}{du}\frac{du}{dy},$$

where u is the argument of F. In this case, $u = (y - b)/a$, and we then obtain

$$f_Y(y) = \frac{1}{a} f_X\left(\frac{y-b}{a}\right) \qquad a > 0$$

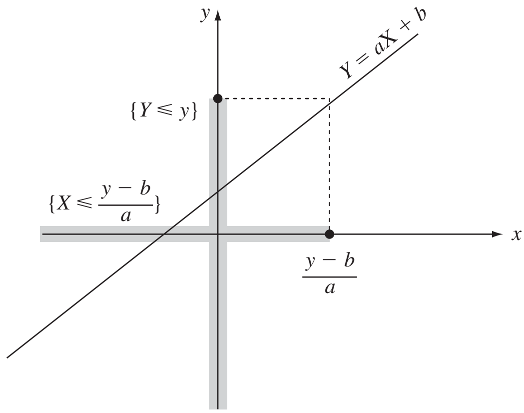


FIGURE 4.11
The equivalent event for $\{Y \leq y\}$ is the event $\{X \leq (y - b)/a\}$, if $a > 0$.

and

$$f_Y(y) = \frac{1}{-a} f_X\left(\frac{y-b}{a}\right) \qquad a < 0.$$

The above two results can be written compactly as

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right). \tag{4.67}$$

Example 4.32 A Linear Function of a Gaussian Random Variable

Let X be a random variable with a Gaussian pdf with mean m and standard deviation σ:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-m)^2/2\sigma^2} \qquad -\infty < x < \infty. \tag{4.68}$$

Let $Y = aX + b$, then find the pdf of Y.

Substitution of Eq. (4.68) into Eq. (4.67) yields

$$f_Y(y) = \frac{1}{\sqrt{2\pi}|a\sigma|} e^{-(y-b-am)^2/2(a\sigma)^2}.$$

Note that Y also has a Gaussian distribution with mean $b + am$ and standard deviation $|a|\,\sigma$. Therefore *a linear function of a Gaussian random variable is also a Gaussian random variable.*

Example 4.33

Let the random variable Y be defined by

$$Y = X^2,$$

where X is a continuous random variable. Find the cdf and pdf of Y.

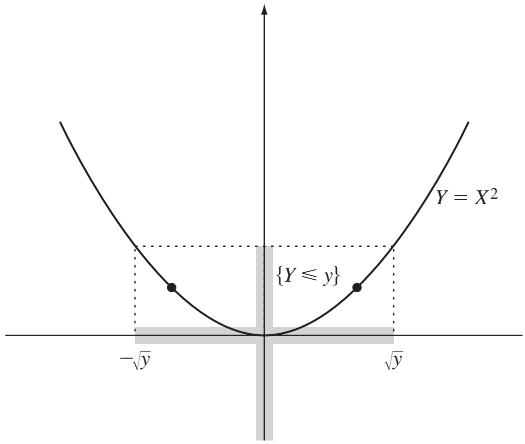


FIGURE 4.12
The equivalent event for $\{Y \leq y\}$ is the event $\{-\sqrt{y} \leq X \leq \sqrt{y}\}$, if $y \geq 0$.

The event $\{Y \leq y\}$ occurs when $\{X^2 \leq y\}$ or equivalently when $\{-\sqrt{y} \leq X \leq \sqrt{y}\}$ for y nonnegative; see Fig. 4.12. The event is null when y is negative. Thus

$$F_Y(y) = \begin{cases} 0 & y < 0 \\ F_X(\sqrt{y}) - F_X(-\sqrt{y}) & y > 0 \end{cases}$$

and differentiating with respect to y,

$$f_Y(y) = \frac{f_X(\sqrt{y})}{2\sqrt{y}} - \frac{f_X(-\sqrt{y})}{-2\sqrt{y}} \qquad y > 0$$
$$= \frac{f_X(\sqrt{y})}{2\sqrt{y}} + \frac{f_X(-\sqrt{y})}{2\sqrt{y}}. \tag{4.69}$$

Example 4.34 A Chi-Square Random Variable

Let X be a Gaussian random variable with mean $m = 0$ and standard deviation $\sigma = 1$. X is then said to be a standard normal random variable. Let $Y = X^2$. Find the pdf of Y.

Substitution of Eq. (4.68) into Eq. (4.69) yields

$$f_Y(y) = \frac{e^{-y/2}}{\sqrt{2y\pi}} \qquad y \geq 0. \tag{4.70}$$

From Table 4.1 we see that $f_Y(y)$ is the pdf of a *chi-square random variable with one degree of freedom.*

The result in Example 4.33 suggests that if the equation $y_0 = g(x)$ has n solutions, $x_0, x_1, \ldots, x_n$, then $f_Y(y_0)$ will be equal to n terms of the type on the right-hand

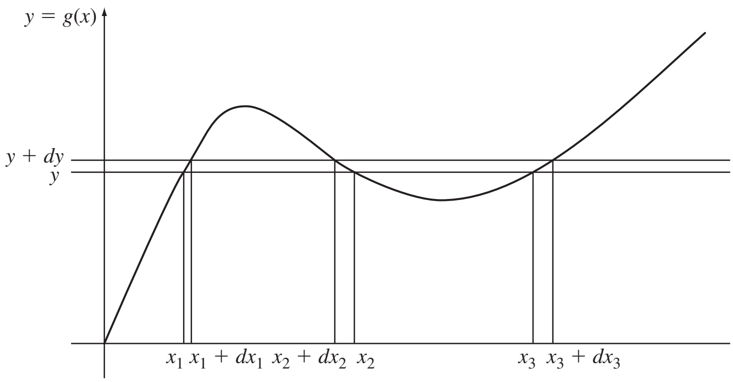


FIGURE 4.13
The equivalent event of $\{y < Y < y + dy\}$ is $\{x_1 < X < x_1 + dx_1\}$ $\cup \{x_2 + dx_2 < X < x_2\} \cup \{x_3 < X < x_3 + dx_3\}$.

side of Eq. (4.69). We now show that this is generally true by using a method for directly obtaining the pdf of Y in terms of the pdf of X.

Consider a nonlinear function $Y = g(X)$ such as the one shown in Fig. 4.13. Consider the event $C_y = \{y < Y < y + dy\}$ and let B_y be its equivalent event. For y indicated in the figure, the equation $g(x) = y$ has three solutions x_1, x_2, and x_3, and the equivalent event B_y has a segment corresponding to each solution:

$$B_y = \{x_1 < X < x_1 + dx_1\} \cup \{x_2 + dx_2 < X < x_2\} \cup \{x_3 < X < x_3 + dx_3\}.$$

The probability of the event C_y is approximately

$$P[C_y] = f_Y(y)|dy|, \tag{4.71}$$

where $|dy|$ is the length of the interval $y < Y \le y + dy$. Similarly, the probability of the event B_y is approximately

$$P[B_y] = f_X(x_1)|dx_1| + f_X(x_2)|dx_2| + f_X(x_3)|dx_3|. \tag{4.72}$$

Since C_y and B_y are equivalent events, their probabilities must be equal. By equating Eqs. (4.71) and (4.72) we obtain

$$f_Y(y) = \sum_k \frac{f_X(x)}{|dy/dx|}\bigg|_{x=x_k} \tag{4.73}$$

$$= \sum_k f_X(x)\left|\frac{dx}{dy}\right|\bigg|_{x=x_k}. \tag{4.74}$$

It is clear that if the equation $g(x) = y$ has n solutions, the expression for the pdf of Y at that point is given by Eqs. (4.73) and (4.74), and contains n terms.

Example 4.35

Let $Y = X^2$ as in Example 4.34. For $y \geq 0$, the equation $y = x^2$ has two solutions, $x_0 = \sqrt{y}$ and $x_1 = -\sqrt{y}$, so Eq. (4.73) has two terms. Since $dy/dx = 2x$, Eq. (4.73) yields

$$f_Y(y) = \frac{f_X(\sqrt{y})}{2\sqrt{y}} + \frac{f_X(-\sqrt{y})}{2\sqrt{y}}.$$

This result is in agreement with Eq. (4.69). To use Eq. (4.74), we note that

$$\frac{dx}{dy} = \frac{d}{dy} \pm \sqrt{y} = \pm\frac{1}{2\sqrt{y}},$$

which when substituted into Eq. (4.74) then yields Eq. (4.69) again.

Example 4.36 Amplitude Samples of a Sinusoidal Waveform

Let $Y = \cos(X)$, where X is uniformly distributed in the interval $(0, 2\pi]$. Y can be viewed as the sample of a sinusoidal waveform at a random instant of time that is uniformly distributed over the period of the sinusoid. Find the pdf of Y.

It can be seen in Fig. 4.14 that for $-1 < y < 1$ the equation $y = \cos(x)$ has two solutions in the interval of interest, $x_0 = \cos^{-1}(y)$ and $x_1 = 2\pi - x_0$. Since (see an introductory calculus textbook)

$$\left.\frac{dy}{dx}\right|_{x_0} = -\sin(x_0) = -\sin(\cos^{-1}(y)) = -\sqrt{1 - y^2},$$

and since $f_X(x) = 1/2\pi$ in the interval of interest, Eq. (4.73) yields

$$f_Y(y) = \frac{1}{2\pi\sqrt{1 - y^2}} + \frac{1}{2\pi\sqrt{1 - y^2}}$$
$$= \frac{1}{\pi\sqrt{1 - y^2}} \qquad \text{for } -1 < y < 1.$$

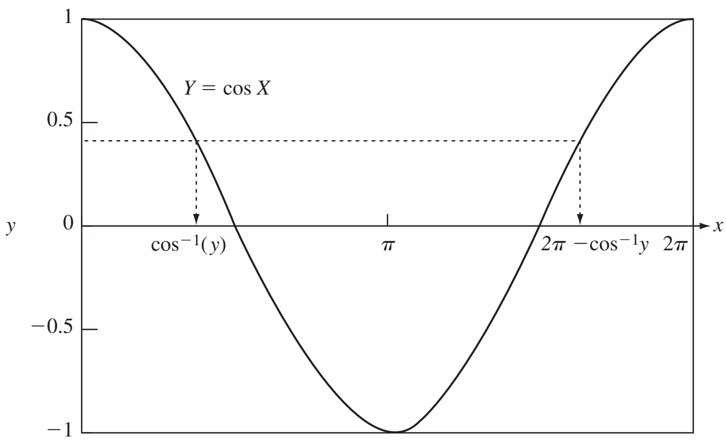


FIGURE 4.14
$y = \cos x$ has two roots in the interval $(0, 2\pi)$.

The cdf of Y is found by integrating the above:

$$F_Y(y) = \begin{cases} 0 & y < -1 \\ \dfrac{1}{2} + \dfrac{\sin^{-1} y}{\pi} & -1 \le y \le 1 \\ 1 & y > 1. \end{cases}$$

Y is said to have the **arcsine distribution**.

4.6 THE MARKOV AND CHEBYSHEV INEQUALITIES

In general, the mean and variance of a random variable do not provide enough information to determine the cdf/pdf. However, the mean and variance of a random variable X do allow us to obtain bounds for probabilities of the form $P[|X| \ge t]$. Suppose first that X is a nonnegative random variable with mean $E[X]$. The **Markov inequality** then states that

$$P[X \ge a] \le \frac{E[X]}{a} \qquad \text{for } X \text{ nonnegative.} \tag{4.75}$$

We obtain Eq. (4.75) as follows:

$$E[X] = \int_0^a t f_X(t)\,dt + \int_a^\infty t f_X(t)\,dt \ge \int_a^\infty t f_X(t)\,dt$$

$$\ge \int_a^\infty a f_X(t)\,dt = aP[X \ge a].$$

The first inequality results from discarding the integral from zero to a; the second inequality results from replacing t with the smaller number a.

Example 4.37

The mean height of children in a kindergarten class is 3 feet, 6 inches. Find the bound on the probability that a kid in the class is taller than 9 feet. The Markov inequality gives $P[H \ge 9] \le 42/108 = .389$.

The bound in the above example appears to be ridiculous. However, a bound, by its very nature, must take the worst case into consideration. One can easily construct a random variable for which the bound given by the Markov inequality is exact. The reason we know that the bound in the above example is ridiculous is that we have knowledge about the variability of the children's height about their mean.

Now suppose that the mean $E[X] = m$ and the variance $\text{VAR}[X] = \sigma^2$ of a random variable are known, and that we are interested in bounding $P[|X - m| \ge a]$. The **Chebyshev inequality** states that

$$P[|X - m| \ge a] \le \frac{\sigma^2}{a^2}. \tag{4.76}$$

The Chebyshev inequality is a consequence of the Markov inequality. Let $D^2 = (X - m)^2$ be the squared deviation from the mean. Then the Markov inequality applied to D^2 gives

$$P[D^2 \geq a^2] \leq \frac{E[(X - m)^2]}{a^2} = \frac{\sigma^2}{a^2}.$$

Equation (4.76) follows when we note that $\{D^2 \geq a^2\}$ and $\{|X - m| \geq a\}$ are equivalent events.

Suppose that a random variable X has zero variance; then the Chebyshev inequality implies that

$$P[X = m] = 1, \tag{4.77}$$

that is, the random variable is equal to its mean with probability one. In other words, X is equal to the constant m in almost all experiments.

Example 4.38

The mean response time and the standard deviation in a multi-user computer system are known to be 15 seconds and 3 seconds, respectively. Estimate the probability that the response time is more than 5 seconds from the mean.

The Chebyshev inequality with $m = 15$ seconds, $\sigma = 3$ seconds, and $a = 5$ seconds gives

$$P[|X - 15| \geq 5] \leq \frac{9}{25} = .36.$$

Example 4.39

If X has mean m and variance σ^2, then the Chebyshev inequality for $a = k\sigma$ gives

$$P[|X - m| \geq k\sigma] \leq \frac{1}{k^2}.$$

Now suppose that we know that X is a Gaussian random variable, then for $k = 2$, $P[|X - m| \geq 2\sigma] = .0456$, whereas the Chebyshev inequality gives the upper bound .25.

Example 4.40 Chebyshev Bound Is Tight

Let the random variable X have $P[X = -v] = P[X = v] = 0.5$. The mean is zero and the variance is $\text{VAR}[X] = E[X^2] = (-v)^2\, 0.5 + v^2\, 0.5 = v^2$.

Note that $P[|X| \geq v] = 1$. The Chebyshev inequality states:

$$P[|X| \geq v] \leq 1 - \frac{\text{VAR}[X]}{v^2} = 1.$$

We see that the bound and the exact value are in agreement, so the bound is tight.

We see from Example 4.38 that for certain random variables, the Chebyshev inequality can give rather loose bounds. Nevertheless, the inequality is useful in situations in which we have no knowledge about the distribution of a given random variable other than its mean and variance. In Section 7.2, we will use the Chebyshev inequality to prove that the arithmetic average of independent measurements of the same random variable is highly likely to be close to the expected value of the random variable when the number of measurements is large. Problems 4.100 and 4.101 give examples of this result.

If more information is available than just the mean and variance, then it is possible to obtain bounds that are tighter than the Markov and Chebyshev inequalities. Consider the Markov inequality again. The region of interest is $A = \{t \geq a\}$, so let $I_A(t)$ be the indicator function, that is, $I_A(t) = 1$ if $t \in A$ and $I_A(t) = 0$ otherwise. The key step in the derivation is to note that $t/a \geq 1$ in the region of interest. In effect we bounded $I_A(t)$ by t/a as shown in Fig. 4.15. We then have:

$$P[X \geq a] = \int_0^\infty I_A(t) f_X(t)\, dt \leq \int_0^\infty \frac{t}{a} f_X(t)\, dt = \frac{E[X]}{a}.$$

By changing the upper bound on $I_A(t)$, we can obtain different bounds on $P[X \geq a]$. Consider the bound $I_A(t) \leq e^{s(t-a)}$, also shown in Fig. 4.15, where $s > 0$. The resulting bound is:

$$P[X \geq a] = \int_0^\infty I_A(t) f_X(t)\, dt \leq \int_0^\infty e^{s(t-a)} f_X(t)\, dt$$

$$= e^{-sa} \int_0^\infty e^{st} f_X(t)\, dt = e^{-sa} E[e^{sX}]. \tag{4.78}$$

This bound is called the **Chernoff bound**, which can be seen to depend on the expected value of an exponential function of X. This function is called the moment generating function and is related to the transforms that are introduced in the next section. We develop the Chernoff bound further in the next section.

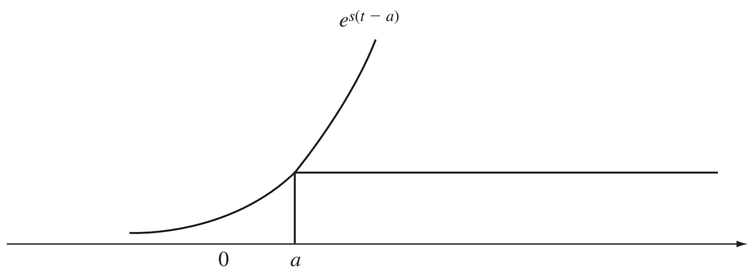


FIGURE 4.15
Bounds on indicator function for $A = \{t \geq a\}$.

4.7 TRANSFORM METHODS

In the old days, before calculators and computers, it was very handy to have logarithm tables around if your work involved performing a large number of multiplications. If you wanted to multiply the numbers x and y, you looked up $\log(x)$ and $\log(y)$, *added* $\log(x)$ and $\log(y)$, and then looked up the inverse logarithm of the result. You probably remember from grade school that longhand multiplication is more tedious and error-prone than addition. Thus logarithms were very useful as a computational aid.

Transform methods are extremely useful computational aids in the solution of equations that involve derivatives and integrals of functions. In many of these problems, the solution is given by the convolution of two functions: $f_1(x) * f_2(x)$. We will define the convolution operation later. For now, all you need to know is that finding the convolution of two functions can be more tedious and error-prone than longhand multiplication! In this section we introduce transforms that map the function $f_k(x)$ into another function $\mathcal{F}_k(\omega)$, and that satisfy the property that $\mathcal{F}[f_1(x) * f_2(x)] = \mathcal{F}_1(\omega)\mathcal{F}_2(\omega)$. In other words, the transform of the convolution is equal to the product of the individual transforms. Therefore transforms allow us to replace the convolution operation by the much simpler multiplication operation. The transform expressions introduced in this section will prove very useful when we consider sums of random variables in Chapter 7.

4.7.1 The Characteristic Function

The **characteristic function** of a random variable X is defined by

$$\Phi_X(\omega) = E[e^{j\omega X}] \tag{4.79a}$$

$$= \int_{-\infty}^{\infty} f_X(x)e^{j\omega x}\,dx, \tag{4.79b}$$

where $j = \sqrt{-1}$ is the imaginary unit number. The two expressions on the right-hand side motivate two interpretations of the characteristic function. In the first expression, $\Phi_X(\omega)$ can be viewed as the expected value of a function of X, $e^{j\omega X}$, in which the parameter ω is left unspecified. In the second expression, $\Phi_X(\omega)$ is simply the Fourier transform of the pdf $f_X(x)$ (with a reversal in the sign of the exponent). Both of these interpretations prove useful in different contexts.

If we view $\Phi_X(\omega)$ as a Fourier transform, then we have from the Fourier transform inversion formula that the pdf of X is given by

$$f_X(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \Phi_X(\omega)e^{-j\omega x}\,d\omega. \tag{4.80}$$

It then follows that every pdf and its characteristic function form a unique Fourier transform pair. Table 4.1 gives the characteristic function of some continuous random variables.

Example 4.41 Exponential Random Variable

The characteristic function for an exponentially distributed random variable with parameter λ is given by

$$\Phi_X(\omega) = \int_0^\infty \lambda e^{-\lambda x} e^{j\omega x}\,dx = \int_0^\infty \lambda e^{-(\lambda - j\omega)x}\,dx$$
$$= \frac{\lambda}{\lambda - j\omega}.$$

If X is a discrete random variable, substitution of Eq. (4.20) into the definition of $\Phi_X(\omega)$ gives

$$\Phi_X(\omega) = \sum_k p_X(x_k) e^{j\omega x_k} \qquad \text{discrete random variables.}$$

Most of the time we deal with discrete random variables that are integer-valued. The characteristic function is then

$$\Phi_X(\omega) = \sum_{k=-\infty}^{\infty} p_X(k) e^{j\omega k} \qquad \text{integer-valued random variables.} \tag{4.81}$$

Equation (4.81) is the **Fourier transform of the sequence** $p_X(k)$. Note that the Fourier transform in Eq. (4.81) is a periodic function of ω with period 2π, since $e^{j(\omega+2\pi)k} = e^{j\omega k}e^{jk2\pi}$ and $e^{jk2\pi} = 1$. Therefore the characteristic function of integer-valued random variables is a periodic function of ω. The following inversion formula allows us to recover the probabilities $p_X(k)$ from $\Phi_X(\omega)$:

$$p_X(k) = \frac{1}{2\pi}\int_0^{2\pi} \Phi_X(\omega) e^{-j\omega k}\,d\omega \qquad k = 0, \pm 1, \pm 2, \ldots \tag{4.82}$$

Indeed, a comparison of Eqs. (4.81) and (4.82) shows that the $p_X(k)$ are simply the coefficients of the Fourier series of the periodic function $\Phi_X(\omega)$.

Example 4.42 Geometric Random Variable

The characteristic function for a geometric random variable is given by

$$\Phi_X(\omega) = \sum_{k=0}^{\infty} pq^k e^{j\omega k} = p\sum_{k=0}^{\infty} (qe^{j\omega})^k$$
$$= \frac{p}{1 - qe^{j\omega}}.$$

Since $f_X(x)$ and $\Phi_X(\omega)$ form a transform pair, we would expect to be able to obtain the moments of X from $\Phi_X(\omega)$. The **moment theorem** states that the moments of

X are given by

$$E[X^n] = \frac{1}{j^n}\frac{d^n}{d\omega^n}\Phi_X(\omega)\bigg|_{\omega=0}. \tag{4.83}$$

To show this, first expand $e^{j\omega x}$ in a power series in the definition of $\Phi_X(\omega)$:

$$\Phi_X(\omega) = \int_{-\infty}^{\infty} f_X(x)\left\{1 + j\omega X + \frac{(j\omega X)^2}{2!} + \cdots\right\} dx.$$

Assuming that all the moments of X are finite and that the series can be integrated term by term, we obtain

$$\Phi_X(\omega) = 1 + j\omega E[X] + \frac{(j\omega)^2 E[X^2]}{2!} + \cdots + \frac{(j\omega)^n E[X^n]}{n!} + \cdots.$$

If we differentiate the above expression once and evaluate the result at $\omega = 0$ we obtain

$$\frac{d}{d\omega}\Phi_X(\omega)\bigg|_{\omega=0} = jE[X].$$

If we differentiate n times and evaluate at $\omega = 0$, we finally obtain

$$\frac{d^n}{d\omega^n}\Phi_X(\omega)\bigg|_{\omega=0} = j^n E[X^n],$$

which yields Eq. (4.83).

Note that when the above power series converges, the characteristic function and hence the pdf by Eq. (4.80) are completely determined by the moments of X.

Example 4.43

To find the mean of an exponentially distributed random variable, we differentiate $\Phi_X(\omega) = \lambda(\lambda - j\omega)^{-1}$ once, and obtain

$$\Phi'_X(\omega) = \frac{\lambda j}{(\lambda - j\omega)^2}.$$

The moment theorem then implies that $E[X] = \Phi'_X(0)/j = 1/\lambda$.

If we take two derivatives, we obtain

$$\Phi''_X(\omega) = \frac{-2\lambda}{(\lambda - j\omega)^3},$$

so the second moment is then $E[X^2] = \Phi''_X(0)/j^2 = 2/\lambda^2$. The variance of X is then given by

$$\text{VAR}[X] = E[X^2] - E[X]^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

Example 4.44 Chernoff Bound for Gaussian Random Variable

Let X be a Gaussian random variable with mean m and variance σ^2. Find the Chernoff bound for X.

The Chernoff bound (Eq. 4.78) depends on the moment generating function:

$$E[e^{sX}] = \Phi_X(-js).$$

In terms of the characteristic function the bound is given by:

$$P[X \geq a] \leq e^{-sa}\Phi_X(-js) \quad \text{for} \quad s \geq 0.$$

The parameter s can be selected to minimize the upper bound.

The bound for the Gaussian random variable is:

$$P[X \geq a] \leq e^{-sa}e^{ms+\sigma^2 s^2/2} = e^{-s(a-m)+\sigma^2 s^2/2} \quad \text{for} \quad s \geq 0.$$

We minimize the upper bound by minimizing the exponent:

$$0 = \frac{d}{ds}(-s(a-m) + \sigma^2 s^2/2) \quad \text{which implies } s = \frac{a-m}{\sigma^2}.$$

The resulting upper bound is:

$$P[X \geq a] = Q\left(\frac{a-m}{\sigma}\right) \leq e^{-(a-m)^2/2\sigma^2}.$$

This bound is much better than the Chebyshev bound and is similar to the estimate given in Eq. (4.54).

4.7.2 The Probability Generating Function

In problems where random variables are nonnegative, it is usually more convenient to use the z-transform or the Laplace transform. The **probability generating function** $G_N(z)$ of a nonnegative integer-valued random variable N is defined by

$$G_N(z) = E[z^N] \tag{4.84a}$$

$$= \sum_{k=0}^{\infty} p_N(k) z^k. \tag{4.84b}$$

The first expression is the expected value of the function of N, z^N. The second expression is the z-transform of the pmf (with a sign change in the exponent). Table 3.1 shows the probability generating function for some discrete random variables. Note that the characteristic function of N is given by $\Phi_N(\omega) = G_N(e^{j\omega})$.

Using a derivation similar to that used in the moment theorem, it is easy to show that the pmf of N is given by

$$p_N(k) = \frac{1}{k!}\frac{d^k}{dz^k}G_N(z)\bigg|_{z=0}. \tag{4.85}$$

This is why $G_N(z)$ is called the probability generating function. By taking the first two derivatives of $G_N(z)$ and evaluating the result at $z = 1$, it is possible to find the first two moments of X:

$$\frac{d}{dz}G_N(z)\bigg|_{z=1} = \sum_{k=0}^{\infty} p_N(k)kz^{k-1}\bigg|_{z=1} = \sum_{k=0}^{\infty} kp_N(k) = E[N]$$

and

$$\frac{d^2}{dz^2}G_N(z)\bigg|_{z=1} = \sum_{k=0}^{\infty} p_N(k)k(k-1)z^{k-2}\bigg|_{z=1}$$

$$= \sum_{k=0}^{\infty} k(k-1)p_N(k) = E[N(N-1)] = E[N^2] - E[N].$$

Thus the mean and variance of X are given by

$$E[N] = G_N'(1) \tag{4.86}$$

and

$$\text{VAR}[N] = G_N''(1) + G_N'(1) - (G_N'(1))^2. \tag{4.87}$$

Example 4.45 Poisson Random Variable

The probability generating function for the Poisson random variable with parameter α is given by

$$G_N(z) = \sum_{k=0}^{\infty} \frac{\alpha^k}{k!}e^{-\alpha}z^k = e^{-\alpha}\sum_{k=0}^{\infty}\frac{(\alpha z)^k}{k!}$$

$$= e^{-\alpha}e^{\alpha z} = e^{\alpha(z-1)}.$$

The first two derivatives of $G_N(z)$ are given by

$$G_N'(z) = \alpha e^{\alpha(z-1)}$$

and

$$G_N''(z) = \alpha^2 e^{\alpha(z-1)}.$$

Therefore the mean and variance of the Poisson are

$$E[N] = \alpha$$

$$\text{VAR}[N] = \alpha^2 + \alpha - \alpha^2 = \alpha.$$

4.7.3 The Laplace Transform of the pdf

In queueing theory one deals with service times, waiting times, and delays. All of these are nonnegative continuous random variables. It is therefore customary to work with the **Laplace transform** of the pdf,

$$X^*(s) = \int_0^{\infty} f_X(x)e^{-sx}\,dx = E[e^{-sX}]. \tag{4.88}$$

Note that $X^*(s)$ can be interpreted as a Laplace transform of the pdf or as an expected value of a function of X, e^{-sX}.

The moment theorem also holds for $X^*(s)$:

$$E[X^n] = (-1)^n \frac{d^n}{ds^n} X^*(s)\bigg|_{s=0}. \tag{4.89}$$

Example 4.46 Gamma Random Variable

The Laplace transform of the gamma pdf is given by

$$\begin{aligned} X^*(s) &= \int_0^\infty \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x} e^{-sx}}{\Gamma(\alpha)}\, dx = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-(\lambda+s)x}\, dx \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{1}{(\lambda+s)^\alpha} \int_0^\infty y^{\alpha-1} e^{-y}\, dy = \frac{\lambda^\alpha}{(\lambda+s)^\alpha}, \end{aligned}$$

where we used the change of variable $y = (\lambda + s)x$. We can then obtain the first two moments of X as follows:

$$E[X] = -\frac{d}{ds} \frac{\lambda^\alpha}{(\lambda+s)^\alpha}\bigg|_{s=0} = \frac{\alpha\lambda^\alpha}{(\lambda+s)^{\alpha+1}}\bigg|_{s=0} = \frac{\alpha}{\lambda}$$

and

$$E[X^2] = \frac{d^2}{ds^2} \frac{\lambda^\alpha}{(\lambda+s)^\alpha}\bigg|_{s=0} = \frac{\alpha(\alpha+1)\lambda^\alpha}{(\lambda+s)^{\alpha+2}}\bigg|_{s=0} = \frac{\alpha(\alpha+1)}{\lambda^2}.$$

Thus the variance of X is

$$\text{VAR}(X) = E[X^2] - E[X]^2 = \frac{\alpha}{\lambda^2}.$$

4.8 BASIC RELIABILITY CALCULATIONS

In this section we apply some of the tools developed so far to the calculation of measures that are of interest in assessing the reliability of systems. We also show how the reliability of a system can be determined in terms of the reliability of its components.

4.8.1 The Failure Rate Function

Let T be the lifetime of a component, a subsystem, or a system. The **reliability** at time t is defined as the probability that the component, subsystem, or system is still functioning at time t:

$$R(t) = P[T > t]. \tag{4.90}$$

The relative frequency interpretation implies that, in a large number of components or systems, $R(t)$ is the fraction that fail after time t. The reliability can be expressed in terms of the cdf of T:

$$R(t) = 1 - P[T \le t] = 1 - F_T(t). \tag{4.91}$$

Note that the derivative of $R(t)$ gives the negative of the pdf of T:

$$R'(t) = -f_T(t). \tag{4.92}$$

The **mean time to failure (MTTF)** is given by the expected value of T:

$$E[T] = \int_0^\infty f_T(t)\,dt = \int_0^\infty R(t)\,dt,$$

where the second expression was obtained using Eqs. (4.28) and (4.91).

Suppose that we know a system is still functioning at time t; what is its future behavior? In Example 4.10, we found that the conditional cdf of T given that $T > t$ is given by

$$F_T(x|T > t) = P[T \le x|T > t]$$
$$= \begin{cases} 0 & x < t \\ \dfrac{F_T(x) - F_T(t)}{1 - F_T(t)} & x \ge t. \end{cases} \tag{4.93}$$

The pdf associated with $F_T(x|T > t)$ is

$$f_T(x|T > t) = \frac{f_T(x)}{1 - F_T(t)} \quad x \ge t. \tag{4.94}$$

Note that the denominator of Eq. (4.94) is equal to $R(t)$.

The **failure rate function** $r(t)$ is defined as $f_T(x|T > t)$ evaluated at $x = t$:

$$r(t) = f_T(t|T > t)$$
$$= \frac{-R'(t)}{R(t)}, \tag{4.95}$$

since by Eq. (4.92), $R'(t) = -f_T(t)$. The failure rate function has the following meaning:

$$P[t < T \le t + dt|T > t] = f_T(t|T > t)\,dt = r(t)\,dt. \tag{4.96}$$

In words, $r(t)\,dt$ is the probability that a component that has functioned up to time t will fail in the next dt seconds.

Example 4.47 Exponential Failure Law

Suppose a component has a constant failure rate function, say $r(t) = \lambda$. Find the pdf and the MTTF for its lifetime T.

Equation (4.95) implies that

$$\frac{R'(t)}{R(t)} = -\lambda. \tag{4.97}$$

Equation (4.97) is a first-order differential equation with initial condition $R(0) = 1$. If we integrate both sides of Eq. (4.97) from 0 to t, we obtain

$$-\int_0^t \lambda\,dt' + k = \int_0^t \frac{R'(t')}{R(t')}dt' = \ln R(t),$$

which implies that

$$R(t) = Ke^{-\lambda t}, \qquad \text{where } K = e^k.$$

The initial condition $R(0) = 1$ implies that $K = 1$. Thus

$$R(t) = e^{-\lambda t} \qquad t > 0 \tag{4.98}$$

and

$$f_T(t) = \lambda e^{-\lambda t} \qquad t > 0.$$

Thus if T has a constant failure rate function, then T is an exponential random variable. This is not surprising, since the exponential random variable satisfies the memoryless property. The MTTF $= E[T] = 1/\lambda$.

The derivation that was used in Example 4.47 can be used to show that, in general, the failure rate function and the reliability are related by

$$R(t) = \exp\left\{-\int_0^t r(t')\,dt'\right\} \tag{4.99}$$

and from Eq. (4.92),

$$f_T(t) = r(t)\exp\left\{-\int_0^t r(t')\,dt'\right\}. \tag{4.100}$$

Figure 4.16 shows the failure rate function for a typical system. Initially there may be a high failure rate due to defective parts or installation. After the "bugs" have been worked out, the system is stable and has a low failure rate. At some later point, ageing and wear effects set in, resulting in an increased failure rate. Equations (4.99) and (4.100) allow us to postulate reliability functions and the associated pdf's in terms of the failure rate function, as shown in the following example.

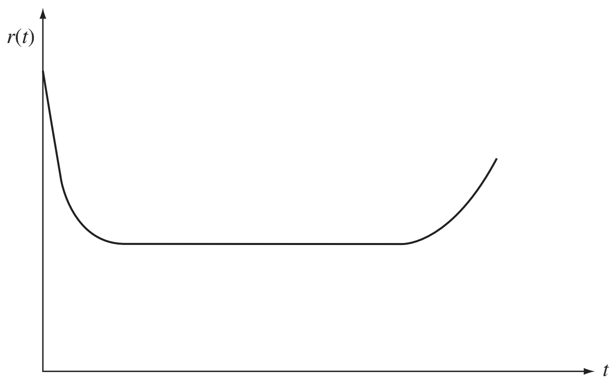


FIGURE 4.16
Failure rate function for a typical system.

Example 4.48 Weibull Failure Law

The Weibull failure law has failure rate function given by

$$r(t) = \alpha\beta t^{\beta-1}, \tag{4.101}$$

where α and β are positive constants. Equation (4.99) implies that the reliability is given by

$$R(t) = e^{-\alpha t^{\beta}}.$$

Equation (4.100) then implies that the pdf for T is

$$f_T(t) = \alpha\beta t^{\beta-1}e^{-\alpha t^{\beta}} \quad t > 0. \tag{4.102}$$

Figure 4.17 shows $f_T(t)$ for $\alpha = 1$ and several values of β. Note that $\beta = 1$ yields the exponential failure law, which has a constant failure rate. For $\beta > 1$, Eq. (4.101) gives a failure rate function that increases with time. For $\beta < 1$, Eq. (4.101) gives a failure rate function that decreases with time. Further properties of the Weibull random variable are developed in the problems.

4.8.2 Reliability of Systems

Suppose that a system consists of several components or subsystems. We now show how the reliability of a system can be computed in terms of the reliability of its subsystems if the components are assumed to fail independently of each other.

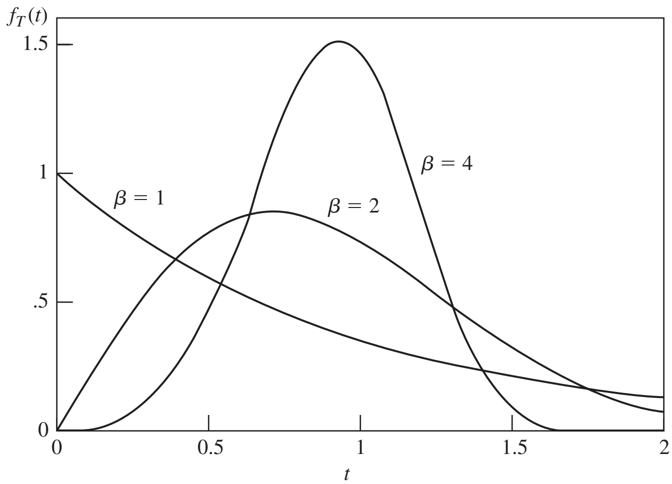


FIGURE 4.17
Probability density function of Weibull random variable, $\alpha = 1$ and $\beta = 1, 2, 4$.

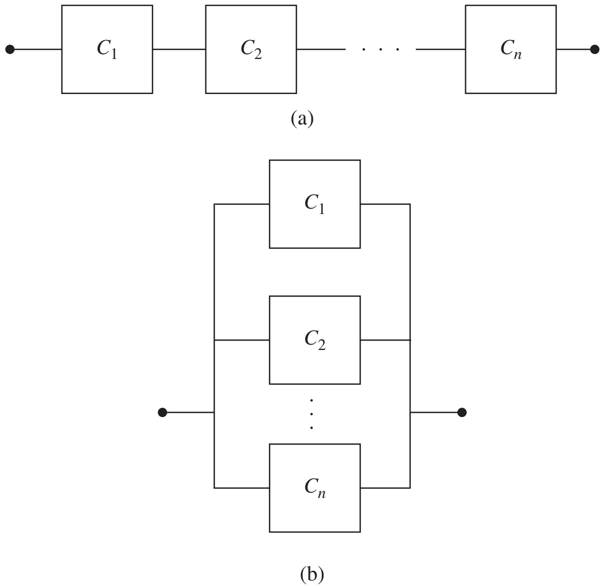


FIGURE 4.18
(a) System consisting of n components in series. (b) System consisting of n components in parallel.

Consider first a system that consists of the series arrangement of n components as shown in Fig. 4.18(a). This system is considered to be functioning only if all the components are functioning. Let A_s be the event "system functioning at time t," and let A_j be the event "jth component is functioning at time t," then the probability that the system is functioning at time t is

$$\begin{aligned} R(t) &= P[A_s] \\ &= P[A_1 \cap A_2 \cap \cdots \cap A_n] = P[A_1]P[A_2]\dots P[A_n] \\ &= R_1(t)R_2(t)\dots R_n(t), \end{aligned} \tag{4.103}$$

since $P[A_j] = R_j(t)$, the reliability function of the jth component. Since probabilities are numbers that are less than or equal to one, we see that $R(t)$ can be no more reliable than the least reliable of the components, that is, $R(t) \le \min_j R_j(t)$.

If we apply Eq. (4.99) to each of the $R_j(t)$ in Eq. (4.103), we then find that the failure rate function of a series system is given by the sum of the component failure rate functions:

$$\begin{aligned} R(t) &= \exp\left\{-\int_0^t r_1(t')\,dt'\right\}\exp\left\{-\int_0^t r_2(t')\,dt'\right\}\dots\exp\left\{-\int_0^t r_n(t')\,dt'\right\} \\ &= \exp\left\{-\int_0^t [r_1(t') + r_2(t') + \cdots + r_n(t')]\,dt'\right\}. \end{aligned}$$

Example 4.49

Suppose that a system consists of n components in series and that the component lifetimes are exponential random variables with rates $\lambda_1, \lambda_2, \dots, \lambda_n$. Find the system reliability.

From Eqs. (4.98) and (4.103), we have

$$R(t) = e^{-\lambda_1 t} e^{-\lambda_2 t} \ldots e^{-\lambda_n t}$$
$$= e^{-(\lambda_1 + \cdots + \lambda_n)t}.$$

Thus the system reliability is exponentially distributed with rate $\lambda_1 + \lambda_2 + \cdots + \lambda_n$.

Now suppose that a system consists of n components in parallel, as shown in Fig. 4.18(b). This system is considered to be functioning as long as at least one of the components is functioning. The system will *not* be functioning if and only if all the components have failed, that is,

$$P[A_s^c] = P[A_1^c]P[A_2^c] \ldots P[A_n^c].$$

Thus

$$1 - R(t) = (1 - R_1(t))(1 - R_2(t)) \ldots (1 - R_n(t)),$$

and finally,

$$R(t) = 1 - (1 - R_1(t))(1 - R_2(t)) \ldots (1 - R_n(t)). \tag{4.104}$$

Example 4.50

Compare the reliability of a single-unit system against that of a system that operates two units in parallel. Assume all units have exponentially distributed lifetimes with rate 1.

The reliability of the single-unit system is

$$R_s(t) = e^{-t}.$$

The reliability of the two-unit system is

$$R_p(t) = 1 - (1 - e^{-t})(1 - e^{-t})$$
$$= e^{-t}(2 - e^{-t}).$$

The parallel system is more reliable by a factor of

$$(2 - e^{-t}) > 1.$$

More complex configurations can be obtained by combining subsystems consisting of series and parallel components. The reliability of such systems can then be computed in terms of the subsystem reliabilities. See Example 2.35 for an example of such a calculation.

4.9 COMPUTER METHODS FOR GENERATING RANDOM VARIABLES

The computer simulation of any random phenomenon involves the generation of random variables with prescribed distributions. For example, the simulation of a queueing system involves generating the time between customer arrivals as well as the service times of each customer. Once the cdf's that model these random quantities have been selected, an algorithm for generating random variables with these cdf's must be found. MATLAB and Octave have built-in functions for generating random variables for all of the well known distributions. In this section we present the methods that are used for generating random variables. All of these methods are based on the availability of random numbers that are uniformly distributed between zero and one. Methods for generating these numbers were discussed in Section 2.7.

All of the methods for generating random variables require the evaluation of either the pdf, the cdf, or the inverse of the cdf of the random variable of interest. We can write programs to perform these evaluations, or we can use the functions available in programs such as MATLAB and Octave. The following example shows some typical evaluations for the Gaussian random variable.

Example 4.51 Evaluation of pdf, cdf, and Inverse cdf

Let X be a Gaussian random variable with mean 1 and variance 2. Find the pdf at $x = 7$. Find the cdf at $x = -2$. Find the value of x at which the cdf $= 0.25$.

The following commands show how these results are obtained using Octave.

```
> normal_pdf (7, 1, 2)
ans = 3.4813e-05
> normal_cdf (−2, 1, 2)
ans = 0.016947
> normal_inv (0.25, 1, 2)
ans = 0.046127
```

4.9.1 The Transformation Method

Suppose that U is uniformly distributed in the interval $[0, 1]$. Let $F_X(x)$ be the cdf of the random variable we are interested in generating. Define the random variable, $Z = F_X^{-1}(U)$; that is, first U is selected and then Z is found as indicated in Fig. 4.19. The cdf of Z is

$$P[Z \le x] = P[F_X^{-1}(U) \le x] = P[U \le F_X(x)].$$

But if U is uniformly distributed in [0, 1] and $0 \le h \le 1$, then $P[U \le h] = h$ (see Example 4.6). Thus

$$P[Z \le x] = F_X(x),$$

and $Z = F_X^{-1}(U)$ has the desired cdf.

Transformation Method for Generating *X*:

1. Generate U uniformly distributed in $[0, 1]$.
2. Let $Z = F_X^{-1}(U)$.

Example 4.52 Exponential Random Variable

To generate an exponentially distributed random variable X with parameter λ, we need to invert the expression $u = F_X(x) = 1 - e^{-\lambda x}$. We obtain

$$X = -\frac{1}{\lambda}\ln(1 - U).$$

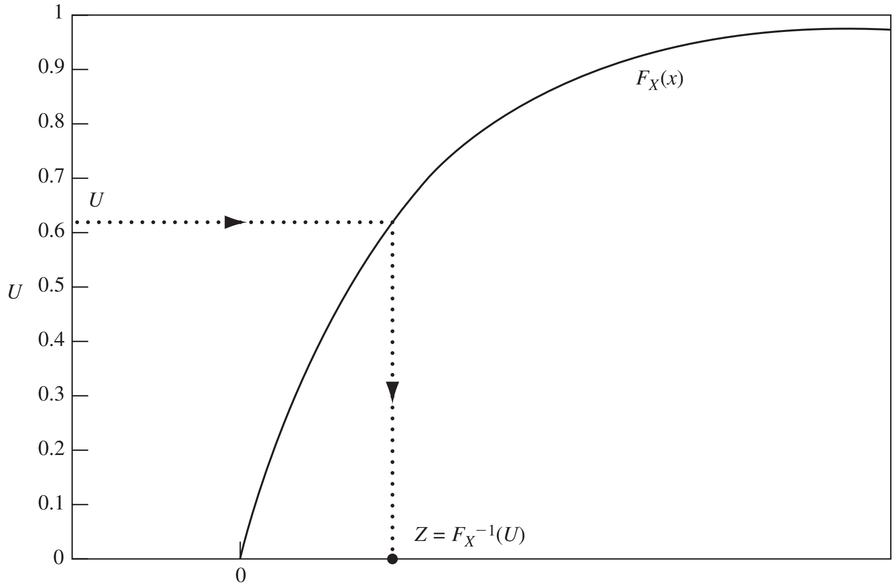


FIGURE 4.19
Transformation method for generating a random variable with cdf $F_X(x)$.

Note that we can use the simpler expression $X = -\ln(U)/\lambda$, since $1 - U$ is also uniformly distributed in $[0, 1]$. The first two lines of the Octave commands below show how to implement the transformation method to generate 1000 exponential random variables with $\lambda = 1$. Figure 4.20 shows the histogram of values obtained. In addition, the figure shows the probability that samples of the random variables fall in the corresponding histogram bins. Good correspondence between the histograms and these probabilities are observed. In Chapter 8 we introduce methods for assessing the goodness-of-fit of data to a given distribution. Both MATLAB and Octave use the transformation method in their function `exponential_rnd`.

```
> U=rand (1, 1000);          % Generate 1000 uniform random variables.
> X=-log(U);                 % Compute 1000 exponential RVs.
> K=0.25:0.5:6;
> P(1)=1-exp(-0.5)
> for i=2:12,                % The remaining lines show how to generate
> P(i)=P(i-1)*exp(-0.5)      % the histogram bins.
> end;
> stem (K, P)
> hold on
> Hist (X, K, 1)
```

4.9.2 The Rejection Method

We first consider the simple version of this algorithm and explain why it works; then we present it in its general form. Suppose that we are interested in generating a random variable Z with pdf $f_X(x)$ as shown in Fig. 4.21. In particular, we assume that: (1) the pdf is nonzero only in the interval $[0, a]$, and (2) the pdf takes on values in the range $[0, b]$. The **rejection method** in this case works as follows:

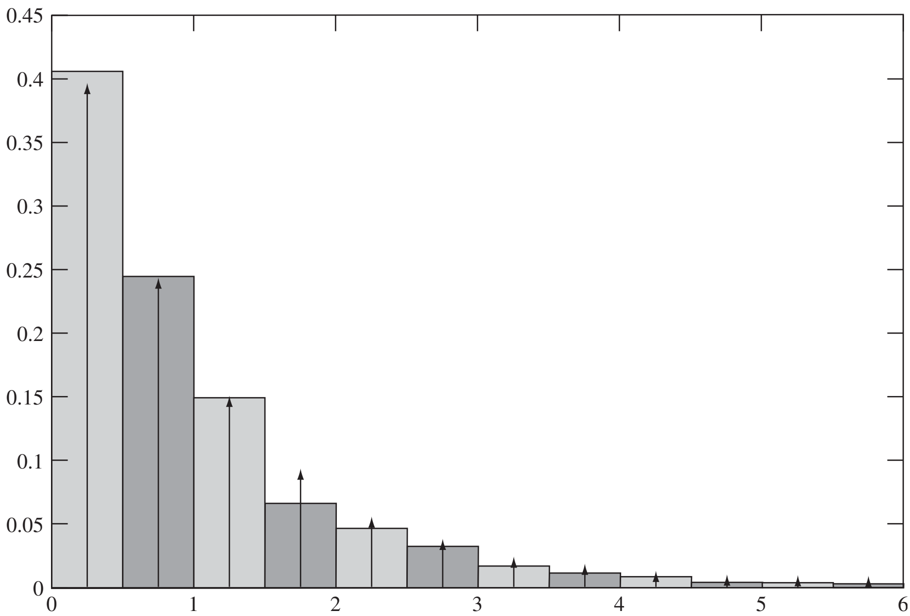


FIGURE 4.20
Histogram of 1000 exponential random variables using transformation method.

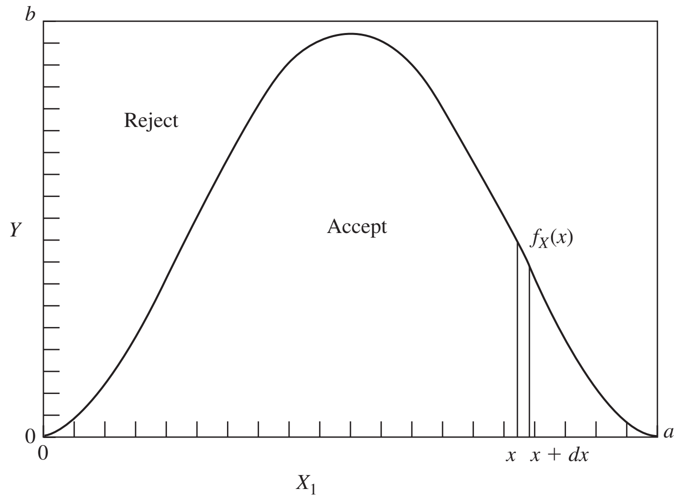


FIGURE 4.21
Rejection method for generating a random variable with pdf $f_X(x)$.

1. Generate X_1 uniform in the interval $[0, a]$.
2. Generate Y uniform in the interval $[0, b]$.
3. If $Y \leq f_X(X_1)$, then output $Z = X_1$; else, reject X_1 and return to step 1.

Note that this algorithm will perform a random number of steps before it produces the output Z.

We now show that the output Z has the desired pdf. Steps 1 and 2 select a point at random in a rectangle of width a and height b. The probability of selecting a point in any region is simply the area of the region divided by the total area of the rectangle, ab. Thus the probability of accepting X_1 is the probability of the region below $f_X(x)$ divided by ab. But the area under any pdf is 1, so we conclude that the probability of success (i.e., acceptance) is $1/ab$. Consider now the following probability:

$$\begin{aligned}
P[x < X_1 \le x + dx | X_1 \text{ is accepted}] \\
&= \frac{P[\{x < X_1 \le x + dx\} \cap \{X_1 \text{ accepted}\}]}{P[X_1 \text{ accepted}]} \\
&= \frac{\text{shaded area}/ab}{1/ab} = \frac{f_X(x)\,dx/ab}{1/ab} \\
&= f_X(x)\,dx.
\end{aligned}$$

Therefore X_1 when accepted has the desired pdf. Thus Z has the desired pdf.

Example 4.53 Generating Beta Random Variables

Show that the beta random variables with $a' = b' = 2$ can be generated using the rejection method.

The pdf of the beta random variable with $a' = b' = 2$ is similar to that shown in Fig. 4.21. This beta pdf is maximum at $x = 1/2$ and the maximum value is:

$$\frac{(1/2)^{2-1}(1/2)^{2-1}}{B(2,2)} = \frac{1/4}{\Gamma(2)\Gamma(2)/\Gamma(4)} = \frac{1/4}{1!1!/3!} = \frac{3}{2}.$$

Therefore we can generate this beta random variable using the rejection method with $b = 1.5$.

The algorithm as stated above can have two problems. First, if the rectangle does not fit snugly around $f_X(x)$, the number of X_1's that need to be generated before acceptance may be excessive. Second, the above method cannot be used if $f_X(x)$ is unbounded or if its range is not finite. The general version of this algorithm overcomes both problems. Suppose we want to generate Z with pdf $f_X(x)$. Let W be a random variable with pdf $f_W(x)$ that is *easy* to generate and such that for some constant $K > 1$,

$$K f_W(x) \ge f_X(x) \qquad \text{for all } x,$$

that is, the region under $K f_W(x)$ contains $f_X(x)$ as shown in Fig. 4.22.

Rejection Method for Generating X:

1. Generate X_1 with pdf $f_W(x)$. Define $B(X_1) = K f_W(X_1)$.
2. Generate Y uniform in $[0, B(X_1)]$.
3. If $Y \le f_X(X_1)$, then output $Z = X_1$; else reject X_1 and return to step 1.

See Problem 4.143 for a proof that Z has the desired pdf.

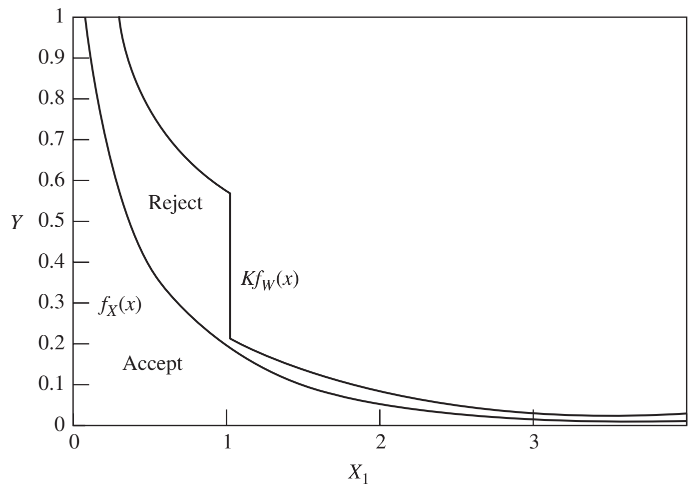


FIGURE 4.22
Rejection method for generating a random variable with gamma pdf and with $0 < \alpha < 1$.

Example 4.54 Gamma Random Variable

We now show how the rejection method can be used to generate X with gamma pdf and parameters $0 < \alpha < 1$ and $\lambda = 1$. A function $Kf_W(x)$ that "covers" $f_X(x)$ is easily obtained (see Fig. 4.22):

$$f_X(x) = \frac{x^{\alpha-1}e^{-x}}{\Gamma(\alpha)} \leq Kf_W(x) = \begin{cases} \dfrac{x^{\alpha-1}}{\Gamma(\alpha)} & 0 \leq x \leq 1 \\ \dfrac{e^{-x}}{\Gamma(\alpha)} & x > 1. \end{cases}$$

The pdf $f_W(x)$ that corresponds to the function on the right-hand side is

$$f_W(x) = \begin{cases} \dfrac{\alpha e x^{\alpha-1}}{\alpha + e} & 0 \leq x \leq 1 \\ \alpha e \dfrac{e^{-x}}{\alpha + e} & x \geq 1. \end{cases}$$

The cdf of W is

$$F_W(x) = \begin{cases} \dfrac{e x^{\alpha}}{\alpha + e} & 0 \leq x \leq 1 \\ 1 - \alpha e \dfrac{e^{-x}}{\alpha + e} & x > 1. \end{cases}$$

W is easy to generate using the transformation method, with

$$F_W^{-1}(u) = \begin{cases} \left[\dfrac{(\alpha + e)u}{e}\right]^{1/\alpha} & u \leq e/(\alpha + e) \\ -\ln\left[(\alpha + e)\dfrac{(1 - u)}{\alpha e}\right] & u > e/(\alpha + e). \end{cases}$$

We can therefore use the transformation method to generate this $f_W(x)$, and then the rejection method to generate any gamma random variable X with parameters $0 < \alpha < 1$ and $\lambda = 1$. Finally we note that if we let $W = \lambda X$, then W will be gamma with parameters α and λ. The generation of gamma random variables with $\alpha > 1$ is discussed in Problem 4.142.

Example 4.55 Implementing Rejection Method for Gamma Random Variables

Given below is an Octave function definition to implement the rejection method using the above transformation.

```
% Generate random numbers from the gamma distribution for 0 ≤ α ≤ 1.
function X = gamma_rejection_method_altone(alpha)
while (true),
    X = special_inverse(alpha);          % Step 1: Generate X with pdf f_X(x).
    B = special_pdf (X, alpha);          % Step 2: Generate Y uniform in [0, K f_X(X)].
    Y = rand.* B;
    if (Y <= fx_gamma_pdf (X, alpha)),   % Step 3: Accept or reject . . .
         break;
    end
end

% Helper function to generate random variables according to K f_Z(x).
function X = special_inverse (alpha)
u = rand;
if (u <= e./(alpha+e)),
    X = ((alpha+e).*u./e). ^ (1./alpha);
elseif (u > e./(alpha+e)),
    X = -log((alpha+e).*(1-u)./(alpha.*e));
end

% Return B in order to generate uniform variables in [0, K f_Z(X)].
function B = special_pdf (X, alpha)
if (X >=0 && X <= 1),
    B = alpha.*e.*X.^(alpha-1)./(alpha + e);
elseif (X > 1),
    B = alpha.*e.*(e. ^(-X)./(alpha + e));
end

% pdf of the gamma distribution.
% Could also use the built in gamma_pdf (X, A, B) function supplied with Octave
setting B = 1
function Y = fx_gamma_pdf (x, alpha)
y = (x.^ (alpha-1)).*(e.^ (-x))./(gamma(alpha));
```

Figure 4.23 shows the histogram of 1000 samples obtained using this function. The figure also shows the probability that the samples fall in the bins of the histogram.

We have presented the most common methods that are used to generate random variables. These methods are incorporated in the functions provided by programs such as MATLAB and Octave, so in practice you do not need to write programs to

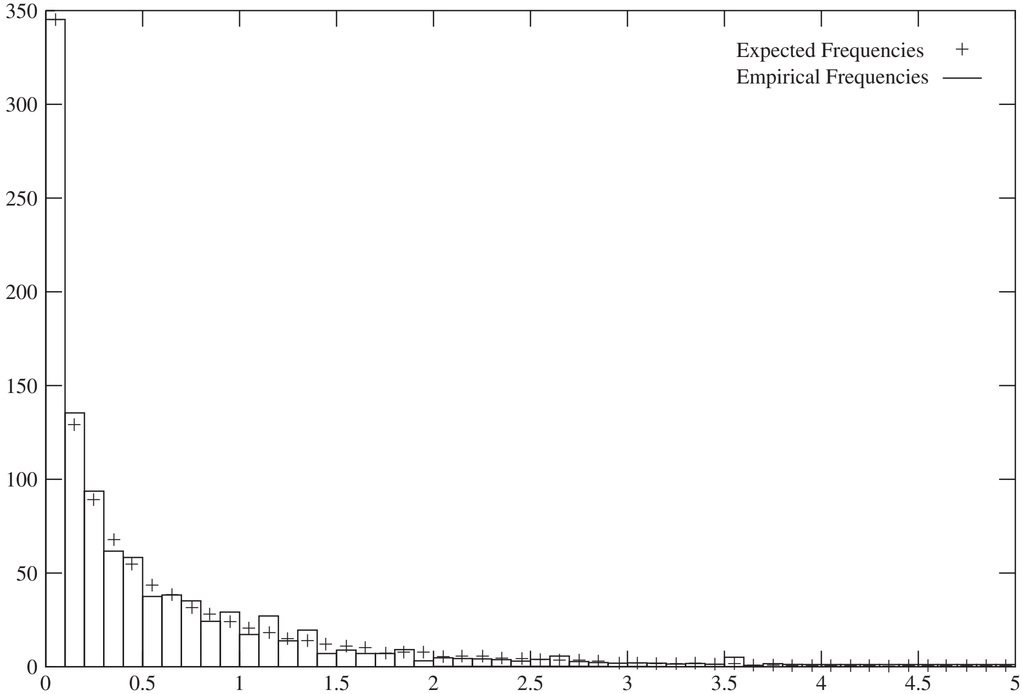


FIGURE 4.23
1000 samples of gamma random variable using rejection method.

generate the most common random variables. You simply need to invoke the appropriate functions.

Example 4.56 Generating Gamma Random Variables

Use Octave to obtain eight Gamma random variables with $\alpha = 0.25$ and $\lambda = 1$.

The Octave command and the corresponding answer are given below:

```
> gamma_rnd (0.25, 1, 1, 8)
ans =
 Columns 1 through 6:
  0.00021529   0.09331491   0.24606757   0.08665787
  0.00013400   0.23384718
 Columns 7 and 8:
  1.72940941   1.29599702
```

4.9.3 Generation of Functions of a Random Variable

Once we have a simple method of generating a random variable X, we can easily generate any random variable that is defined by $Y = g(X)$ or even $Z = h(X_1, X_2, \ldots, X_n)$, where $X_1, \ldots, X_n$ are n outputs of the random variable generator.

Example 4.57 ***m*-Erlang Random Variable**

Let $X_1, X_2, \ldots$ be independent, exponentially distributed random variables with parameter λ. In Chapter 7 we show that the random variable

$$Y = X_1 + X_2 + \cdots + X_m$$

has an m-Erlang pdf with parameter λ. We can therefore generate an m-Erlang random variable by first generating m exponentially distributed random variables using the transformation method, and then taking the sum. Since the m-Erlang random variable is a special case of the gamma random variable, for large m it may be preferable to use the rejection method described in Problem 4.142.

4.9.4 Generating Mixtures of Random Variables

We have seen in previous sections that sometimes a random variable consists of a mixture of several random variables. In other words, the generation of the random variable can be viewed as first selecting a random variable type according to some pmf, and then generating a random variable from the selected pdf type. This procedure can be simulated easily.

Example 4.58 **Hyperexponential Random Variable**

A two-stage hyperexponential random variable has pdf

$$f_X(x) = pae^{-ax} + (1 - p)be^{-bx}.$$

It is clear from the above expression that X consists of a mixture of two exponential random variables with parameters a and b, respectively. X can be generated by first performing a Bernoulli trial with probability of success p. If the outcome is a success, we then use the transformation method to generate an exponential random variable with parameter a. If the outcome is a failure, we generate an exponential random variable with parameter b instead.

*4.10 ENTROPY

Entropy is a measure of the uncertainty in a random experiment. In this section, we first introduce the notion of the entropy of a random variable and develop several of its fundamental properties. We then show that entropy quantifies uncertainty by the amount of information required to specify the outcome of a random experiment. Finally, we discuss the method of maximum entropy, which has found wide use in characterizing random variables when only some parameters, such as the mean or variance, are known.

4.10.1 The Entropy of a Random Variable

Let X be a discrete random variable with $S_X = \{1, 2, \ldots, K\}$ and pmf $p_k = P[X = k]$. We are interested in quantifying the uncertainty of the event $A_k = \{X = k\}$. Clearly, the uncertainty of A_k is low if the probability of A_k is close to one, and it is high if the probability of A_k is small. The following measure of uncertainty satisfies these two properties:

Incerteza ←
$$I(X = k) = \ln \frac{1}{P[X = k]} = -\ln P[X = k]. \tag{4.105}$$

Note from Fig. 4.24 that $I(X = k) = 0$ if $P[X = k] = 1$, and $I(X = k)$ increases with decreasing $P[X = k]$. The **entropy of a random variable** $\boldsymbol{X}$ is defined as the expected value of the uncertainty of its outcomes:

Valor esperado de la incerteza ←
$$H_X = E[I(X)] = \sum_{k=1}^{K} P[X = k] \ln \frac{1}{P[X = k]}$$
$$= -\sum_{k=1}^{K} P[X = k] \ln P[X = k]. \tag{4.106}$$

Note that in the above definition we have used $I(X)$ as a function of a random variable. We say that entropy is in units of "bits" when the logarithm is base 2. In the above expression we are using the natural logarithm, so we say the units are in "nats." Changing the base of the logarithm is equivalent to multiplying entropy by a constant, since $\ln(x) = \ln 2 \log_2 x$.

Example 4.59 Entropy of a Binary Random Variable

Suppose that $S_X = \{0, 1\}$ and $p = P[X = 0] = 1 - P[X = 1]$. Figure 4.25 shows $-p\ln(p)$, $-(1 - p)\ln(1 - p)$, and the entropy of the binary random variable $H_X = h(p) = -p \ln(p) - (1 - p)\ln(1 - p)$ as functions of p. Note that $h(p)$ is symmetric about $p = 1/2$ and that it achieves its maximum at $p = 1/2$. Note also how the uncertainties of the events $\{X = 0\}$ and $\{X = 1\}$ vary together in complementary fashion: When $P[X = 0]$ is very small (i.e., highly uncertain), then $P[X = 1]$ is close to one (i.e., highly certain), and vice versa. Thus the highest average uncertainty occurs when $P[X = 0] = P[X = 1] = 1/2$.

H_X can be viewed as the average uncertainty that is resolved by observing X. This suggests that if we are designing a binary experiment (for example, a yes/no question), then the average uncertainty that is resolved will be maximized if the two outcomes are designed to be equiprobable.

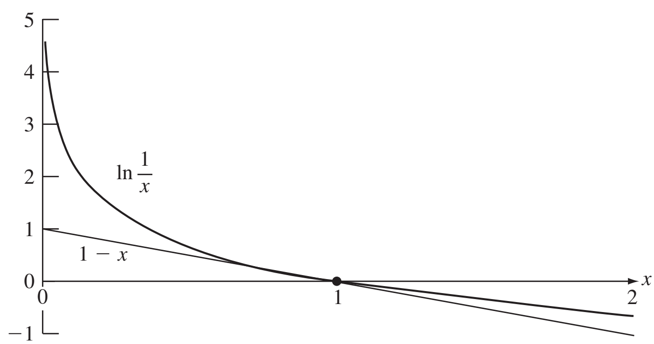
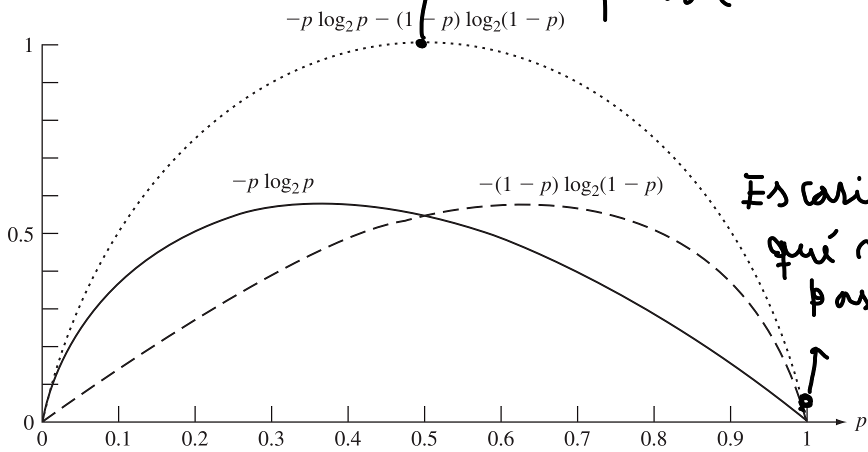


FIGURE 4.24
$\ln(1/x) \geq 1 - x$

Muy difícil saber qué va a pasar



Es casi obvio qué va a pasar

FIGURE 4.25
Entropy of binary random variable.

Example 4.60 Reduction of Entropy Through Partial Information

The binary representation of the random variable X takes on values from the set $\{000, 001, 010, \ldots, 111\}$ with equal probabilities. Find the reduction in the entropy of X given the event $A = \{X \text{ begins with a } 1\}$.

The entropy of X is

$$H_X = -\frac{1}{8}\log_2\frac{1}{8} - \frac{1}{8}\log_2\frac{1}{8} - \cdots - \frac{1}{8}\log_2\frac{1}{8} = 3 \text{ bits.}$$

Tiene todo el sentido del mundo

The event A implies that X is in the set $\{100, 101, 110, 111\}$, so the entropy of X given A is

$$H_{X|A} = -\frac{1}{4}\log_2\frac{1}{4} - \cdots - \frac{1}{4}\log_2\frac{1}{4} = 2 \text{ bits.}$$

Thus the reduction in entropy is $H_X - H_{X|A} = 3 - 2 = 1$ bit.

Let $\boldsymbol{p} = (p_1, p_2, \ldots, p_K)$, and $\boldsymbol{q} = (q_1, q_2, \ldots, q_K)$ be two pmf's. The **relative entropy** of $\boldsymbol{q}$ with respect to $\boldsymbol{p}$ is defined by

¿?

$+ \sum p_x \ln(p_x)$

$$H(p; q) = \sum_{k=1}^{K} p_k \ln\frac{1}{q_k} \cancel{-} + H_X = \sum_{k=1}^{K} p_k \ln\frac{p_k}{q_k}. \qquad (4.107)$$

The relative entropy is nonnegative, and equal to zero if and only if $p_k = q_k$ for all k:

$$H(p; q) \geq 0 \qquad \text{with equality iff} \qquad p_k = q_k \qquad \text{for } k = 1, \ldots, K. \qquad (4.108)$$

We will use this fact repeatedly in the remainder of this section.

¿? Para qué sirve?

OBS: Si $\vec{p} = \vec{q} \Rightarrow H(p,p) = 0$

To show that the relative entropy is nonnegative, we use the inequality $\ln(1/x) \geq 1 - x$ with equality iff $x = 1$, as shown in Fig. 4.24. Equation (4.107) then becomes

$$H(p; q) = \sum_{k=1}^{K} p_k \ln \frac{p_k}{q_k} \geq \sum_{k=1}^{K} p_k \left(1 - \frac{q_k}{p_k}\right) = \sum_{k=1}^{K} p_k - \sum_{k=1}^{K} q_k = 0. \tag{4.109}$$

In order for equality to hold in the above expression, we must have $p_k = q_k$ for $k = 1, \ldots, K$.

Let X be any random variable with $S_X = \{1, 2, \ldots, K\}$ and pmf $\boldsymbol{p}$. If we let $q_k = 1/K$ in Eq. (4.108), then

$$H(p; q) = \ln K - H_X = \sum_{k=1}^{K} p_k \ln \frac{p_k}{1/K} \geq 0,$$

which implies that for any random variable X with $S_X = \{1, 2, \ldots, K\}$,

$$H_X \leq \ln K \qquad \text{with equality iff} \qquad p_k = \frac{1}{K} \qquad k = 1, \ldots, K. \tag{4.110}$$

Thus *the maximum entropy attainable by the random variable X is* ln K, *and this maximum is attained when all the outcomes are equiprobable.*

Equation (4.110) shows that the entropy of random variables with finite S_X is always finite. On the other hand, it also shows that as the size of S_X is increased, the entropy can increase without bound. The following example shows that some countably infinite random variables have finite entropy.

Example 4.61 Entropy of a Geometric Random Variable

The entropy of the geometric random variable with $S_X = \{0, 1, 2, \ldots\}$ is:

$$\begin{aligned}
H_X &= -\sum_{k=0}^{\infty} p(1-p)^k \ln(p(1-p)^k) \\
&= -\ln p - \ln(1-p) \sum_{k=0}^{\infty} k p (1-p)^k \\
&= -\ln p - \frac{(1-p)\ln(1-p)}{p} \\
&= \frac{-p \ln p - (1-p)\ln(1-p)}{p} = \frac{h(p)}{p},
\end{aligned} \tag{4.111}$$

where $h(p)$ is the entropy of a binary random variable. Note that $H_X = 2$ bits when $p = 1/2$.

For continuous random variables we have that $P[X = x] = 0$ for all x. Therefore by Eq. (4.105) the uncertainty for every event $\{X = x\}$ is infinite, and it follows from

Eq. (4.106) that the *entropy of continuous random variables is infinite.* The next example takes a look at how the notion of entropy may be applied to continuous random variables.

Example 4.62 Entropy of a Quantized Continuous Random Variable

Let X be a continuous random variable that takes on values in the interval $[a, b]$. Suppose that the interval $[a, b]$ is divided into a large number K of subintervals of length Δ. Let $Q(X)$ be the midpoint of the subinterval that contains X. Find the entropy of Q.

Let x_k be the midpoint of the kth subinterval, then $P[Q = x_k] = P[X \text{ is in } k\text{th subinterval}] = P[x_k - \Delta/2 < X < x_k + \Delta/2] \simeq f_X(x_k)\Delta$, and thus

$$\begin{aligned} H_Q &= \sum_{k=1}^{K} P[Q = x_k] \ln P[Q = x_k] \\ &\simeq -\sum_{k=1}^{K} f_X(x_k)\Delta \ln(f_X(x_k)\Delta) \\ &= \overbrace{-\ln(\Delta)}^{\to -\infty} - \overbrace{\sum_{k=1}^{K} f_X(x_k) \ln(f_X(x_k))\Delta}^{\to \int}. \end{aligned} \tag{4.112}$$

The above equation shows that there is a tradeoff between the entropy of Q and the quantization error $X - Q(X)$. As Δ is decreased the error decreases, but the entropy increases without bound, once again confirming the fact that the entropy of continuous random variables is infinite.

In the final expression for H_X in Eq. (4.112), as Δ approaches zero, the first expression approaches infinity, but the second expression approaches an integral which may be finite in some cases. The **differential entropy** is defined by this integral:

¿Y para $\mathbb{R}^n$?

$$H_X = -\int_{-\infty}^{\infty} f_X(x) \ln f_X(x)\, dx = -E[\ln f_X(X)]. \tag{4.113}$$

In the above expression, we reuse the term H_X with the understanding that we deal with differential entropy when dealing with continuous random variables.

Example 4.63 Differential Entropy of a Uniform Random Variable

The differential entropy for X uniform in $[a, b]$ is

$$H_X = -E\left[\ln\left(\frac{1}{b-a}\right)\right] = \ln(b - a). \tag{4.114}$$

Example 4.64 Differential Entropy of a Gaussian Random Variable

The differential entropy for X, a Gaussian random variable (see Eq. 4.47), is

$$
\begin{aligned}
H_X &= -E[\ln f_X(X)] \\
&= -E\left[\ln\frac{1}{\sqrt{2\pi\sigma^2}} - \frac{(X-m)^2}{2\sigma^2}\right] = -\ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) + \frac{1}{2\sigma^2}\overbrace{E\left[(X-m)^2\right]}^{\sigma^2} \\
&= \frac{1}{2}\ln(2\pi\sigma^2) + \frac{1}{2} \\
&= \frac{1}{2}\ln(2\pi e\sigma^2). \qquad (4.115)
\end{aligned}
$$

$H_X \propto \sigma_X$ → ¿y para qué sirve la entropía?

Si sigma crece, crece la entropía, es difícil saber con precisión qué va a pasar xq la gaussiana se dispersa

The entropy function and the differential entropy function differ in several fundamental ways. In the next section we will see that the entropy of a random variable has a very well defined operational interpretation as the average number of information bits required to specify the value of the random variable. Differential entropy does not possess this operational interpretation. In addition, the entropy function does not change when the random variable X is mapped into Y by an invertible transformation. Again, the differential entropy does not possess this property. (See Problems 4.153 and 4.160.) Nevertheless, the differential entropy does possess some useful properties. The differential entropy appears naturally in problems involving entropy reduction, as demonstrated in Problem 4.159. In addition, the relative entropy of continuous random variables, which is defined by

¿?

$$H(f_X; f_Y) = \int_{-\infty}^{\infty} f_X(x)\ln\frac{f_X(x)}{f_Y(x)}\,dx,$$

does not change under invertible transformations.

4.10.2 Entropy as a Measure of Information

Let X be a discrete random variable with $S_X = \{1, 2, \ldots, K\}$ and pmf $p_k = P[X = k]$. Suppose that the experiment that produces X is performed by John, and that he attempts to communicate the outcome to Mary by answering a series of yes/no questions. We are interested in characterizing the minimum average number of questions required to identify X.

Example 4.65

An urn contains 16 balls: 4 balls are labeled "1", 4 are labeled "2", 2 are labeled "3", 2 are labeled "4", and the remaining balls are labeled "5", "6", "7", and "8." John picks a ball from the urn at random, and he notes the number. Discuss what strategies Mary can use to find out the number

1 → x4 3 → x2 5, 6, 7, 8 → x1
2 → x4 4 → x2

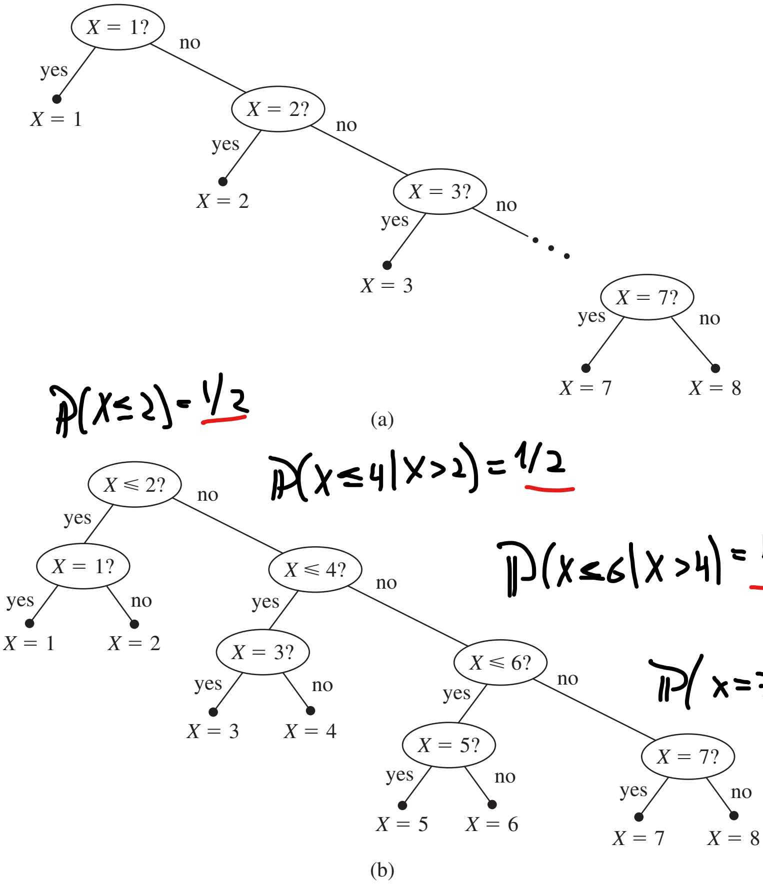
of the ball through a series of yes/no questions. Compare the average number of questions asked to the entropy of X.

If we let X be the random variable denoting the number of the ball, then $S_X = \{1, 2, \ldots, 8\}$ and the pmf is $\boldsymbol{p} = (1/4, 1/4, 1/8, 1/8, 1/16, 1/16, 1/16, 1/16)$. We will compare the two strategies shown in Figs. 4.26(a) and (b).

The series of questions in Fig. 4.26(a) uses the fact that the probability of $\{X = k\}$ decreases with k. Thus it is reasonable to ask the question {"Was X equal to 1?"}, {"Was X equal to 2?"}, and so on, until the answer is yes. Let L be the number of questions asked until the answer is yes, then the average number of questions asked is

$$E[L] = 1\left(\tfrac{1}{4}\right) + 2\left(\tfrac{1}{4}\right) + 3\left(\tfrac{1}{8}\right) + 4\left(\tfrac{1}{8}\right) + 5\left(\tfrac{1}{16}\right) + 6\left(\tfrac{1}{16}\right) + 7\left(\tfrac{1}{16}\right) + 7\left(\tfrac{1}{16}\right)$$
$$= 51/16.$$

The series of questions in Fig. 4.26(b) uses the observation made in Example 4.57 that yes/no questions should be designed so that the two answers are equiprobable. The questions in



$\mathbb{P}(X \leq 2) = 1/2$

$\mathbb{P}(X \leq 4 | X > 2) = 1/2$

$\mathbb{P}(X \leq 6 | X > 4) = 1/2$

$\mathbb{P}(X = 7 | X > 6) = 1/2$

FIGURE 4.26
Two strategies for finding out the value of X through a series of yes/no questions.

Fig. 4.26(b) meet this requirement. The average number of questions asked is

$$E[L] = 2\left(\tfrac{1}{4}\right) + 2\left(\tfrac{1}{4}\right) + 3\left(\tfrac{1}{8}\right) + 3\left(\tfrac{1}{8}\right) + 4\left(\tfrac{1}{16}\right) + 4\left(\tfrac{1}{16}\right) + 4\left(\tfrac{1}{16}\right) + 4\left(\tfrac{1}{16}\right)$$
$$= 44/16.$$

Thus the second series of questions has the better performance.

Finally, we find that the entropy of X is

$$H_X = -\frac{1}{4}\log_2\frac{1}{4} - \frac{1}{4}\log_2\frac{1}{4} - \frac{1}{8}\log_2\frac{1}{8} - \cdots - \frac{1}{16}\log_2\frac{1}{16} = 44/16,$$

which is equal to the performance of the second series of questions.

The problem of designing the series of questions to identify the random variable X is exactly the same as the problem of encoding the output of an information source. Each output of an information source is a random variable X, and the task of the encoder is to map each possible output into a unique string of binary digits. We can see this correspondence by taking the trees in Fig. 4.26 and identifying each yes/no answer with a 0/1. The sequence of 0's and 1's from the top node to each terminal node then defines the binary string ("codeword") for each outcome. It then follows that the problem of finding the best series of yes/no questions is the same as finding the binary tree code that minimizes the average codeword length. → DOU

In the remainder of this section we develop the following fundamental results from information theory. First, the average codeword length of *any* code cannot be less than the entropy. Second, if the pmf of X consists of powers of 1/2, then there is a tree code that achieves the entropy. And finally, by encoding groups of outcomes of X we can achieve average codeword length arbitrarily close to the entropy. Thus *the entropy of X represents the minimum average number of bits required to establish the outcome of X*. Pero ¿qué significa promedio mínimo?

First, let's show that the average codeword length of *any* tree code cannot be less than the entropy. Note from Fig. 4.26 that the set of lengths $\{l_k\}$ of the codewords for every complete binary tree must satisfy

$$\sum_{k=1}^{K} 2^{-l_k} = 1. \tag{4.116}$$

To see this, extend the tree to the same depth as the longest codeword, as shown in Fig. 4.27. If we then "prune" the tree at a node of depth l_k, we remove a fraction 2^{-l_k} of the nodes at the bottom of the tree. Note that the converse result is also true: If a set of codeword lengths satisfies Eq. (4.116), then we can construct a tree code with these lengths.

Consider next the difference between the entropy and $E[L]$ for any binary tree code:

$$\begin{aligned} E[L] - H_X &= \sum_{k=1}^{K} l_k P[X = k] + \sum_{k=1}^{K} P[X = k] \log_2 P[X = k] \\ &= \sum_{k=1}^{K} P[X = k] \log_2 \frac{P[X = k]}{2^{-l_k}}, \end{aligned} \tag{4.117}$$

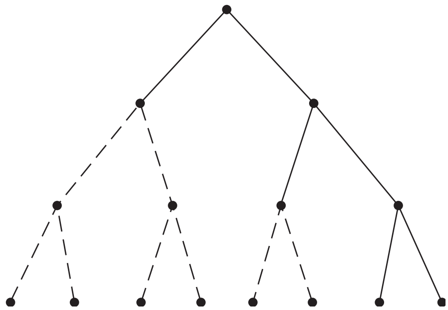


FIGURE 4.27
Extension of a binary tree code to a full tree.

where we have expressed the entropy in bits. Equation (4.17) is the relative entropy of Eq. (4.107) with $q_k = 2^{-l_k}$. Thus by Eq. (4.108)

$$E[L] \geq H_X \qquad \text{with equality iff} \qquad P[X = k] = 2^{-l_k}. \tag{4.118}$$

Thus the average number of questions for *any* tree code (and in particular the *best* tree code) cannot be less than the entropy of X. Therefore we can use the entropy H_X as a baseline against which to test any code.

Equation (4.118) also implies that if the outcomes of X all have probabilities that are integer powers of 1/2 (as in Example 4.63), then we can find a tree code that achieves the entropy. If $P[X = k] = 2^{-l_k}$, then we assign the outcome k a binary codeword of length l_k. We can show that we can always find a tree code with these lengths by using the fact that the probabilities add to one, and hence the codeword lengths satisfy Eq. (4.116). Equation (4.118) then implies that $E[L] = H$.

It is clear that Eq. (4.117) will be nonzero if the p_k's are not integer powers of 1/2. Thus in general the best tree code does not always have $E[L] = H_X$. However, it is possible to show that the approach of grouping outcomes into sets that are approximately equiprobable leads to tree codes with lengths that are close to the entropy. Furthermore, by encoding vectors of outcomes of X, it is possible to obtain average codeword lengths that are arbitrarily close to the entropy. Problem 4.165 discusses how this is done.

We have now reached our objective of showing that the entropy of a random variable X represents the minimum average number of bits required to identify its value. Before proceeding, let's reconsider continuous random variables. A continuous random variable can assume values from an uncountably infinite set, so in general an infinite number of bits is required to specify its value. Thus, the interpretation of entropy as the average number of bits required to specify a random variable immediately implies that continuous random variables have infinite entropy. This implies that any representation of a continuous random variable that uses a finite number of bits will inherently involve some approximation error.

4.10.3 The Method of Maximum Entropy

Let X be a random variable with $S_X = \{x_1, x_2, \ldots, x_K\}$ and unknown pmf $p_k = P[X = x_k]$. Suppose that we are asked to estimate the pmf of X given the expected value of some function $g(X)$ of X:

$$\sum_{k=1}^{K} g(x_k)P[X = x_k] = c. \tag{4.119}$$

For example, if $g(X) = X$ then $c = E[g(X)] = E[X]$, and if $g(X) = (X - E[X])^2$ then $c = \text{VAR}[X]$. Clearly, this problem is underdetermined since knowledge of these parameters is not sufficient to specify the pmf uniquely. The **method of maximum entropy** approaches this problem by seeking the pmf that maximizes the entropy subject to the constraint in Eq. (4.119).

Suppose we set up this maximization problem by using Lagrange multipliers:

$$H_X + \lambda\left(\sum_{k=1}^{K} P[X = x_k]g(x_k) - c\right) = -\sum_{k=1}^{K} P[X = x_k] \ln \frac{P[X = x_k]}{Ce^{-\lambda g(x_k)}}, \tag{4.120}$$

where $C = e^c$. Note that if $\{Ce^{-\lambda g(x_k)}\}$ forms a pmf, then the above expression is the negative value of the relative entropy of this pmf with respect to $\boldsymbol{p}$. Equation (4.108) then implies that the expression in Eq. (4.120) is always less than or equal to zero with equality iff $P[X = x_k] = Ce^{-\lambda g(x_k)}$. We now show that this does indeed lead to the maximum entropy solution.

Suppose that the random variable X has pmf $p_k = Ce^{-\lambda g(x_k)}$, where C and λ are chosen so that Eq. (4.119) is satisfied and so that $\{p_k\}$ is a pmf. X then has entropy

$$\begin{aligned} H_X &= E[-\ln P[X]] = [-\ln Ce^{-\lambda g(x_k)}] = -\ln C + \lambda E[g(X)] \\ &= -\ln C + \lambda c. \end{aligned} \tag{4.121}$$

Now let's compare the entropy in Eq. (4.121) to that of some other pmf q_k that also satisfies the constraint in Eq. (4.119). Consider the relative entropy of $\boldsymbol{p}$ with respect to $\boldsymbol{q}$:

$$\begin{aligned} 0 \le H(\boldsymbol{q}; \boldsymbol{p}) &= \sum_{k=1}^{K} q_k \ln \frac{q_k}{p_k} = \sum_{k=1}^{K} q_k \ln q_k + \sum_{k=1}^{K} q_k(-\ln C + \lambda g(x_k)) \\ &= -\ln C + \lambda c - H(q) = H_X - H(q). \end{aligned} \tag{4.122}$$

Thus $H_X \ge H(\boldsymbol{q})$, and $\boldsymbol{p}$ achieves the highest entropy.

Example 4.66

Let X be a random variable with $S_X = \{0, 1, \ldots\}$ and expected value $E[X] = m$. Find the pmf of X that maximizes the entropy.

In this example $g(X) = X$, so

$$p_k = Ce^{-\lambda k} = C\alpha^k,$$

where $\alpha = e^{-\lambda}$. Clearly, X is a geometric random variable with mean $m = \alpha/(1 - \alpha)$ and thus $\alpha = m/(m + 1)$. It then follows that $C = 1 - \alpha = 1/(m + 1)$.

When dealing with continuous random variables, the method of maximum entropy maximizes the differential entropy:

$$-\int_{-\infty}^{\infty} f_X(x) \ln f_X(x)\, dx. \tag{4.123}$$

The parameter information is in the form

$$c = E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx. \tag{4.124}$$

The relative entropy expression in Eq. (4.115) and the approach used for discrete random variables can be used to show that the pdf $f_X(x)$ that maximizes the differential entropy will have the form

$$f_X(x) = Ce^{-\lambda g(x)}, \tag{4.125}$$

where C and λ must be chosen so that Eq. (4.125) integrates to one and so that Eq. (4.124) is satisfied.

Example 4.67

Suppose that the continuous random variable X has known variance $\sigma^2 = E[(X - m)^2]$, where the mean m is not specified. Find the pdf that maximizes the entropy of X.

Equation (4.125) implies that the pdf has the form

$$f_X(x) = Ce^{-\lambda(x-m)^2}.$$

We can meet the constraint in Eq. (4.124) by picking

$$\lambda = \frac{1}{2\sigma^2} \qquad C = \frac{1}{\sqrt{2\pi\sigma^2}}.$$

We thus obtain a Gaussian pdf with variance σ^2. Note that the mean m is arbitrary; that is, any choice of m yields a pdf that maximizes the differential entropy.

The method of maximum entropy can be extended to the case where several parameters of the random variable X are known. It can also be extended to the case of vectors and sequences of random variables.

SUMMARY

- The cumulative distribution function $F_X(x)$ is the probability that X falls in the interval $(-\infty, x]$. The probability of any event consisting of the union of intervals can be expressed in terms of the cdf.
- A random variable is continuous if its cdf can be written as the integral of a non-negative function. A random variable is mixed if it is a mixture of a discrete and a continuous random variable.
- The probability of events involving a continuous random variable X can be expressed as integrals of the probability density function $f_X(x)$.
- If X is a random variable, then $Y = g(X)$ is also a random variable. The notion of equivalent events allows us to derive expressions for the cdf and pdf of Y in terms of the cdf and pdf of X.
- The cdf and pdf of the random variable X are sufficient to compute all probabilities involving X alone. The mean, variance, and moments of a random variable summarize some of the information about the random variable X. These parameters are useful in practice because they are easier to measure and estimate than the cdf and pdf.
- Conditional cdf's or pdf's incorporate partial knowledge about the outcome of an experiment in the calculation of probabilities of events.
- The Markov and Chebyshev inequalities allow us to bound probabilities involving X in terms of its first two moments only.
- Transforms provide an alternative but equivalent representation of the pmf and pdf. In certain types of problems it is preferable to work with the transforms rather than the pmf or pdf. The moments of a random variable can be obtained from the corresponding transform.
- The reliability of a system is the probability that it is still functioning after t hours of operation. The reliability of a system can be determined from the reliability of its subsystems.
- There are a number of methods for generating random variables with prescribed pmf's or pdf's in terms of a random variable that is uniformly distributed in the unit interval. These methods include the transformation and the rejection methods as well as methods that simulate random experiments (e.g., functions of random variables) and mixtures of random variables.
- The entropy of a random variable X is a measure of the uncertainty of X in terms of the average amount of information required to identify its value.
- The maximum entropy method is a procedure for estimating the pmf or pdf of a random variable when only partial information about X, in the form of expected values of functions of X, is available.

CHECKLIST OF IMPORTANT TERMS

Characteristic function
Chebyshev inequality
Chernoff bound
Conditional cdf, pdf
Continuous random variable
Cumulative distribution function
Differential entropy
Discrete random variable
Entropy
Equivalent event
Expected value of X
Failure rate function
Function of a random variable
Laplace transform of the pdf
Markov inequality
Maximum entropy method
Mean time to failure (MTTF)
Moment theorem
nth moment of X
Probability density function
Probability generating function
Probability mass function
Random variable
Random variable of mixed type
Rejection method
Reliability
Standard deviation of X
Transformation method
Variance of X

ANNOTATED REFERENCES

Reference [1] is the standard reference for electrical engineers for the material on random variables. Reference [2] is entirely devoted to continuous distributions. Reference [3] discusses some of the finer points regarding the concept of a random variable at a level accessible to students of this course. Reference [4] presents detailed discussions of the various methods for generating random numbers with specified distributions. Reference [5] also discusses the generation of random variables. Reference [9] is focused on signal processing. Reference [11] discusses entropy in the context of information theory.

1. A. Papoulis and S. Pillai, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, 2002.
2. N. Johnson et al., *Continuous Univariate Distributions,* vol. 2, Wiley, New York, 1995.
3. K. L. Chung, *Elementary Probability Theory*, Springer-Verlag, New York, 1974.
4. A. M. Law and W. D. Kelton, *Simulation Modeling and Analysis*, McGraw-Hill, New York, 2000.
5. S. M. Ross, *Introduction to Probability Models*, Academic Press, New York, 2003.
6. H. Cramer, *Mathematical Methods of Statistics*, Princeton University Press, Princeton, N.J., 1946.
7. M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions*, National Bureau of Standards, Washington, D.C., 1964. Downloadable: *www.math.sfu.ca/~cbm/aands/*.
8. R. C. Cheng, "The Generation of Gamma Variables with Nonintegral Shape Parameter," *Appl. Statist.*, 26: 71–75, 1977.
9. R. Gray and L.D. Davisson, *An Introduction to Statistical Signal Processing,* Cambridge Univ. Press, Cambridge, UK, 2005.

10. P. O. Börjesson and C. E. W. Sundberg, "Simple Approximations of the Error Function $Q(x)$ for Communications Applications," *IEEE Trans. on Communications*, March 1979, 639–643.

11. R. G. Gallager, *Information Theory and Reliable Communication*, Wiley, New York, 1968.

PROBLEMS

Section 4.1: The Cumulative Distribution Function

4.1. An information source produces binary pairs that we designate as $S_X = \{1, 2, 3, 4\}$ with the following pmf's:

(i) $p_k = p_1/k$ for all k in S_X.

(ii) $p_{k+1} = p_k/2$ for $k = 2, 3, 4$.

(iii) $p_{k+1} = p_k/2^k$ for $k = 2, 3, 4$.

(a) Plot the cdf of these three random variables.

(b) Use the cdf to find the probability of the events: $\{X \le 1\}, \{X < 2.5\}, \{0.5 < X \le 2\}, \{1 < X < 4\}$.

4.2. A die is tossed. Let X be the number of full pairs of dots in the face showing up, and Y be the number of full or partial pairs of dots in the face showing up. Find and plot the cdf of X and Y.

4.3. The loose minute hand of a clock is spun hard. The coordinates (x, y) of the point where the tip of the hand comes to rest is noted. Z is defined as the sgn function of the product of x and y, where $\text{sgn}(t)$ is 1 if $t > 0$, 0 if $t = 0$, and -1 if $t < 0$.

(a) Find and plot the cdf of the random variable X.

(b) Does the cdf change if the clock hand has a propensity to stop at 3, 6, 9, and 12 o'clock?

4.4. An urn contains 8 \$1 bills and two \$5 bills. Let X be the total amount that results when two bills are drawn from the urn without replacement, and let Y be the total amount that results when two bills are drawn from the urn *with* replacement.

(a) Plot and compare the cdf's of the random variables.

(b) Use the cdf to compare the probabilities of the following events in the two problems: $\{X = \$2\}, \{X < \$7\}, \{X \ge 6\}$.

4.5. Let Y be the difference between the number of heads and the number of tails in the 3 tosses of a fair coin.

(a) Plot the cdf of the random variable Y.

(b) Express $P[|Y| < y]$ in terms of the cdf of Y.

4.6. A dart is equally likely to land at any point inside a circular target of radius 2. Let R be the distance of the landing point from the origin.

(a) Find the sample space S and the sample space of R, S_R.

(b) Show the mapping from S to S_R.

(c) The "bull's eye" is the central disk in the target of radius 0.25. Find the event A in S_R corresponding to "dart hits the bull's eye." Find the equivalent event in S and $P[A]$.

(d) Find and plot the cdf of R.

4.7. A point is selected at random inside a square defined by $\{(x, y): 0 \le x \le b, 0 \le y \le b\}$. Assume the point is equally likely to fall anywhere in the square. Let the random variable Z be given by the minimum of the two coordinates of the point where the dart lands.

(a) Find the sample space S and the sample space of Z, S_Z.

(b) Show the mapping from S to S_Z.

(c) Find the region in the square corresponding to the event $\{Z \leq z\}$.

(d) Find and plot the cdf of Z.

(e) Use the cdf to find: $P[Z > 0]$, $P[Z > b]$, $P[Z \leq b/2]$, $P[Z > b/4]$.

4.8. Let ζ be a point selected at random from the unit interval. Consider the random variable $X = (1 - \zeta)^{-1/2}$.

(a) Sketch X as a function of ζ.

(b) Find and plot the cdf of X.

(c) Find the probability of the events $\{X > 1\}$, $\{5 < X < 7\}$, $\{X \leq 20\}$.

4.9. The loose hand of a clock is spun hard and the outcome ζ is the angle in the range $[0, 2\pi)$ where the hand comes to rest. Consider the random variable $X(\zeta) = 2\sin(\zeta/4)$.

(a) Sketch X as a function of ζ.

(b) Find and plot the cdf of X.

(c) Find the probability of the events $\{X > 1\}$, $\{-1/2 < X < 1/2\}$, $\{X \leq 1/\sqrt{2}\}$.

4.10. Repeat Problem 4.9 if 80% of the time the hand comes to rest anywhere in the circle, but 20% of the time the hand comes to rest at 3, 6, 9, or 12 o'clock.

4.11. The random variable X is uniformly distributed in the interval $[-1, 2]$.

(a) Find and plot the cdf of X.

(b) Use the cdf to find the probabilities of the following events: $\{X \leq 0\}$, $\{|X - 0.5| < 1\}$, and $C = \{X > -0.5\}$.

4.12. The cdf of the random variable X is given by:

$$F_X(x) = \begin{cases} 0 & x < -1 \\ 0.5 & -1 \leq x \leq 0 \\ (1 + x)/2 & 0 \leq x \leq 1 \\ 1 & x \geq 1. \end{cases}$$

(a) Plot the cdf and identify the type of random variable.

(b) Find $P[X \leq -1]$, $P[X = -1]$, $P[X < 0.5]$, $P[-0.5 < X < 0.5]$, $P[X > -1]$, $P[X \leq 2]$, $P[X > 3]$.

4.13. A random variable X has cdf:

$$F_X(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 - \dfrac{1}{4}e^{-2x} & \text{for } x \geq 0. \end{cases}$$

(a) Plot the cdf and identify the type of random variable.

(b) Find $P[X \leq 2]$, $P[X = 0]$, $P[X < 0]$, $P[2 < X < 6]$, $P[X > 10]$.

4.14. The random variable X has cdf shown in Fig. P4.1.

(a) What type of random variable is X?

(b) Find the following probabilities: $P[X < -1]$, $P[X \leq -1]$, $P[-1 < X < -0.75]$, $P[-0.5 \leq X < 0]$, $P[-0.5 \leq X \leq 0.5]$, $P[|X - 0.5| < 0.5]$.

4.15. For $\beta > 0$ and $\lambda > 0$, the Weibull random variable Y has cdf:

$$F_X(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 - e^{-(x/\lambda)^{\beta}} & \text{for } x \geq 0. \end{cases}$$

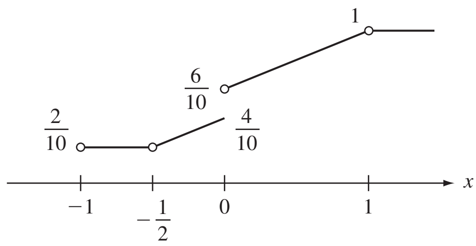


FIGURE P4.1

(a) Plot the cdf of Y for $\beta = 0.5, 1,$ and 2.

(b) Find the probability $P[j\lambda < X < (j+1)\lambda]$ and $P[X > j\lambda]$.

(c) Plot $\log P[X > x]$ vs. $\log x$.

4.16. The random variable X has cdf:

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 0.5 + c\sin^2(\pi x/2) & 0 \le x \le 1 \\ 1 & x > 1. \end{cases}$$

(a) What values can c assume?

(b) Plot the cdf.

(c) Find $P[X > 0]$.

Section 4.2: The Probability Density Function

4.17. A random variable X has pdf:

$$f_X(x) = \begin{cases} c(1 - x^2) & -1 \le x \le 1 \\ 0 & \text{elsewhere.} \end{cases}$$

(a) Find c and plot the pdf.

(b) Plot the cdf of X.

(c) Find $P[X = 0]$, $P[0 < X < 0.5]$, and $P[|X - 0.5| < 0.25]$.

4.18. A random variable X has pdf:

$$f_X(x) = \begin{cases} cx(1 - x^2) & 0 \le x \le 1 \\ 0 & \text{elsewhere.} \end{cases}$$

(a) Find c and plot the pdf.

(b) Plot the cdf of X.

(c) Find $P[0 < X < 0.5]$, $P[X = 1]$, $P[.25 < X < 0.5]$.

4.19. **(a)** In Problem 4.6, find and plot the pdf of the random variable R, the distance from the dart to the center of the target.

(b) Use the pdf to find the probability that the dart is outside the bull's eye.

4.20. **(a)** Find and plot the pdf of the random variable Z in Problem 4.7.

(b) Use the pdf to find the probability that the minimum is greater than $b/3$.

4.21. **(a)** Find and plot the pdf in Problem 4.8.

(b) Use the pdf to find the probabilities of the events: $\{X > a\}$ and $\{X > 2a\}$.

4.22. **(a)** Find and plot the pdf in Problem 4.12.

(b) Use the pdf to find $P[-1 \le X < 0.25]$.

4.23. **(a)** Find and plot the pdf in Problem 4.13.

(b) Use the pdf to find $P[X = 0]$, $P[X > 8]$.

4.24. **(a)** Find and plot the pdf of the random variable in Problem 4.14.

(b) Use the pdf to calculate the probabilities in Problem 4.14b.

4.25. Find and plot the pdf of the Weibull random variable in Problem 4.15a.

4.26. Find the cdf of the Cauchy random variable which has pdf:

$$f_X(x) = \frac{\alpha/\pi}{x^2 + \alpha^2} \qquad -\infty < x < \infty.$$

4.27. A voltage X is uniformly distributed in the set $\{-3, -2, \ldots, 3, 4\}$.

(a) Find the pdf and cdf of the random variable X.

(b) Find the pdf and cdf of the random variable $Y = -2X^2 + 3$.

(c) Find the pdf and cdf of the random variable $W = \cos(\pi X/8)$.

(d) Find the pdf and cdf of the random variable $Z = \cos^2(\pi X/8)$.

4.28. Find the pdf and cdf of the Zipf random variable in Problem 3.70.

4.29. Let C be an event for which $P[C] > 0$. Show that $F_X(x|C)$ satisfies the eight properties of a cdf.

4.30. **(a)** In Problem 4.13, find $F_X(x|C)$ where $C = \{X > 0\}$.

(b) Find $F_X(x \mid C)$ where $C = \{X = 0\}$.

4.31. **(a)** In Problem 4.10, find $F_X(x \mid B)$ where $B = \{\text{hand does not stop at 3, 6, 9, or 12 o'clock}\}$.

(b) Find $F_X(x \mid B^c)$.

4.32. In Problem 4.13, find $f_X(x \mid B)$ and $F_X(x \mid B)$ where $B = \{X > 0.25\}$.

4.33. Let X be the exponential random variable.

(a) Find and plot $F_X(x \mid X > t)$. How does $F_X(x \mid X > t)$ differ from $F_X(x)$?

(b) Find and plot $f_X(x \mid X > t)$.

(c) Show that $P[X > t + x \mid X > t] = P[X > x]$. Explain why this is called the memoryless property.

4.34. The Pareto random variable X has cdf:

$$F_X(x) = \begin{cases} 0 & x < x_m \\ 1 - \dfrac{x_m^\alpha}{x^\alpha} & x \ge x_m. \end{cases}$$

(a) Find and plot the pdf of X.

(b) Repeat Problem 4.33 parts a and b for the Pareto random variable.

(c) What happens to $P[X > t + x \mid X > t]$ as t becomes large? Interpret this result.

4.35. **(a)** Find and plot $F_X(x \mid a \le X \le b)$. Compare $F_X(x \mid a \le X \le b)$ to $F_X(x)$.

(b) Find and plot $f_X(x \mid a \le X \le b)$.

4.36. In Problem 4.6, find $F_R(r \mid R > 1)$ and $f_R(r \mid R > 1)$.

4.37. **(a)** In Problem 4.7, find $F_Z(z \mid b/4 \le Z \le b/2)$ and $f_Z(z \mid b/4 \le Z \le b/2)$.

(b) Find $F_Z(z \mid B)$ and $f_Z(z \mid B)$, where $B = \{x > b/2\}$.

4.38. A binary transmission system sends a "0" bit using a -1 voltage signal and a "1" bit by transmitting a $+1$. The received signal is corrupted by noise N that has a Laplacian distribution with parameter α. Assume that "0" bits and "1" bits are equiprobable.

(a) Find the pdf of the received signal $Y = X + N$, where X is the transmitted signal, given that a "0" was transmitted; that a "1" was transmitted.

(b) Suppose that the receiver decides a "0" was sent if $Y < 0$, and a "1" was sent if $Y \ge 0$. What is the probability that the receiver makes an error given that a $+1$ was transmitted? a -1 was transmitted?

(c) What is the overall probability of error?

Section 4.3: The Expected Value of X

4.39. Find the mean and variance of X in Problem 4.17.

4.40. Find the mean and variance of X in Problem 4.18.

4.41. Find the mean and variance of Y, the distance from the dart to the origin, in Problem 4.19.

4.42. Find the mean and variance of Z, the minimum of the coordinates in a square, in Problem 4.20.

4.43. Find the mean and variance of $X = (1 - \zeta)^{-1/2}$ in Problem 4.21. Find $E[X]$ using Eq. (4.28).

4.44. Find the mean and variance of X in Problems 4.12 and 4.22.

4.45. Find the mean and variance of X in Problems 4.13 and 4.23. Find $E[X]$ using Eq. (4.28).

4.46. Find the mean and variance of the Gaussian random variable by direct integration of Eqs. (4.27) and (4.34).

4.47. Prove Eqs. (4.28) and (4.29).

4.48. Find the variance of the exponential random variable.

4.49. **(a)** Show that the mean of the Weibull random variable in Problem 4.15 is $\Gamma(1 + 1/\beta)$ where $\Gamma(x)$ is the gamma function defined in Eq. (4.56).

(b) Find the second moment and the variance of the Weibull random variable.

4.50. Explain why the mean of the Cauchy random variable does not exist.

4.51. Show that $E[X]$ does not exist for the Pareto random variable with $\alpha = 1$ and $x_m = 1$.

4.52. Verify Eqs. (4.36), (4.37), and (4.38).

4.53. Let $Y = A\cos(\omega t) + c$ where A has mean m and variance σ^2 and ω and c are constants. Find the mean and variance of Y. Compare the results to those obtained in Example 4.15.

4.54. A limiter is shown in Fig. P4.2.

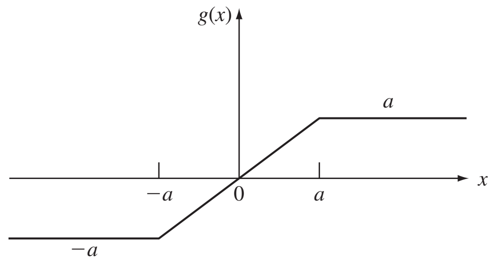


FIGURE P4.2

(a) Find an expression for the mean and variance of $Y = g(X)$ for an arbitrary continuous random variable X.

(b) Evaluate the mean and variance if X is a Laplacian random variable with $\lambda = a = 1$.

(c) Repeat part (b) if X is from Problem 4.17 with $a = 1/2$.

(d) Evaluate the mean and variance if $X = U^3$ where U is a uniform random variable in the unit interval, $[-1, 1]$ and $a = 1/2$.

4.55. A limiter with center-level clipping is shown in Fig. P4.3.

(a) Find an expression for the mean and variance of $Y = g(X)$ for an arbitrary continuous random variable X.

(b) Evaluate the mean and variance if X is Laplacian with $\lambda = a = 1$ and $b = 2$.

(c) Repeat part (b) if X is from Problem 4.22, $a = 1/2$, $b = 3/2$.

(d) Evaluate the mean and variance if $X = b\cos(2\pi U)$ where U is a uniform random variable in the unit interval $[-1, 1]$ and $a = 3/4$, $b = 1/2$.

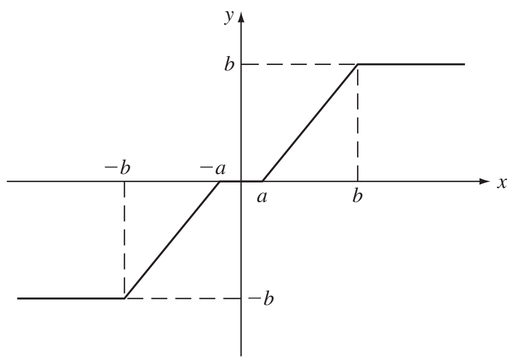


FIGURE P4.3

4.56. Let $Y = 3X + 2$.

(a) Find the mean and variance of Y in terms of the mean and variance of X.

(b) Evaluate the mean and variance of Y if X is Laplacian.

(c) Evaluate the mean and variance of Y if X is an arbitrary Gaussian random variable.

(d) Evaluate the mean and variance of Y if $X = b\cos(2\pi U)$ where U is a uniform random variable in the unit interval.

4.57. Find the nth moment of U, the uniform random variable in the unit interval. Repeat for X uniform in $[a, b]$.

4.58. Consider the quantizer in Example 4.20.

(a) Find the conditional pdf of X given that X is in the interval $(d, 2d)$.

(b) Find the conditional expected value and conditional variance of X given that X is in the interval $(d, 2d)$.

(c) Now suppose that when X falls in $(d, 2d)$, it is mapped onto the point c where $d < c < 2d$. Find an expression for the expected value of the mean square error: $E[(X - c)^2 \mid d < X < 2d]$.

(d) Find the value c that minimizes the above mean square error. Is c the midpoint of the interval? Explain why or why not by sketching possible conditional pdf shapes.

(e) Find an expression for the overall mean square error using the approach in parts c and d.

Section 4.4: Important Continuous Random Variables

4.59. Let X be a uniform random variable in the interval $[-2, 2]$. Find and plot $P[|X| > x]$.

4.60. In Example 4.20, let the input to the quantizer be a uniform random variable in the interval $[-4d, 4d]$. Show that $Z = X - Q(X)$ is uniformly distributed in $[-d/2, d/2]$.

4.61. Let X be an exponential random variable with parameter λ.

(a) For $d > 0$ and k a nonnegative integer, find $P[kd < X < (k + 1)d]$.

(b) Segment the positive real line into four equiprobable disjoint intervals.

4.62. The rth percentile, $\pi(r)$, of a random variable X is defined by $P[X \le \pi(r)] = r/100$.

(a) Find the 90%, 95%, and 99% percentiles of the exponential random variable with parameter λ.

(b) Repeat part a for the Gaussian random variable with parameters $m = 0$ and σ^2.

4.63. Let X be a Gaussian random variable with $m = 5$ and $\sigma^2 = 16$.

(a) Find $P[X > 4]$, $P[X \ge 7]$, $P[6.72 < X < 10.16]$, $P[2 < X < 7]$, $P[6 \le X \le 8]$.

(b) $P[X < a] = 0.8869$, find a.

(c) $P[X > b] = 0.11131$, find b.

(d) $P[13 < X \le c] = 0.0123$, find c.

4.64. Show that the Q-function for the Gaussian random variable satisfies $Q(-x) = 1 - Q(x)$.

4.65. Use Octave to generate Tables 4.2 and 4.3.

4.66. Let X be a Gaussian random variable with mean m and variance σ^2.

(a) Find $P[X \le m]$.

(b) Find $P[|X - m| < k\sigma]$, for $k = 1, 2, 3, 4, 5, 6$.

(c) Find the value of k for which $Q(k) = P[X > m + k\sigma] = 10^{-j}$ for $j = 1, 2, 3, 4, 5, 6$.

4.67. A binary transmission system transmits a signal X (-1 to send a "0" bit; $+1$ to send a "1" bit). The received signal is $Y = X + N$ where noise N has a zero-mean Gaussian distribution with variance σ^2. Assume that "0" bits are three times as likely as "1" bits.

(a) Find the conditional pdf of Y given the input value: $f_Y(y \mid X = +1)$ and $f_Y(y \mid X = -1)$.

(b) The receiver decides a "0" was transmitted if the observed value of y satisfies

$$f_Y(y \mid X = -1)P[X = -1] > f_Y(y \mid X = +1)P[X = +1]$$

and it decides a "1" was transmitted otherwise. Use the results from part a to show that this decision rule is equivalent to: If $y < T$ decide "0"; if $y \ge T$ decide "1".

(c) What is the probability that the receiver makes an error given that a $+1$ was transmitted? a -1 was transmitted? Assume $\sigma^2 = 1/16$.

(d) What is the overall probability of error?

4.68. Two chips are being considered for use in a certain system. The lifetime of chip 1 is modeled by a Gaussian random variable with mean 20,000 hours and standard deviation 5000 hours. (The probability of negative lifetime is negligible.) The lifetime of chip 2 is also a Gaussian random variable but with mean 22,000 hours and standard deviation 1000 hours. Which chip is preferred if the target lifetime of the system is 20,000 hours? 24,000 hours?

4.69. Passengers arrive at a taxi stand at an airport at a rate of one passenger per minute. The taxi driver will not leave until seven passengers arrive to fill his van. Suppose that passenger interarrival times are exponential random variables, and let X be the time to fill a van. Find the probability that more than 10 minutes elapse until the van is full.

4.70. **(a)** Show that the gamma random variable has mean:

$$E[X] = \alpha/\lambda.$$

(b) Show that the gamma random variable has second moment, and variance given by:

$$E[X^2] = \alpha(\alpha + 1)/\lambda^2 \text{ and } \mathrm{VAR}[X] = \alpha/\lambda^2.$$

(c) Use parts a and b to obtain the mean and variance of an m-Erlang random variable.

(d) Use parts a and b to obtain the mean and variance of a chi-square random variable.

4.71. The time X to complete a transaction in a system is a gamma random variable with mean 4 and variance 8. Use Octave to plot $P[X > x]$ as a function of x. Note: Octave uses $\beta = 1/2$.

4.72. **(a)** Plot the pdf of an m-Erlang random variable for $m = 1, 2, 3$ and $\lambda = 1$.

(b) Plot the chi-square pdf for $k = 1, 2, 3$.

4.73. A repair person keeps four widgets in stock. What is the probability that the widgets in stock will last 15 days if the repair person needs to replace widgets at an average rate of one widget every three days, where the time between widget failures is an exponential random variable?

4.74. **(a)** Find the cdf of the m-Erlang random variable by integration of the pdf. *Hint:* Use integration by parts.

(b) Show that the derivative of the cdf given by Eq. (4.58) gives the pdf of an m-Erlang random variable.

4.75. Plot the pdf of a beta random variable with: $a = b = 1/4, 1, 4, 8$; $a = 5, b = 1$; $a = 1, b = 3$; $a = 2, b = 5$.

Section 4.5: Functions of a Random Variable

4.76. Let X be a Gaussian random variable with mean 2 and variance 4. The reward in a system is given by $Y = (X)^+$. Find the pdf of Y.

4.77. The amplitude of a radio signal X is a Rayleigh random variable with pdf:

$$f_X(x) = \frac{x}{\alpha^2} e^{-x^2/2\alpha^2} \qquad x > 0, \quad \alpha > 0.$$

(a) Find the pdf of $Z = (X - r)^+$.

(b) Find the pdf of $Z = X^2$.

4.78. A wire has length X, an exponential random variable with mean 5π cm. The wire is cut to make rings of diameter 1 cm. Find the probability for the number of complete rings produced by each length of wire.

4.79. A signal that has amplitudes with a Gaussian pdf with zero mean and unit variance is applied to the quantizer in Example 4.27.

(a) Pick d so that the probability that X falls outside the range of the quantizer is 1%.

(b) Find the probability of the output levels of the quantizer.

4.80. The signal X is amplified and shifted as follows: $Y = 2X + 3$, where X is the random variable in Problem 4.12. Find the cdf and pdf of Y.

4.81. The net profit in a transaction is given by $Y = 2 - 4X$ where X is the random variable in Problem 4.13. Find the cdf and pdf of Y.

4.82. Find the cdf and pdf of the output of the limiter in Problem 4.54 parts b, c, and d.

4.83. Find the cdf and pdf of the output of the limiter with center-level clipping in Problem 4.55 parts b, c, and d.

4.84. Find the cdf and pdf of $Y = 3X + 2$ in Problem 4.56 parts b, c, and d.

4.85. The exam grades in a certain class have a Gaussian pdf with mean m and standard deviation σ. Find the constants a and b so that the random variable $y = aX + b$ has a Gaussian pdf with mean m' and standard deviation σ'.

4.86. Let $X = U^n$ where n is a positive integer and U is a uniform random variable in the unit interval. Find the cdf and pdf of X.

4.87. Repeat Problem 4.86 if U is uniform in the interval $[-1, 1]$.

4.88. Let $Y = |X|$ be the output of a full-wave rectifier with input voltage X.

(a) Find the cdf of Y by finding the equivalent event of $\{Y \le y\}$. Find the pdf of Y by differentiation of the cdf.

(b) Find the pdf of Y by finding the equivalent event of $\{y < Y \le y + dy\}$. Does the answer agree with part a?

(c) What is the pdf of Y if the $f_X(x)$ is an even function of x?

4.89. Find and plot the cdf of Y in Example 4.34.

4.90. A voltage X is a Gaussian random variable with mean 1 and variance 2. Find the pdf of the power dissipated by an R-ohm resistor $P = RX^2$.

4.91. Let $Y = e^X$.

(a) Find the cdf and pdf of Y in terms of the cdf and pdf of X.

(b) Find the pdf of Y when X is a Gaussian random variable. In this case Y is said to be a lognormal random variable. Plot the pdf and cdf of Y when X is zero-mean with variance 1/8; repeat with variance 8.

4.92. Let a radius be given by the random variable X in Problem 4.18.

(a) Find the pdf of the area covered by a disc with radius X.

(b) Find the pdf of the volume of a sphere with radius X.

(c) Find the pdf of the volume of a sphere in R^n:

$$Y = \begin{cases} (2\pi)^{(n-1)/2}\, X^n/(2 \times 4 \times \cdots \times n) & \text{for } n \text{ even} \\ 2(2\pi)^{(n-1)/2}\, X^n/(1 \times 3 \times \cdots \times n) & \text{for } n \text{ odd.} \end{cases}$$

4.93. In the quantizer in Example 4.20, let $Z = X - q(X)$. Find the pdf of Z if X is a Laplacian random variable with parameter $\alpha = d/2$.

4.94. Let $Y = \alpha \tan \pi X$, where X is uniformly distributed in the interval $(-1, 1)$.

(a) Show that Y is a Cauchy random variable.

(b) Find the pdf of $Y = 1/X$.

4.95. Let X be a Weibull random variable in Problem 4.15. Let $Y = (X/\lambda)^\beta$. Find the cdf and pdf of Y.

4.96. Find the pdf of $X = -\ln(1 - U)$, where U is a uniform random variable in $(0, 1)$.

Section 4.6: The Markov and Chebyshev Inequalities

4.97. Compare the Markov inequality and the exact probability for the event $\{X > c\}$ as a function of c for:

(a) X is a uniform random variable in the interval $[0, b]$.

(b) X is an exponential random variable with parameter λ.

(c) X is a Pareto random variable with $\alpha > 1$.

(d) X is a Rayleigh random variable.

4.98. Compare the Markov inequality and the exact probability for the event $\{X > c\}$ as a function of c for:

(a) X is a uniform random variable in $\{1, 2, \ldots, L\}$.

(b) X is a geometric random variable.

(c) X is a Zipf random variable with $L = 10$; $L = 100$.

(d) X is a binomial random variable with $n = 10$, $p = 0.5$; $n = 50$, $p = 0.5$.

4.99. Compare the Chebyshev inequality and the exact probability for the event $\{|X - m| > c\}$ as a function of c for:

(a) X is a uniform random variable in the interval $[-b, b]$.

(b) X is a Laplacian random variable with parameter α.

(c) X is a zero-mean Gaussian random variable.

(d) X is a binomial random variable with $n = 10$, $p = 0.5$; $n = 50$, $p = 0.5$.

4.100. Let X be the number of successes in n Bernoulli trials where the probability of success is p. Let $Y = X/n$ be the average number of successes per trial. Apply the Chebyshev inequality to the event $\{|Y - p| > a\}$. What happens as $n \to \infty$?

4.101. Suppose that light bulbs have exponentially distributed lifetimes with unknown mean $E[X]$. Suppose we measure the lifetime of n light bulbs, and we estimate the mean $E[X]$ by the arithmetic average Y of the measurements. Apply the Chebyshev inequality to the event $\{|Y - E[X]| > a\}$. What happens as $n \to \infty$? *Hint:* Use the m-Erlang random variable.

Section 4.7: Transform Methods

4.102. **(a)** Find the characteristic function of the uniform random variable in $[-b, b]$.

(b) Find the mean and variance of X by applying the moment theorem.

4.103. **(a)** Find the characteristic function of the Laplacian random variable.

(b) Find the mean and variance of X by applying the moment theorem.

4.104. Let $\Phi_X(\omega)$ be the characteristic function of an exponential random variable. What random variable does $\Phi_X^n(\omega)$ correspond to?

4.105. Find the mean and variance of the Gaussian random variable by applying the moment theorem to the characteristic function given in Table 4.1.

4.106. Find the characteristic function of $Y = aX + b$ where X is a Gaussian random variable. *Hint:* Use Eq. (4.79).

4.107. Show that the characteristic function for the Cauchy random variable is $e^{-|\omega|}$.

4.108. Find the Chernoff bound for the exponential random variable with $\lambda = 1$. Compare the bound to the exact value for $P[X > 5]$.

4.109. **(a)** Find the probability generating function of the geometric random variable.
(b) Find the mean and variance of the geometric random variable from its pgf.

4.110. **(a)** Find the pgf for the binomial random variable X with parameters n and p.
(b) Find the mean and variance of X from the pgf.

4.111. Let $G_X(z)$ be the pgf for a binomial random variable with parameters n and p, and let $G_Y(z)$ be the pgf for a binomial random variable with parameters m and p. Consider the function $G_X(z)\, G_Y(z)$. Is this a valid pgf? If so, to what random variable does it correspond?

4.112. Let $G_N(z)$ be the pgf for a Poisson random variable with parameter α, and let $G_M(z)$ be the pgf for a Poisson random variable with parameters β. Consider the function $G_N(z)\, G_M(z)$. Is this a valid pgf? If so, to what random variable does it correspond?

4.113. Let N be a Poisson random variable with parameter $\alpha = 1$. Compare the Chernoff bound and the exact value for $P[X \geq 5]$.

4.114. **(a)** Find the pgf $G_U(z)$ for the discrete uniform random variable U.
(b) Find the mean and variance from the pgf.
(c) Consider $G_U(z)^2$. Does this function correspond to a pgf? If so, find the mean of the corresponding random variable.

4.115. **(a)** Find $P[X = r]$ for the negative binomial random variable from the pgf in Table 3.1.
(b) Find the mean of X.

4.116. Derive Eq. (4.89).

4.117. Obtain the nth moment of a gamma random variable from the Laplace transform of its pdf.

4.118. Let X be the mixture of two exponential random variables (see Example 4.58). Find the Laplace transform of the pdf of X.

4.119. The Laplace transform of the pdf of a random variable X is given by:

$$X^*(s) = \frac{a}{s+a}\frac{b}{s+b}.$$

Find the pdf of X. *Hint:* Use a partial fraction expansion of $X^*(s)$.

4.120. Find a relationship between the Laplace transform of a gamma random variable pdf with parameters α and λ and the Laplace transform of a gamma random variable with parameters $\alpha - 1$ and λ. What does this imply if X is an m-Erlang random variable?

4.121. **(a)** Find the Chernoff bound for $P[X > t]$ for the gamma random variable.
(b) Compare the bound to the exact value of $P[X \geq 9]$ for an $m = 3$, $\lambda = 1$ Erlang random variable.

Section 4.8: Basic Reliability Calculations

4.122. The lifetime T of a device has pdf

$$f_T(t) = \begin{cases} 1/10T_0 & 0 < t < T_0 \\ 0.9\lambda e^{-\lambda(t-T_0)} & t \geq T_0 \\ 0 & t < T_0. \end{cases}$$

(a) Find the reliability and MTTF of the device.

(b) Find the failure rate function.

(c) How many hours of operation can be considered to achieve 99% reliability?

4.123. The lifetime T of a device has pdf

$$f_T(t) = \begin{cases} 1/T_0 & a \leq t \leq a + T_0 \\ 0 & \text{elsewhere.} \end{cases}$$

(a) Find the reliability and MTTF of the device.

(b) Find the failure rate function.

(c) How many hours of operation can be considered to achieve 99% reliability?

4.124. The lifetime T of a device is a Rayleigh random variable.

(a) Find the reliability of the device.

(b) Find the failure rate function. Does $r(t)$ increase with time?

(c) Find the reliability of two devices that are in series.

(d) Find the reliability of two devices that are in parallel.

4.125. The lifetime T of a device is a Weibull random variable.

(a) Plot the failure rates for $\alpha = 1$ and $\beta = 0.5$; for $\alpha = 1$ and $\beta = 2$.

(b) Plot the reliability functions in part a.

(c) Plot the reliability of two devices that are in series.

(d) Plot the reliability of two devices that are in parallel.

4.126. A system starts with m devices, 1 active and $m - 1$ on standby. Each device has an exponential lifetime. When a device fails it is immediately replaced with another device (if one is still available).

(a) Find the reliability of the system.

(b) Find the failure rate function.

4.127. Find the failure rate function of the memory chips discussed in Example 2.28. Plot $\text{In}(r(t))$ versus αt.

4.128. A device comes from two sources. Devices from source 1 have mean m and exponentially distributed lifetimes. Devices from source 2 have mean m and Pareto-distributed lifetimes with $\alpha > 1$. Assume a fraction p is from source 1 and a fraction $1 - p$ from source 2.

(a) Find the reliability of an arbitrarily selected device.

(b) Find the failure rate function.

4.129. A device has the failure rate function:

$$r(t) = \begin{cases} 1 + 9(1 - t) & 0 \le t < 1 \\ 1 & 1 \le t < 10 \\ 1 + 10(t - 10) & t \ge 10. \end{cases}$$

Find the reliability function and the pdf of the device.

4.130. A system has three identical components and the system is functioning if two or more components are functioning.

(a) Find the reliability and MTTF of the system if the component lifetimes are exponential random variables with mean 1.

(b) Find the reliability of the system if one of the components has mean 2.

4.131. Repeat Problem 4.130 if the component lifetimes are Weibull distributed with $\beta = 3$.

4.132. A system consists of two processors and three peripheral units. The system is functioning as long as one processor and two peripherals are functioning.

(a) Find the system reliability and MTTF if the processor lifetimes are exponential random variables with mean 5 and the peripheral lifetimes are Rayleigh random variables with mean 10.

(b) Find the system reliability and MTTF if the processor lifetimes are exponential random variables with mean 10 and the peripheral lifetimes are exponential random variables with mean 5.

4.133. An operation is carried out by a subsystem consisting of three units that operate in a series configuration.

(a) The units have exponentially distributed lifetimes with mean 1. How many subsystems should be operated in parallel to achieve a reliability of 99% in T hours of operation?

(b) Repeat part a with Rayleigh-distributed lifetimes.

(c) Repeat part a with Weibull-distributed lifetimes with $\beta = 3$.

Section 4.9: Computer Methods for Generating Random Variables

4.134. Octave provides function calls to evaluate the pdf and cdf of important continuous random variables. For example, the functions `\normal_cdf(x, m, var)` and `normal_pdf(x, m, var)` compute the cdf and pdf, respectively, at `x` for a Gaussian random variable with mean `m` and variance `var`.

(a) Plot the conditional pdfs in Example 4.11 if $v = \pm 2$ and the noise is zero-mean and unit variance.

(b) Compare the cdf of the Gaussian random variable with the Chernoff bound obtained in Example 4.44.

4.135. Plot the pdf and cdf of the gamma random variable for the following cases.

(a) $\lambda = 1$ and $\alpha = 1, 2, 4$.

(b) $\lambda = 1/2$ and $\alpha = 1/2, 1, 3/2, 5/2$.

4.136. The random variable X has the triangular pdf shown in Fig. P4.4.

(a) Find the transformation needed to generate X.

(b) Use Octave to generate 100 samples of X. Compare the empirical pdf of the samples with the desired pdf.

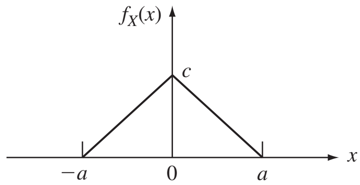


FIGURE P4.4

4.137. For each of the following random variables: Find the transformation needed to generate the random variable X; use Octave to generate 1000 samples of X; Plot the sequence of outcomes; compare the empirical pdf of the samples with the desired pdf.

(a) Laplacian random variable with $\alpha = 1$.

(b) Pareto random variable with $\alpha = 1.5, 2, 2.5$.

(c) Weibull random variable with $\beta = 0.5, 2, 3$ and $\lambda = 1$.

4.138. A random variable Y of mixed type has pdf

$$f_Y(x) = p\delta(x) + (1 - p)f_Y(x),$$

where X is a Laplacian random variable and p is a number between zero and one. Find the transformation required to generate Y.

4.139. Specify the transformation method needed to generate the geometric random variable with parameter $p = 1/2$. Find the average number of comparisons needed in the search to determine each outcome.

4.140. Specify the transformation method needed to generate the Poisson random variable with small parameter α. Compute the average number of comparisons needed in the search.

4.141. The following rejection method can be used to generate Gaussian random variables:

1. Generate U_1, a uniform random variable in the unit interval.
2. Let $X_1 = -\ln(U_1)$.
3. Generate U_2, a uniform random variable in the unit interval. If $U_2 \le \exp\{-(\mathrm{X}_1 - 1)^2/2\}$, accept X_1. Otherwise, reject X_1 and go to step 1.
4. Generate a random sign $(+ \text{ or } -)$ with equal probability. Output X equal to X_1 with the resulting sign.

(a) Show that if X_1 is accepted, then its pdf corresponds to the pdf of the absolute value of a Gaussian random variable with mean 0 and variance 1.

(b) Show that X is a Gaussian random variable with mean 0 and variance 1.

4.142. Cheng (1977) has shown that the function $Kf_Z(x)$ bounds the pdf of a gamma random variable with $\alpha > 1$, where

$$f_Z(x) = \frac{\lambda\alpha^\lambda x^{\lambda-1}}{(\alpha^\lambda + x^\lambda)^2} \qquad \text{and} \qquad K = (2\alpha - 1)^{1/2}.$$

Find the cdf of $f_Z(x)$ and the corresponding transformation needed to generate Z.

4.143. **(a)** Show that in the modified rejection method, the probability of accepting X_1 is $1/K$. *Hint:* Use conditional probability.

(b) Show that Z has the desired pdf.

4.144. Two methods for generating binomial random variables are: (1) Generate n Bernoulli random variables and add the outcomes; (2) Divide the unit interval according to binomial probabilities. Compare the methods under the following conditions:

(a) $p = 1/2, n = 5, 25, 50$;

(b) $p = 0.1, n = 5, 25, 50$.

(c) Use Octave to implement the two methods by generating 1000 binomially distributed samples.

4.145. Let the number of event occurrences in a time interval be a Poisson random variable. In Section 3.4, it was found that the time between events for a Poisson random variable is an exponentially distributed random variable.

(a) Explain how one can generate Poisson random variables from a sequence of exponentially distributed random variables.

(b) How does this method compare with the one presented in Problem 4.140?

(c) Use Octave to implement the two methods when $\alpha = 3$, $\alpha = 25$, and $\alpha = 100$.

4.146. Write a program to generate the gamma pdf with $\alpha > 1$ using the rejection method discussed in Problem 4.142. Use this method to generate m-Erlang random variables with $m = 2, 10$ and $\lambda = 1$ and compare the method to the straightforward generation of m exponential random variables as discussed in Example 4.57.

*Section 4.10: Entropy

4.147. Let X be the outcome of the toss of a fair die.

(a) Find the entropy of X.

(b) Suppose you are told that X is even. What is the reduction in entropy?

4.148. A biased coin is tossed three times.

(a) Find the entropy of the outcome if the sequence of heads and tails is noted.

(b) Find the entropy of the outcome if the number of heads is noted.

(c) Explain the difference between the entropies in parts a and b.

4.149. Let X be the number of tails until the first heads in a sequence of tosses of a biased coin.

(a) Find the entropy of X given that $X \geq k$.

(b) Find the entropy of X given that $X \leq k$.

4.150. One of two coins is selected at random: Coin A has $P[\text{heads}] = 1/10$ and coin B has $P[\text{heads}] = 9/10$.

(a) Suppose the coin is tossed once. Find the entropy of the outcome.

(b) Suppose the coin is tossed twice and the sequence of heads and tails is observed. Find the entropy of the outcome.

4.151. Suppose that the randomly selected coin in Problem 4.150 is tossed until the first occurrence of heads. Suppose that heads occurs in the kth toss. Find the entropy regarding the identity of the coin.

4.152. A communication channel accepts input I from the set $\{0, 1, 2, 3, 4, 5, 6\}$. The channel output is $X = I + N \bmod 7$, where N is equally likely to be $+1$ or -1.

(a) Find the entropy of I if all inputs are equiprobable.

(b) Find the entropy of I given that $X = 4$.

4.153. Let X be a discrete random variable with entropy H_X.

(a) Find the entropy of $Y = 2X$.

(b) Find the entropy of any invertible transformation of X.

4.154. Let (X, Y) be the pair of outcomes from two independent tosses of a die.

(a) Find the entropy of X.

(b) Find the entropy of the pair (X, Y).

(c) Find the entropy in n independent tosses of a die. Explain why entropy is additive in this case.

4.155. Let X be the outcome of the toss of a die, and let Y be a randomly selected integer less than or equal to X.

(a) Find the entropy of Y.

(b) Find the entropy of the pair (X, Y) and denote it by $H(X, Y)$.

(c) Find the entropy of Y given $X = k$ and denote it by $g(k) = H(Y \mid X = k)$. Find $E[g(X)] = E[H(Y \mid X)]$.

(d) Show that $H(X, Y) = H_X + E[H(Y \mid X)]$. Explain the meaning of this equation.

4.156. Let X take on values from $\{1, 2, \ldots, K\}$. Suppose that $P[X = K] = p$, and let H_Y be the entropy of X given that X is not equal to K. Show that $H_X = -p \ln p - (1 - p) \ln(1 - p) + (1 - p)H_Y$.

4.157. Let X be a uniform random variable in Example 4.62. Find and plot the entropy of Q as a function of the variance of the error $X - Q(X)$. *Hint:* Express the variance of the error in terms of d and substitute into the expression for the entropy of Q.

4.158. A communication channel accepts as input either 000 or 111. The channel transmits each binary input correctly with probability $1 - p$ and erroneously with probability p. Find the entropy of the input given that the output is 000; given that the output is 010.

4.159. Let X be a uniform random variable in the interval $[-a, a]$. Suppose we are told that the X is positive. Use the approach in Example 4.62 to find the reduction in entropy. Show that this is equal to the difference of the differential entropy of X and the differential entropy of X given $\{X > 0\}$.

4.160. Let X be uniform in $[a, b]$, and let $Y = 2X$. Compare the differential entropies of X and Y. How does this result differ from the result in Problem 4.153?

4.161. Find the pmf for the random variable X for which the sequence of questions in Fig. 4.26(a) is optimum.

4.162. Let the random variable X have $S_X = \{1, 2, 3, 4, 5, 6\}$ and pmf (3/8, 3/8, 1/8, 1/16, 1/32, 1/32). Find the entropy of X. What is the best code you can find for X?

4.163. Seven cards are drawn from a deck of 52 distinct cards. How many bits are required to represent all possible outcomes?

4.164. Find the optimum encoding for the geometric random variable with $p = 1/2$.

4.165. An urn experiment has 10 equiprobable distinct outcomes. Find the performance of the best tree code for encoding (a) a single outcome of the experiment; (b) a sequence of n outcomes of the experiment.

4.166. A binary information source produces n outputs. Suppose we are told that there are k 1's in these n outputs.

(a) What is the best code to indicate which pattern of k 1's and $n - k$ 0's occurred?

(b) How many bits are required to specify the value of k using a code with a fixed number of bits?

4.167. The random variable X takes on values from the set $\{1, 2, 3, 4\}$. Find the maximum entropy pmf for X given that $E[X] = 2$.

4.168. The random variable X is nonnegative. Find the maximum entropy pdf for X given that $E[X] = 10$.

4.169. Find the maximum entropy pdf of X given that $E[X^2] = c$.

4.170. Suppose we are given two parameters of the random variable X, $E[g_1(X)] = c_1$ and $E[g_2(X)] = c_2$.

(a) Show that the maximum entropy pdf for X has the form

$$f_X(x) = Ce^{-\lambda_1 g_1(x) - \lambda_2 g_2(x)}.$$

(b) Find the entropy of X.

4.171. Find the maximum entropy pdf of X given that $E[X] = m$ and $\text{VAR}[X] = \sigma^2$.

Problems Requiring Cumulative Knowledge

4.172. Three types of customers arrive at a service station. The time required to service type 1 customers is an exponential random variable with mean 2. Type 2 customers have a Pareto distribution with $\alpha = 3$ and $x_m = 1$. Type 3 customers require a constant service time of 2 seconds. Suppose that the proportion of type 1, 2, and 3 customers is 1/2, 1/8, and 3/8, respectively. Find the probability that an arbitrary customer requires more than 15 seconds of service time. Compare the above probability to the bound provided by the Markov inequality.

4.173. The lifetime X of a light bulb is a random variable with

$$P[X > t] = 2/(2 + t) \text{ for } t > 0.$$

Suppose three new light bulbs are installed at time $t = 0$. At time $t = 1$ all three light bulbs are still working. Find the probability that at least one light bulb is still working at time $t = 9$.

4.174. The random variable X is uniformly distributed in the interval $[0, a]$. Suppose a is unknown, so we estimate a by the maximum value observed in n independent repetitions of the experiment; that is, we estimate a by $Y = \max\{X_1, X_2, \ldots, X_n\}$.

(a) Find $P[Y \le y]$.

(b) Find the mean and variance of Y, and explain why Y is a good estimate for a when N is large.

4.175. The sample X of a signal is a Gaussian random variable with $m = 0$ and $\sigma^2 = 1$. Suppose that X is quantized by a nonuniform quantizer consisting of four intervals: $(-\infty, -a]$, $(-a, 0]$, $(0, a]$, and (a, ∞).

(a) Find the value of a so that X is equally likely to fall in each of the four intervals.

(b) Find the representation point $x_i = q(X)$ for X in $(0, a]$ that minimizes the mean-squared error, that is,

$$\int_0^a (x - x_1)^2 f_X(x)\, dx \text{ is minimized.}$$

Hint: Differentiate the above expression with respect to x_i. Find the representation points for the other intervals.

(c) Evaluate the mean-squared error of the quantizer $E[(X - q(X)^2]$.

4.176. The output Y of a binary communication system is a unit-variance Gaussian random with mean zero when the input is "0" and mean one when the input is "one". Assume the input is 1 with probability p.

(a) Find $P[\text{input is } 1 \mid y < Y < y + h]$ and $P[\text{input is } 0 \mid y < Y < y + h]$.

(b) The receiver uses the following decision rule:

If $P[\text{input is } 1 \mid y < Y < y + h] > P[\text{input is } 0 \mid y < Y < y + h]$, decide input was 1; otherwise, decide input was 0.

Show that this decision rule leads to the following threshold rule:

If $Y > T$, decide input was 1; otherwise, decide input was 0.

(c) What is the probability of error for the above decision rule?