# 2

# PROBABILITY AND STOCHASTIC PROCESSES

The theory of probability and stochastic processes is an essential mathematical tool in the design of digital communication systems. This subject is important in the statistical modeling of sources that generate the information, in the digitization of the source output, in the characterization of the channel through which the digital information is transmitted, in the design of the receiver that processes the information-bearing signal from the channel, and in the evaluation of the performance of the communication system. Our coverage of this rich and interesting subject is brief and limited in scope. We present a number of definitions and basic concepts in the theory of probability and stochastic processes and we derive several results that are important in the design of efficient digital communication systems and in the evaluation of their performance.

We anticipate that most readers have had some prior exposure to the theory of probability and stochastic processes, so that our treatment serves primarily as a review. Some readers, however, who have had no previous exposure may find the presentation in this chapter extremely brief. These readers will benefit from additional reading of engineering-level treatments of the subject found in the texts by Davenport and Root (1958), Davenport (1970), Papoulis (1984), Helstrom (1991), and Leon-Garcia (1994).

## 2-1 PROBABILITY

Let us consider an experiment, such as the rolling of a die, with a number of possible outcomes. The sample space $S$ of the experiment consists of the set of all possible outcomes. In the case of the die,

$$S = \{1, 2, 3, 4, 5, 6\} \tag{2-1-1}$$

where the integers $1, \ldots, 6$ represent the number of dots on the six faces of the die. These six possible outcomes are the sample points of the experiment. An event is a subset of $S$, and may consist of any number of sample points. For example, the event $A$ defined as

$$A = \{2, 4\} \tag{2-1-2}$$

consists of the outcomes 2 and 4. The complement of the event $A$, denoted by $\bar{A}$, consists of all the sample points in $S$ that are not in $A$ and, hence,

$$\bar{A} = \{1, 3, 5, 6\} \tag{2-1-3}$$

Two events are said to be mutally exclusive if they have no sample points in common—that is, if the occurrence of one event excludes the occurrence of the other. For example, if $A$ is defined as in (2-1-2) and the event $B$ is defined as

$$B = \{1, 3, 6\} \tag{2-1-4}$$

then $A$ and $B$ are mutually exclusive events. Similarly, $A$ and $\bar{A}$ are mutually exclusive events.

The union (sum) of two events is an event that consists of all the sample points in the two events. For example, if $B$ is the event defined in (2-1-4) and $C$ is the event defined as

$$C = \{1, 2, 3\} \tag{2-1-5}$$

then, the union of $B$ and $C$, denoted by $B \cup C$, is the event

$$\begin{aligned} D &= B \cup C \\ &= \{1, 2, 3, 6\} \end{aligned} \tag{2-1-6}$$

Similarly, $A \cup \bar{A} = S$, where $S$ is the entire sample space or the certain event. On the other hand, the intersection of two events is an event that consists of the points that are common to the two events. Thus, if $E = B \cap C$ represents the intersection of the events $B$ and $C$, defined by (2-1-4) and (2-1-5), respectively, then

$$E = \{1, 3\}$$

When the events are mutually exclusive, the intersection is the null event, denoted as $\varnothing$. For example, $A \cap B = \varnothing$, and $A \cap \bar{A} = \varnothing$. The definitions of union and intersection are extended to more than two events in a straightforward manner.

Associated with each event $A$ contained in $S$ is its probability $P(A)$. In the assignment of probabilities to events, we adopt an axiomatic viewpoint. That is, we postulate that the probability of the event $A$ satisfies the condition $P(A) \geqslant 0$. We also postulate that the probability of the sample space (certain event) is $P(S) = 1$. The third postulate deals with the probability of mutually exclusive events. Suppose that $A_i$, $i = 1, 2, \ldots$, are a (possibly infinite) number of events in the sample space $S$ such that

$$A_i \cap A_j = \varnothing \quad i \neq j = 1, 2, \ldots$$

Then the probability of the union of these mutually exclusive events satisfies the condition

$$P\Big(\bigcup_i A_i\Big) = \sum_i P(A_i) \tag{2-1-7}$$

For example, in a roll of a fair die, each possible outcome is assigned the probability $\frac{1}{6}$. The event $A$ defined by (2-1-2) consists of two mutually exclusive subevents or outcomes, and, hence, $P(A) = \frac{2}{6} = \frac{1}{3}$. Also, the probability of the event $A \cup B$, where $A$ and $B$ are the mutually exclusive events defined by (2-1-2) and (2-1-4), respectively, is $P(A) + P(B) = \frac{1}{3} + \frac{1}{2} = \frac{5}{6}$.

**Joint Events and Joint Probabilities** Instead of dealing with a single experiment, let us perform two experiments and consider their outcomes. For example, the two experiments may be two separate tosses of a single die or a single toss of two dice. In either case, the sample space $S$ consists of the 36 two-tuples $(i, j)$ where $i, j = 1, 2, \ldots, 6$. If the dice are fair, each point in the sample space is assigned the probability $\frac{1}{36}$. We may now consider joint events, such as $\{i \text{ is even}, j = 3\}$, and determine the associated probabilities of such events from knowledge of the probabilities of the sample points.

In general, if one experiment has the possible outcomes $A_i$, $i = 1, 2, \ldots, n$, and the second experiment has the possible outcomes $B_j$, $j = 1, 2, \ldots, m$, then the combined experiment has the possible joint outcomes $(A_i, B_j)$, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m$. Associated with each joint outcome $(A_i, B_j)$ is the joint probability $P(A_i, B_j)$ which satisfies the condition

$$0 \leqslant P(A_i, B_j) \leqslant 1$$

Assuming that the outcomes $B_j$, $j = 1, 2, \ldots, m$, are mutually exclusive, it follows that

$$\sum_{j=1}^{m} P(A_i, B_j) = P(A_i) \tag{2-1-8}$$

Similarly, if the outcomes $A_i$, $i = 1, 2, \ldots, n$, are mutually exclusive then

$$\sum_{i=1}^{n} P(A_i, B_j) = P(B_j) \tag{2-1-9}$$

Furthermore, if all the outcomes of the two experiments are mutually exclusive then

$$\sum_{i=1}^{n} \sum_{j=1}^{m} P(A_i, B_j) = 1 \tag{2-1-10}$$

The generalization of the above treatment to more than two experiments is straightforward.

**Conditional Probabilities** Consider a combined experiment in which a joint event occurs with probability $P(A, B)$. Suppose that the event $B$ has occurred and we wish to determine the probability of occurrence of the event $A$. This is called the *conditional probability* of the event $A$ given the occurrence of the event $B$ and is defined as

$$P(A \mid B) = \frac{P(A, B)}{P(B)} \tag{2-1-11}$$

provided $P(B) > 0$. In a similar manner, the probability of the event $B$ conditioned on the occurrence of the event $A$ is defined as

$$P(B \mid A) = \frac{P(A, B)}{P(A)} \tag{2-1-12}$$

provided $P(A) > 0$. The relations in (2-1-11) and (2-1-12) may also be expressed as

$$P(A, B) = P(A \mid B)P(B) = P(B \mid A)P(A) \tag{2-1-13}$$

The relations in (2-1-11), (2-1-12), and (2-1-13) also apply to a single experiment in which $A$ and $B$ are any two events defined on the sample space $S$ and $P(A, B)$ is interpreted as the probability of the $A \cap B$. That is, $P(A, B)$ denotes the simultaneous occurrence of $A$ and $B$. For example, consider the events $B$ and $C$ given by (2-1-4) and (2-1-5), respectively, for the single toss of a die. The joint event consists of the sample points $\{1, 3\}$. The conditional probability of the event $C$ given that $B$ occurred is

$$P(C \mid B) = \frac{\frac{2}{6}}{\frac{3}{6}} = \frac{2}{3}$$

In a single experiment, we observe that when two events $A$ and $B$ are mutually exclusive, $A \cap B = \varnothing$ and, hence, $P(A \mid B) = 0$. Also, if $A$ is a subset of $B$ then $A \cap B = A$ and, hence,

$$P(A \mid B) = \frac{P(A)}{P(B)}$$

On the other hand, if $B$ is a subset of $A$, we have $A \cap B = B$ and, hence,

$$P(A \mid B) = \frac{P(B)}{P(B)} = 1$$

An extremely useful relationship for conditional probabilities is Bayes' theorem, which states that if $A_i$, $i = 1, 2, \ldots, n$, are mutually exclusive events such that

$$\bigcup_{i=1}^{n} A_i = S$$

and $B$ is an arbitrary event with nonzero probability then

$$\begin{aligned} P(A_i \mid B) &= \frac{P(A_i, B)}{P(B)} \\ &= \frac{P(B \mid A_i)P(A_i)}{\sum_{j=1}^{n} P(B \mid A_j)P(A_j)} \end{aligned} \tag{2-1-14}$$

We use this formula in Chapter 5 to derive the structure of the optimum receiver for a digital communication system in which the events $A_i$, $i = 1, 2, \ldots, n$, represent the possible transmitted messages in a given time interval, $P(A_i)$ represent their *a priori* probabilities, $B$ represents the received signal, which consists of the transmitted message (one of the $A_i$) corrupted by noise, and $P(A_i \mid B)$ is the *a posteriori* probability of $A_i$ conditioned on having observed the received signal $B$.

**Statistical Independence** The statistical independence of two or more events is another important concept in probability theory. It usually arises when we consider two or more experiments or repeated trials of a single experiment. To explain this concept, we consider two events $A$ and $B$ and their conditional probability $P(A \mid B)$, which is the probability of occurrence of $A$ given that $B$ has occurred. Suppose that the occurrence of $A$ does not depend on the occurrence of $B$. That is,

$$P(A \mid B) = P(A) \tag{2-1-15}$$

Substitution of (2-1-15) into (2-1-13) yields the result

$$P(A, B) = P(A)P(B) \tag{2-1-16}$$

That is, the joint probability of the events $A$ and $B$ factors into the product of the elementary or marginal probabilities $P(A)$ and $P(B)$. When the events $A$ and $B$ satisfy the relation in (2-1-16), they are said to be *statistically independent.*

For example, consider two successive experiments in tossing a die. Let $A$ represent the even-numbered sample points $\{2, 4, 6\}$ in the first toss and $B$ represent the even-numbered possible outcomes $\{2, 4, 6\}$ in the second toss. In a fair die, we assign the probabilities $P(A) = \frac{1}{2}$ and $P(B) = \frac{1}{2}$. Now, the joint probability of the joint event "even-numbered outcome on the first toss and even-numbered outcome on the second toss" is just the probability of the nine pairs of outcomes $(i, j)$, $i = 2, 4, 6$; $j = 2, 4, 6$, which is $\frac{1}{4}$. Also,

$$P(A, B) = P(A)P(B) = \tfrac{1}{4}$$

Thus, the events $A$ and $B$ are statistically independent. Similarly, we may say that the outcomes of the two experiments are statistically independent.

The definition of statistical independence can be extended to three or more events. Three statistically independent events $A_1$, $A_2$, and $A_3$ must satisfy the following conditions:

$$\begin{aligned} P(A_1, A_2) &= P(A_1)P(A_2) \\ P(A_1, A_3) &= P(A_1)P(A_3) \\ P(A_2, A_3) &= P(A_2)P(A_3) \\ P(A_1, A_2, A_3) &= P(A_1)P(A_2)P(A_3) \end{aligned} \tag{2-1-17}$$

In the general case, the events $A_i$, $i = 1, 2, \ldots, n$, are statistically independent provided that the probabilities of the joint events taken $2, 3, 4, \ldots$, and $n$ at a time factor into the product of the probabilities of the individual events.

### 2-1-1 Random Variables, Probability Distributions, and Probability Densities

Given an experiment having a sample space $S$ and elements $s \in S$, we define a function $X(s)$ whose domain is $S$ and whose range is a set of numbers on the real line. The function $X(s)$ is called a *random variable.* For example, if we flip a coin the possible outcomes are head (H) and tail (T), so $S$ contains two points labeled H and T. Suppose we define a function $X(s)$ such that

$$X(s) = \begin{cases} 1 & (s = H) \\ -1 & (s = T) \end{cases} \tag{2-1-18}$$

Thus we have mapped the two possible outcomes of the coin-flipping experiment into the two points ($\pm 1$) on the real line. Another experiment is the toss of a die with possible outcomes $S = \{1, 2, 3, 4, 5, 6\}$. A random variable defined on this sample space may be $X(s) = s$, in which case the outcomes of the experiment are mapped into the integers $1, \ldots, 6$, or, perhaps, $X(s) = s^2$, in which case the possible outcomes are mapped into the integers $\{1, 4, 9, 16, 25, 36\}$. These are examples of discrete random variables.

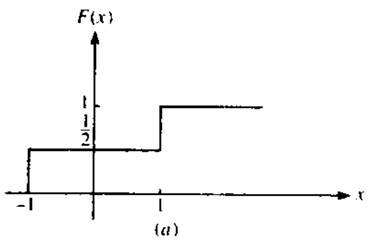
Although we have used as examples experiments that have a finite set of possible outcomes, there are many physical systems (experiments) that generate continuous outputs (outcomes). For example, the noise voltage generated by an electronic amplifier has a continuous amplitude. Consequently, the sample space $S$ of voltage amplitudes $v \in S$ is continuous and so is the mapping $X(v) = v$. In such a case, the random variable† $X$ is said to be a *continuous random variable.*

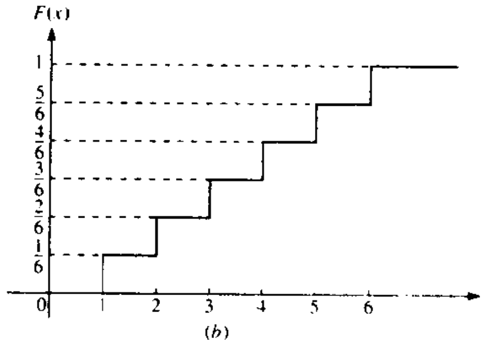
Given a random variable $X$, let us consider the event $\{X \leq x\}$ where $x$ is any real number in the interval $(-\infty, \infty)$. We write the probability of this event as $P(X \leq x)$ and denote it simply by $F(x)$, i.e.,

$$F(x) = P(X \leq x) \quad (-\infty < x < \infty) \tag{2-1-19}$$

The function $F(x)$ is called the *probability distribution function* of the random variable $X$. It is also called the *cumulative distribution function* (cdf). Since $F(x)$ is a probability, its range is limited to the interval $0 \leq F(x) \leq 1$. In fact, $F(-\infty) = 0$ and $F(\infty) = 1$. For example, the discrete random variable generated by flipping a fair coin and defined by (2-1-18) has the cdf shown in Fig. 2-1-1($a$). There are two discontinuities or jumps in $F(x)$, one at $x = -1$ and one at $x = 1$. Similarly, the random variable $X(s) = s$ generated by tossing a fair die has the cdf shown in Fig. 2-1-1($b$). In this case $F(x)$ has six jumps, one at each of the points $x = 1, \ldots, 6$.

**FIGURE 2-1-1** Examples of the cumulative distribution functions of two discrete random variables.




† The random variable $X(s)$ will be written simply as $X$.

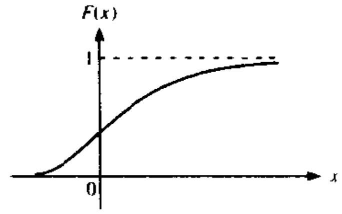



**FIGURE 2-1-2** An example of the cumulative distribution function of a continuous random variable.

The cdf of a continuous random variable typically appears as shown in Fig. 2-1-2. This is a smooth, nondecreasing function of $x$. In some practical problems, we may also encounter a random variable of a mixed type. The cdf of such a random variable is a smooth, nondecreasing function in certain parts of the real line and contains jumps at a number of discrete values of $x$. An example of such a cdf is illustrated in Fig. 2-1-3.

The derivative of the cdf $F(x)$, denoted as $p(x)$, is called the *probability density function* (pdf) of the random variable $X$. Thus, we have

$$p(x) = \frac{dF(x)}{dx} \quad (-\infty < x < \infty) \tag{2-1-20}$$

or, equivalently

$$F(x) = \int_{-\infty}^{x} p(u)\, du \quad (-\infty < x < \infty) \tag{2-1-21}$$

Since $F(x)$ is a nondecreasing function, it follows that $p(x) \geq 0$. When the random variable is discrete or of a mixed type, the pdf contains impulses at the points of discontinuity of $F(x)$. In such cases, the discrete part of $p(x)$ may be expressed as

$$p(x) = \sum_{i=1}^{n} P(X = x_i)\, \delta(x - x_i) \tag{2-1-22}$$

where $x_i$, $i = 1, 2, \ldots, n$, are the possible discrete values of the random

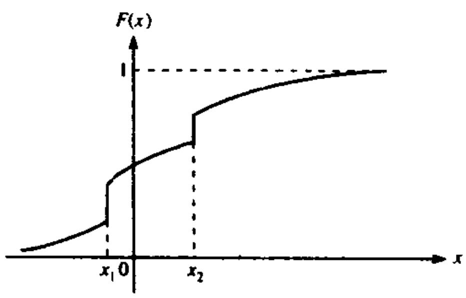



**FIGURE 2-1-3** An example of the cumulative distribution function of a random variable of a mixed type.

variable; $P(X=x_i)$, $i=1, 2, \ldots, n$, are the probabilities, and $\delta(x)$ denotes an impulse at $x=0$.

Often we are faced with the problem of determining the probability that a random variable $X$ falls in an interval $(x_1, x_2)$, where $x_2 > x_1$. To determine the probability of this event, let us begin with the event $\{X \leq x_2\}$. The event can always be expressed as the union of two mutually exclusive events $\{X \leq x_1\}$ and $\{x_1 < X \leq x_2\}$. Hence the probability of the event $\{X \leq x_2\}$ can be expressed as the sum of the probabilities of the mutually exclusive events. Thus we have

$$P(X \leq x_2) = P(X \leq x_1) + P(x_1 < X \leq x_2)$$

$$F(x_2) = F(x_1) + P(x_1 < X \leq x_2)$$

or, equivalently,

$$\begin{aligned} P(x_1 < X \leq x_2) &= F(x_2) - F(x_1) \\ &= \int_{x_1}^{x_2} p(x)\,dx \end{aligned} \tag{2-1-23}$$

In other words, the probability of the event $\{x_1 < X \leq x_2\}$ is simply the area under the pdf in the range $x_1 < X \leq x_2$.

**Multiple Random Variables, Joint Probability Distributions, and Joint Probability Densities** In dealing with combined experiments or repeated trials of a single experiment, we encounter multiple random variables and their cdfs and pdfs. Multiple random variables are basically multidimensional functions defined on a sample space of a combined experiment. Let us begin with two random variables $X_1$ and $X_2$, each of which may be continuous, discrete, or mixed. The joint cumulative distribution function (joint cdf) for the two random variables is defined as

$$\begin{aligned} F(x_1, x_2) &= P(X_1 \leq x_1, X_2 \leq x_2) \\ &= \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} p(u_1, u_2)\,du_1\,du_2 \end{aligned} \tag{2-1-24}$$

where $p(x_1, x_2)$ is the joint probability density function (joint pdf). The latter may also be expressed in the form

$$p(x_1, x_2) = \frac{\partial^2}{\partial x_1\,\partial x_2} F(x_1, x_2) \tag{2-1-25}$$

When the joint pdf $p(x_1, x_2)$ is integrated over one of the variables, we obtain the pdf of the other variable. That is,

$$\begin{aligned} \int_{-\infty}^{\infty} p(x_1, x_2)\,dx_1 &= p(x_2) \\ \int_{-\infty}^{\infty} p(x_1, x_2)\,dx_2 &= p(x_1) \end{aligned} \tag{2-1-26}$$

The pdfs $p(x_1)$ and $p(x_2)$ obtained from integrating over one of the variables are called *marginal pdf*s. Furthermore, if $p(x_1, x_2)$ is integrated over both variables, we obtain

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x_1, x_2)\, dx_1\, dx_2 = F(\infty, \infty) = 1 \tag{2-1-27}$$

We also note that $F(-\infty, -\infty) = F(-\infty, x_2) = F(x_1, -\infty) = 0$.

The generalization of the above expressions to multidimensional random variables is straightforward. Suppose that $X_i$, $i = 1, 2, \ldots, n$, are random variables with a joint cdf defined as

$$\begin{aligned} F(x_1, x_2, \ldots, x_n) &= P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n) \\ &= \int_{-\infty}^{x_1}\int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_n} p(u_1, u_2, \ldots, u_n)\, du_1\, du_2 \cdots du_n \end{aligned} \tag{2-1-28}$$

where $p(x_1, x_2, \ldots, x_n)$ is the joint pdf. By taking the partial derivatives of $F(x_1, x_2, \ldots, x_n)$ given by (2-1-28), we obtain

$$p(x_1, x_2, \ldots, x_n) = \frac{\partial^n}{\partial x_1\, \partial x_2 \cdots \partial x_n} F(x_1, x_2, \ldots, x_n) \tag{2-1-29}$$

Any number of variables in $p(x_1, x_2, \ldots, x_n)$ can be eliminated by integrating over these variables. For example, integration over $x_2$ and $x_3$ yields

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x_1, x_2, x_3, \ldots, x_n)\, dx_2\, dx_3 = p(x_1, x_4, \ldots, x_n) \tag{2-1-30}$$

It also follows that $F(x_1, \infty, \infty, x_4, \ldots, x_n) = F(x_1, x_4, x_5, \ldots, x_n)$ and

$$F(x_1, -\infty, -\infty, x_4, \ldots, x_n) = 0.$$

**Conditional Probability Distribution Functions** Let us consider two random variables $X_1$ and $X_2$ with joint pdf $p(x_1, x_2)$. Suppose that we wish to determine the probability that the random variable $X_1 \leq x_1$ conditioned on

$$x_2 - \Delta x_2 < X_2 \leq x_2$$

where $\Delta x_2$ is some positive increment. That is, we wish to determine the probability of the event $(X_1 \leq x_1 \mid x_2 - \Delta x_2 < X_2 \leq x_2)$. Using the relations established earlier for the conditional probability of an event, the probability of the event $(X_1 \leq x_1 \mid x_2 - \Delta x_2 < X_2 \leq x_2)$ can be expressed as the probability of the joint event $(X_1 \le x_1, x_2 - \Delta x_2 < X_2 \le x_2)$ divided by the probability of the event $(x_2 - \Delta x_2 < X_2 \le x_2)$. Thus

$$P(X_1 \le x_1 \mid x_2 - \Delta x_2 < X_2 \le x_2) = \frac{\int_{-\infty}^{x_1} \int_{x_2-\Delta x_2}^{x_2} p(u_1, u_2)\, du_1\, du_2}{\int_{x_2-\Delta x_2}^{x_2} p(u_2)\, du_2}$$

$$= \frac{F(x_1, x_2) - F(x_1, x_2 - \Delta x_2)}{F(x_2) - F(x_2 - \Delta x_2)} \tag{2-1-31}$$

Assuming that the pdfs $p(x_1, x_2)$ and $p(x_2)$ are continuous functions over the interval $(x_2 - \Delta x_2, x_2)$, we may divide both numerator and denominator in (2-1-31) by $\Delta x_2$ and take the limit as $\Delta x_2 \to 0$. Thus we obtain

$$P(X_1 \le x_1 \mid X_2 = x_2) \equiv F(x_1 \mid x_2) = \frac{\partial F(x_1, x_2)/\partial x_2}{\partial F(x_2)/\partial x_2}$$

$$= \frac{\partial[\int_{-\infty}^{x_1} \int_{-\infty}^{x_2} p(u_1, u_2)\, du_1 du_2]/\partial x_2}{\partial[\int_{-\infty}^{x_2} p(u_2)\, du_2]/\partial x_2}$$

$$= \frac{\int_{-\infty}^{x_1} p(u_1, x_2)\, du_1}{p(x_2)} \tag{2-1-32}$$

which is the conditional cdf of the random variable $X_1$ given the random variable $X_2$. We observe that $F(-\infty \mid x_2) = 0$ and $F(\infty \mid x_2) = 1$. By differentiating (2-1-32) with respect to $x_1$, we obtain the corresponding pdf $p(x_1 \mid x_2)$ in the form

$$p(x_1 \mid x_2) = \frac{p(x_1, x_2)}{p(x_2)} \tag{2-1-33}$$

Alternatively, we may express the joint pdf $p(x_1, x_2)$ in terms of the conditional pdfs, $p(x_1 \mid x_2)$ or $p(x_2 \mid x_1)$, as

$$p(x_1, x_2) = p(x_1 \mid x_2) p(x_2)$$

$$= p(x_2 \mid x_1) p(x_1) \tag{2-1-34}$$

The extension of the relations given above to multidimensional random variables is also easily accomplished. Beginning with the joint pdf of the random variables $X_i$, $i = 1, 2, \ldots, n$, we may write

$$p(x_1, x_2, \ldots, x_n) = p(x_1, x_2, \ldots, x_k \mid x_{k+1}, \ldots, x_n) p(x_{k+1}, \ldots, x_n) \tag{2-1-35}$$

where $k$ is any integer in the range $1 < k < n$. The joint conditional cdf corresponding to the pdf $p(x_1, x_2, \ldots, x_k \mid x_{k+1}, \ldots, x_n)$ is

$$F(x_1, x_2, \ldots, x_k \mid x_{k+1}, \ldots, x_n)$$

$$= \frac{\int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_k} p(u_1, u_2, \ldots, u_k, x_{k+1}, \ldots, x_n)\, du_1\, du_2 \cdots du_k}{p(x_{k+1}, \ldots, x_n)} \tag{2-1-36}$$

This conditional cdf satisfies the properties previously established for these functions, such as

$$F(\infty, x_2, \ldots, x_k \mid x_{k+1}, \ldots, x_n) = F(x_2, x_3, \ldots, x_k \mid x_{k+1}, \ldots, x_n)$$
$$F(-\infty, x_2, \ldots, x_k \mid x_{k+1}, \ldots, x_n) = 0$$

**Statistically Independent Random Variables.** We have already defined statistical independence of two or more events of a sample space $S$. The concept of statistical independence can be extended to random variables defined on a sample space generated by a combined experiment or by repeated trials of a single experiment. If the experiments result in mutually exclusive outcomes, the probability of an outcome in one experiment is independent of an outcome in any other experiment. That is, the joint probability of the outcomes factors into a product of the probabilities corresponding to each outcome. Consequently, the random variables corresponding to the outcomes in these experiments are independent in the sense that their joint pdf factors into a product of marginal pdfs. Hence the multidimensional random variables are statistically independent if and only if

$$F(x_1, x_2, \ldots, x_n) = F(x_1)F(x_2)\cdots F(x_n) \tag{2-1-37}$$

or, alternatively,

$$p(x_1, x_2, \ldots, x_n) = p(x_1)p(x_2)\cdots p(x_n) \tag{2-1-38}$$

### 2-1-2 Functions of Random Variables

A problem that arises frequently in practical applications of probability is the following. Given a random variable $X$, which is characterized by its pdf $p(x)$, determine the pdf of the random variable $Y = g(X)$, where $g(X)$ is some given function of $X$. When the mapping $g$ from $X$ to $Y$ is one-to-one, the determination of $p(y)$ is relatively straightforward. However, when the mapping is not one-to-one, as is the case, for example, when $Y = X^2$, we must be very careful in our derivation of $p(y)$.

**Example 2-1-1**

Consider the random variable $Y$ defined as

$$Y = aX + b \tag{2-1-39}$$

where $a$ and $b$ are constants. We assume that $a > 0$. If $a < 0$, the approach is similar (see Problem 2-3). We note that this mapping, illustrated in Fig. 2-1-4($a$) is linear and monotonic. Let $F_X(x)$ and $F_Y(y)$ denote the cdfs for $X$ and $Y$, respectively.† Then

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(aX + b \leq y) = P\left(X \leq \frac{y-b}{a}\right) \\ &= \int_{-\infty}^{(y-b)/a} p_X(x)\,dx = F_X\left(\frac{y-b}{a}\right) \end{aligned} \tag{2-1-40}$$

† To avoid confusion in changing variables, subscripts are used in the respective pdfs and cdfs.

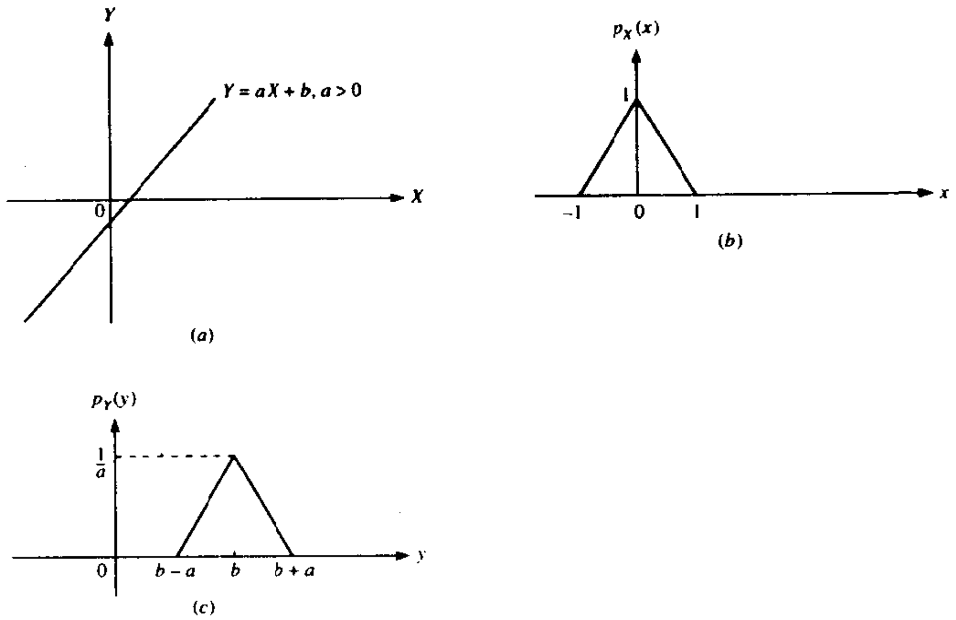



**FIGURE 2-1-4** A linear transformation of a random variable $X$ and an example of the corresponding pdfs of $X$ and $Y$.

By differentiating (2-1-40) with respect to $y$, we obtain the relationship between the respective pdfs. It is

$$p_Y(y) = \frac{1}{a} p_X\left(\frac{y-b}{a}\right) \tag{2-1-41}$$

Thus (2-1-40) and (2-1-41) specify the cdf and pdf of the random variable $Y$ in terms of the cdf and pdf of the random variable $X$ for the linear transformation in (2-1-39). To illustrate this mapping for a specific pdf $p_X(x)$, consider the one shown in Fig. 2-1-4($b$). The pdf $p_Y(y)$ that results from the mapping in (2-1-39) is shown in Fig. 2-1-4($c$).

**Example 2-1-2**

Consider the random variable $Y$ defined as

$$Y = aX^3 + b, \qquad a > 0 \tag{2-1-42}$$

As in Example 2-1-1, the mapping between $X$ and $Y$ is one-to-one. Hence

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(aX^3 + b \leq y) \\ &= P\left[X \leq \left(\frac{y-b}{a}\right)^{1/3}\right] = F_X\left[\left(\frac{y-b}{a}\right)^{1/3}\right] \end{aligned} \tag{2-1-43}$$

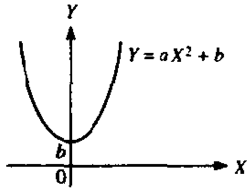



FIGURE 2-1-5 A quadratic transformation of the random variable $X$.

Differentiation of (2-1-43) with respect to $y$ yields the desired relationship between the two pdfs as

$$p_Y(y) = \frac{1}{3a[(y-b)/a]^{2/3}} p_X\left[\left(\frac{y-b}{a}\right)^{1/3}\right] \tag{2-1-44}$$

**Example 2-1-3**

The random variable $Y$ is defined as

$$Y = aX^2 + b, \qquad a > 0 \tag{2-1-45}$$

In contrast to Examples 2-1-1 and 2-1-2, the mapping between $X$ and $Y$, illustrated in Fig. 2-1-5, is not one-to-one. To determine the cdf of $Y$, we observe that

$$\begin{aligned} F_Y(y) &= P(Y \le y) = P(aX^2 + b \le y) \\ &= P\left(|X| \le \sqrt{\frac{y-b}{a}}\right) \end{aligned}$$

Hence

$$F_Y(y) = F_X\left(\sqrt{\frac{y-b}{a}}\right) - F_X\left(-\sqrt{\frac{y-b}{a}}\right) \tag{2-1-46}$$

Differentiating (2-1-46) with respect to $y$, we obtain the pdf of $Y$ in terms of the pdf of $X$ in the form

$$p_Y(y) = \frac{p_X[\sqrt{(y-b)/a}]}{2a\sqrt{[(y-b)/a]}} + \frac{p_X[-\sqrt{(y-b)/a}]}{2a\sqrt{[(y-b)/a]}} \tag{2-1-47}$$

In Example 2-1-3, we observe that the equation $g(x) = ax^2 + b = y$ has two real solutions,

$$x_1 = \sqrt{\frac{y-b}{a}}$$

$$x_2 = -\sqrt{\frac{y-b}{a}}$$

and that $p_Y(y)$ consists of two terms corresponding to these two solutions. That is,

$$p_Y(y) = \frac{p_X[x_1 = \sqrt{(y-b)/a}]}{|g'[x_1 = \sqrt{(y-b)/a}]|} + \frac{p_X[x_2 = -\sqrt{(y-b)/a}]}{|g'[x_2 = -\sqrt{(y-b)/a}]|} \tag{2-1-48}$$

where $g'(x)$ denotes the first derivative of $g(x)$.

In the general case, suppose that $x_1, x_2, \ldots, x_n$ are the real roots of the equation $g(x) = y$. Then the pdf of the random variable $Y = g(X)$ may be expressed as

$$p_Y(y) = \sum_{i=1}^{n} \frac{p_X(x_i)}{|g'(x_i)|} \tag{2-1-49}$$

where the roots $x_i$, $i = 1, 2, \ldots, n$, are functions of $y$.

Now let us consider functions of multidimensional random variables. Suppose that $X_i$, $i = 1, 2, \ldots, n$, are random variables with joint pdf $p_X(x_1, x_2, \ldots, x_n)$, and let $Y_i$, $i = 1, 2, \ldots, n$, be another set of $n$ random variables related to the $X_i$ by the functions

$$Y_i = g_i(X_1, X_2, \ldots, X_n), \qquad i = 1, 2, \ldots, n \tag{2-1-50}$$

We assume that the $g_i(X_1, X_2, \ldots, X_n)$, $i = 1, 2, \ldots, n$, are single-valued functions with continuous partial derivatives and invertible. By "invertible" we mean that the $X_i$, $i = 1, 2, \ldots, n$, can be expressed as functions of $Y_i$, $i = 1, 2, \ldots, n$, in the form

$$X_i = g_i^{-1}(Y_1, Y_2, \ldots, Y_n), \qquad i = 1, 2, \ldots, n \tag{2-1-51}$$

where the inverse functions are also assumed to be single-valued with continuous partial derivatives. The problem is to determine the joint pdf of $Y_i$, $i = 1, 2, \ldots, n$, denoted by $p_Y(y_1, y_2, \ldots, y_n)$, given the joint pdf $p_X(x_1, x_2, \ldots, x_n)$.

To determine the desired relation, let $R_X$ be the region in the $n$-dimensional space of the random variables $X_i$, $i = 1, 2, \ldots, n$, and let $R_Y$ be the (one-to-one) mapping of $R_X$ defined by the functions $Y_i = g_i(X_1, X_2, \ldots, X_n)$. Clearly,

$$\int\int_{R_Y} \cdots \int p_Y(y_1, y_2, \ldots, y_n)\, dy_1\, dy_2 \cdots dy_n$$

$$= \int\int_{R_X} \cdots \int p_X(x_1, x_2, \ldots, x_n)\, dx_1\, dx_2 \cdots dx_n \tag{2-1-52}$$

By making a change in variables in the multiple integral on the right-hand side of (2-1-52) with the substitution

$$x_i = g_i^{-1}(y_1, y_2, \ldots, y_n) \equiv g_i^{-1}, \qquad i = 1, 2, \ldots, n$$

we obtain

$$\iint_{R_Y} \cdots \int p_Y(y_1, y_2, \ldots, y_n)\, dy_1\, dy_2 \cdots dy_n$$
$$= \iint_{R_Y} \cdots \int p_X(x_1 = g_1^{-1}, x_2 = g_2^{-1}, \ldots, x_n = g_n^{-1})\, |J|\, dy_1\, dy_2 \cdots dy_n \tag{2-1-53}$$

where $J$ denotes the jacobian of the transformation, defined by the determinant

$$J = \begin{vmatrix} \dfrac{\partial g_1^{-1}}{\partial y_1} & \dfrac{\partial g_2^{-1}}{\partial y_1} & \cdots & \dfrac{\partial g_n^{-1}}{\partial y_1} \\ \vdots & \vdots & & \vdots \\ \dfrac{\partial g_1^{-1}}{\partial y_n} & \dfrac{\partial g_2^{-1}}{\partial y_n} & \cdots & \dfrac{\partial g_n^{-1}}{\partial y_n} \end{vmatrix} \tag{2-1-54}$$

Consequently, the desired relation for the joint pdf of the $Y_i$, $i = 1, 2, \ldots, n$, is

$$p_Y(y_1, y_2, \ldots, y_n) = p_X(x_1 = g_1^{-1}, x_2 = g_2^{-1}, \ldots, x_n = g_n^{-1})\, |J| \tag{2-1-55}$$

**Example 2-1-4**

An important functional relation between two sets of $n$-dimensional random variables that frequently arises in practice is the linear transformation

$$Y_i = \sum_{j=1}^{n} a_{ij} X_j, \qquad i = 1, 2, \ldots, n \tag{2-1-56}$$

where the $\{a_{ij}\}$ are constants. It is convenient to employ the matrix form for the transformation, which is

$$\mathbf{Y} = \mathbf{AX} \tag{2-1-57}$$

where $\mathbf{X}$ and $\mathbf{Y}$ are $n$-dimensional vectors and $\mathbf{A}$ is an $n \times n$ matrix. We assume that $\mathbf{A}$ is nonsingular. Then $\mathbf{A}$ is invertible and, hence,

$$\mathbf{X} = \mathbf{A}^{-1}\mathbf{Y} \tag{2-1-58}$$

Equivalently, we have

$$X_i = \sum_{j=1}^{n} b_{ij} Y_j, \qquad i = 1, 2, \ldots, n \tag{2-1-59}$$

where $\{b_{ij}\}$ are the elements of the inverse matrix $\mathbf{A}^{-1}$. The jacobian of this transformation is $J = 1/\det \mathbf{A}$. Hence

$$p_Y(y_1, y_2, \ldots, y_n)$$
$$= p_X\left(x_1 = \sum_{j=1}^{n} b_{1j} y_j, x_2 = \sum_{j=1}^{n} b_{2j} y_j, \ldots, x_n = \sum_{j=1}^{n} b_{nj} y_j\right) \frac{1}{|\det \mathbf{A}|} \tag{2-1-60}$$

### 2-1-3 Statistical Averages of Random Variables

Averages play an important role in the characterization of the outcomes of experiments and the random variables defined on the sample space of the experiments. Of particular interest are the first and second moments of a single random variable and the joint moments, such as the correlation and covariance, between any pair of random variables in a multidimensional set of random variables. Also of great importance are the characteristic function for a single random variable and the joint characteristic function for a multidimensional set of random variables. This section is devoted to the definition of these important statistical averages.

First we consider a single random variable $X$ characterized by its pdf $p(x)$. The *mean* or *expected value* of $X$ is defined as

$$E(X) \equiv m_x = \int_{-\infty}^{\infty} x p(x)\, dx \tag{2-1-61}$$

where $E(\ )$ denotes expectation (statistical averaging). This is the first moment of the random variable $X$. In general, the $n$th moment is defined as

$$E(X^n) = \int_{-\infty}^{\infty} x^n p(x)\, dx \tag{2-1-62}$$

Now, suppose that we define a random variable $Y = g(X)$, where $g(X)$ is some arbitrary function of the random variable $X$. The expected value of $Y$ is

$$E(Y) = E[g(X)] = \int_{-\infty}^{\infty} g(x) p(x)\, dx \tag{2-1-63}$$

In particular, if $Y = (X - m_x)^n$ where $m_x$ is the mean value of $X$, then

$$E(Y) = E[(X - m_x)^n] = \int_{-\infty}^{\infty} (x - m_x)^n p(x)\, dx \tag{2-1-64}$$

This expected value is called the $n$th *central moment* of the random variable $X$, because it is a moment taken relative to the mean. When $n = 2$, the central moment is called the *variance* of the random variable and denoted as $\sigma_x^2$. That is,

$$\sigma_x^2 = \int_{-\infty}^{\infty} (x - m_x)^2 p(x)\, dx \tag{2-1-65}$$

This parameter provides a measure of the dispersion of the random variable $X$. By expanding the term $(x - m_x)^2$ in the integral of (2-1-65) and noting that the expected value of a constant is equal to the constant, we obtain the expression that relates the variance to the first and second moments, namely,

$$\begin{aligned} \sigma_x^2 &= E(X^2) - [E(X)]^2 \\ &= E(X^2) - m_x^2 \end{aligned} \tag{2-1-66}$$

In the case of two random variables, $X_1$ and $X_2$, with joint pdf $p(x_1, x_2)$, we define the *joint moment* as

$$E(X_1^k X_2^n) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1^k x_2^n p(x_1, x_2)\, dx_1\, dx_2 \tag{2-1-67}$$

and the *joint central moment* as

$$\begin{aligned} &E[(X_1 - m_1)^k (X_2 - m_2)^n] \\ &\qquad = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_1 - m_1)^k (x_2 - m_2)^n p(x_1, x_2)\, dx_1\, dx_2 \end{aligned} \tag{2-1-68}$$

where $m_i = E(X_i)$. Of particular importance to us are the joint moment and joint central moment corresponding to $k = n = 1$. These joint moments are called the *correlation* and the *covariance* of the random variables $X_1$ and $X_2$, respectively.

In considering multidimensional random variables, we can define joint moments of any order. However, the moments that are most useful in practical applications are the correlations and covariances between pairs of random variables. To elaborate, suppose that $X_i$, $i = 1, 2, \ldots, n$, are random variables with joint pdf $p(x_1, x_2, \ldots, x_n)$. Let $p(x_i, x_j)$ be the joint pdf of the random variables $X_i$ and $X_j$. Then the correlation between $X_i$ and $X_j$ is given by the joint moment

$$E(X_i X_j) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_i x_j p(x_i, x_j)\, dx_i\, dx_j \tag{2-1-69}$$

and the covariance of $X_i$ and $X_j$ is

$$\begin{aligned} \mu_{ij} &\equiv E[(X_i - m_i)(X_j - m_j)] \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_i - m_i)(x_j - m_j) p(x_i, x_j)\, dx_i\, dx_j \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_i x_j p(x_i, x_j)\, dx_i\, dx_j - m_i m_j \\ &= E(X_i X_j) - m_i m_j \end{aligned} \tag{2-1-70}$$

The $n \times n$ matrix with elements $\mu_{ij}$ is called the *covariance matrix* of the random variables $X_i$, $i = 1, 2, \ldots, n$. We shall encounter the covariance matrix in our discussion of jointly gaussian random variables in Section 2-1-4.

Two random variables are said to be *uncorrelated* if $E(X_i X_j) = E(X_i)E(X_j) = m_i m_j$. In that case, the covariance $\mu_{ij} = 0$. We note that when $X_i$ and $X_j$ are statistically independent, they are also uncorrelated. However, if $X_i$ and $X_j$ are uncorrelated, they are not necessarily statistically independent.

Two random variables are said to be *orthogonal* if $E(X_i X_j) = 0$. We note that this condition holds when $X_i$ and $X_j$ are uncorrelated and either one or both of the random variables have zero mean.

**Characteristic Functions** The *characteristic function* of a random variable $X$ is defined as the statistical average

$$E(e^{jvX}) \equiv \psi(jv) = \int_{-\infty}^{\infty} e^{jvx} p(x)\, dx \tag{2-1-71}$$

where the variable $v$ is real and $j = \sqrt{-1}$. We note that $\psi(jv)$ may be described as the Fourier transform† of the pdf $p(x)$. Hence the inverse Fourier transform is

$$p(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \psi(jv) e^{-jvx}\, dv \tag{2-1-72}$$

One useful property of the characteristic function is its relation to the moments of the random variable. We note that the first derivative of (2-1-71) with respect to $v$ yields

$$\frac{d\psi(jv)}{dv} = j \int_{-\infty}^{\infty} x e^{jvx} p(x)\, dx$$

By evaluating the derivative at $v = 0$, we obtain the first moment (mean)

$$E(X) = m_x = -j \left. \frac{d\psi(jv)}{dv} \right|_{v=0} \tag{2-1-73}$$

The differentiation process can be repeated, so that the $n$th derivative of $\psi(jv)$ evaluated at $v = 0$ yields the $n$th moment

$$E(X^n) = (-j)^n \left. \frac{d^n \psi(jv)}{dv^n} \right|_{v=0} \tag{2-1-74}$$

Thus the moments of a random variable can be determined from the characteristic function. On the other hand, suppose that the characteristic function can be expanded in a Taylor series about the point $v = 0$. That is,

$$\psi(jv) = \sum_{n=0}^{\infty} \left[ \frac{d^n \psi(jv)}{dv^n} \right]_{v=0} \frac{v^n}{n!} \tag{2-1-75}$$

Using the relation in (2-1-74) to eliminate the derivative in (2-1-75), we obtain

† Usually the Fourier transform of a function $g(u)$ is defined as $G(v) = \int_{-\infty}^{\infty} g(u) e^{-juv}\, du$, which differs from (2-1-71) by the negative sign in the exponential. This is a trivial difference, however, so we call the integral in (2-1-71) a Fourier transform.

an expression for the characteristic function in terms of its moments in the form

$$\psi(jv) = \sum_{n=0}^{\infty} E(X^n) \frac{(jv)^n}{n!} \tag{2-1-76}$$

The characteristic function provides a simple method for determining the pdf of a sum of statistically independent random variables. To illustrate this point, let $X_i$, $i = 1, 2, \ldots, n$, be a set of $n$ statistically independent random variables and let

$$Y = \sum_{i=1}^{n} X_i \tag{2-1-77}$$

The problem is to determine the pdf of $Y$. We shall determine the pdf of $Y$ by first finding its characteristic function and then computing the inverse Fourier transform. Thus

$$\begin{aligned}
\psi_Y(jv) &= E(e^{jvY}) \\
&= E\left[\exp\left(jv \sum_{i=1}^{n} X_i\right)\right] \\
&= E\left[\prod_{i=1}^{n} (e^{jvX_i})\right] \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(\prod_{i=1}^{n} e^{jvx_i}\right) p(x_1, x_2, \ldots, x_n)\, dx_1\, dx_2 \cdots dx_n
\end{aligned} \tag{2-1-78}$$

Since the random variables are statistically independent, $p(x_1, x_2, \ldots, x_n) = p(x_1)p(x_2)\cdots p(x_n)$, and, hence, the $n$th-order integral in (2-1-78) reduces to a product of $n$ single integrals, each corresponding to the characteristic function of one of the $X_i$. Hence,

$$\psi_Y(jv) = \prod_{i=1}^{n} \psi_{X_i}(jv) \tag{2-1-79}$$

If, in addition to their statistical independence, the $X_i$ are identically distributed then all the $\psi_{X_i}(jv)$ are identical. Consequently,

$$\psi_Y(jv) = [\psi_X(jv)]^n \tag{2-1-80}$$

Finally, the pdf of $Y$ is determined from the inverse Fourier transform of $\psi_Y(jv)$, given by (2-1-72).

Since the characteristic function of the sum of $n$ statistically independent random variables is equal to the product of the characteristic functions of the individual random variables $X_i$, $i = 1, 2, \ldots, n$, it follows that, in the transform domain, the pdf of $Y$ is the $n$-fold convolution of the pdfs of the $X_i$. Usually the $n$-fold convolution is more difficult to perform than the characteristic function method described above in determining the pdf of $Y$.

When working with $n$-dimensional random variables, it is appropriate to define and $n$-dimensional Fourier transform of the joint pdf. In particular, if

$X_i$, $i = 1, 2, \ldots, n$, are random variables with pdf $p(x_1, x_2, \ldots, x_n)$, the *n-dimensional characteristic function* is defined as

$$\begin{aligned}\psi(jv_1, jv_2, \ldots, jv_n) &= E\left[\exp\left(j\sum_{i=1}^{n} v_i X_i\right)\right] \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left(j\sum_{i=1}^{n} v_i x_i\right) p(x_1, x_2, \ldots, x_n)\, dx_1\, dx_2 \cdots dx_n \qquad (2\text{-}1\text{-}81)\end{aligned}$$

Of special interest is the two-dimensional characteristic function

$$\psi(jv_1, jv_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{j(v_1x_1 + v_2x_2)} p(x_1, x_2)\, dx_1\, dx_2 \qquad (2\text{-}1\text{-}82)$$

We observe that the partial derivatives of $\psi(jv_1, jv_2)$ with respect to $v_1$ and $v_2$ can be used to generate the joint moments. For example, it is easy to show that

$$E(X_1X_2) = -\left.\frac{\partial^2 \psi(jv_1, jv_2)}{\partial v_1\, \partial v_2}\right|_{v_1 = v_2 = 0} \qquad (2\text{-}1\text{-}83)$$

Higher-order moments are generated in a straightforward manner.

## 2-1-4 Some Useful Probability Distributions

In subsequent chapters, we shall encounter several different types of random variables. In this section we list these frequently encountered random variables, their pdfs, their cdfs, and their moments. We begin with the binomial distribution, which is the distribution of a discrete random variable, and then we present the distributions of several continuous random variables.

**Binomial Distribution** Let $X$ be a discrete random variable that has two possible values, say $X = 1$ or $X = 0$, with probabilities $p$ and $1 - p$, respectively. The pdf of $X$ is shown in Fig. 2-1-6. Now, suppose that

$$Y = \sum_{i=1}^{n} X_i$$

where the $X_i$, $i = 1, 2, \ldots, n$, are statistically independent and identically

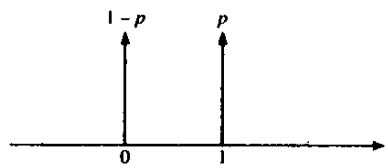



**FIGURE 2-1-6** The probability distribution function of $X$.

distributed random variables with the pdf shown in Fig. 2-1-6. What is the probability distribution function of $Y$?

To answer this question, we observe that the range of $Y$ is the set of integers from 0 to $n$. The probability that $Y=0$ is simply the probability that all the $X_i = 0$. Since the $X_i$ are statistically independent,

$$P(Y=0)=(1-p)^n$$

The probability that $Y=1$ is simply the probability that one $X_i = 1$ and the rest of the $X_i = 0$. Since this event can occur in $n$ different ways,

$$P(Y=1)=np(1-p)^{n-1}$$

To generalize, the probability that $Y=k$ is the probability that $k$ of the $X_i$ are equal to one and $n-k$ are equal to zero. Since there are

$$\binom{n}{k} \equiv \frac{n!}{k!\,(n-k)!} \tag{2-1-84}$$

different combinations that result in the event $\{Y=k\}$, it follows that

$$P(Y=k)=\binom{n}{k}p^k(1-p)^{n-k} \tag{2-1-85}$$

where $\binom{n}{k}$ is the binomial coefficient. Consequently, the pdf of $Y$ may be expressed as

$$\begin{aligned} p(y) &= \sum_{k=0}^{n} P(Y=k)\,\delta(y-k) \\ &= \sum_{k=0}^{n} \binom{n}{k} p^k(1-p)^{n-k}\,\delta(y-k) \end{aligned} \tag{2-1-86}$$

The cdf of $Y$ is

$$\begin{aligned} F(y) &= P(Y \leqslant y) \\ &= \sum_{k=0}^{[y]} \binom{n}{k} p^k(1-p)^{n-k} \end{aligned} \tag{2-1-87}$$

where $[y]$ denotes the largest integer $m$ such that $m \leqslant y$. The cdf in (2-1-87) characterizes a binomially distributed random variable.

The first two moments of $Y$ are

$$\begin{aligned} E(Y) &= np \\ E(Y^2) &= np(1-p)+n^2p^2 \\ \sigma^2 &= np(1-p) \end{aligned} \tag{2-1-88}$$

and the characteristic function is

$$\psi(jv)=(1-p+pe^{jv})^n \tag{2-1-89}$$

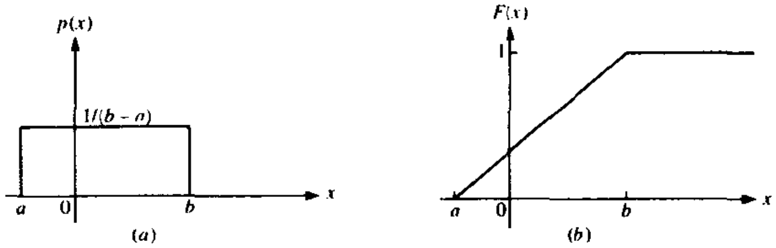



**FIGURE 2-1-7** The pdf and cdf of a uniformly distributed random variable.

**Uniform Distribution** The pdf and cdf of a uniformly distributed random variable $X$ are shown in Fig. 2-1-7. The first two moments of $X$ are

$$\begin{aligned} E(X) &= \tfrac{1}{2}(a+b) \\ E(X^2) &= \tfrac{1}{3}(a^2+b^2+ab) \\ \sigma^2 &= \tfrac{1}{12}(a-b)^2 \end{aligned} \tag{2-1-90}$$

and the characteristic function is

$$\psi(jv) = \frac{e^{jvb} - e^{jva}}{jv(b-a)} \tag{2-1-91}$$

**Gaussian (Normal) Distribution** The pdf of a gaussian or normally distributed random variable is

$$p(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-m_x)^2/2\sigma^2} \tag{2-1-92}$$

where $m_x$ is the mean and $\sigma^2$ is the variance of the random variable. The cdf is

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} p(u)\,du \\ &= \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{x} e^{-(u-m_x)^2/2\sigma^2}\,du \\ &= \frac{1}{2}\frac{2}{\sqrt{\pi}}\int_{-\infty}^{(x-m_x)/\sqrt{2}\sigma} e^{-t^2}\,dt \\ &= \tfrac{1}{2} + \tfrac{1}{2}\operatorname{erf}\left(\frac{x-m_x}{\sqrt{2}\,\sigma}\right) \end{aligned} \tag{2-1-93}$$

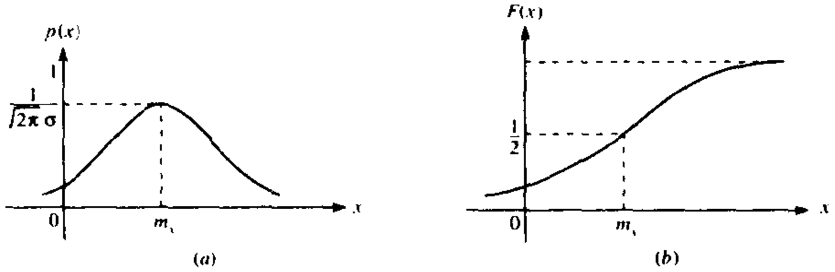



**FIGURE 2-1-8** The pdf and cdf of a gaussian-distributed random variable.

where erf $(x)$ denotes the error function, defined as

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\, dt \tag{2-1-94}$$

The pdf and cdf are illustrated in Fig. 2-1-8.

The cdf $F(x)$ may also be expressed in terms of the complementary error function. That is,

$$F(x) = 1 - \tfrac{1}{2} \operatorname{erfc}\left(\frac{x - m_x}{\sqrt{2}\,\sigma}\right)$$

where

$$\begin{aligned} \operatorname{erfc}(x) &= \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2}\, dt \\ &= 1 - \operatorname{erf}(x) \end{aligned} \tag{2-1-95}$$

We note that erf $(-x) = -$erf $(x)$, erfc $(-x) = 2 -$ erfc $(x)$, erf $(0) =$ erfc $(\infty) = 0$, and erf $(\infty) =$ erfc $(0) = 1$. For $x > m_x$, the complementary error function is proportional to the area under the tail of the gaussian pdf. For large values of $x$, the complementary error function erfc $(x)$ may be approximated by the asymptotic series

$$\operatorname{erfc}(x) = \frac{e^{-x^2}}{x\sqrt{\pi}}\left(1 - \frac{1}{2x^2} + \frac{1 \cdot 3}{2^2 x^4} - \frac{1 \cdot 3 \cdot 5}{2^3 x^6} + \cdots\right) \tag{2-1-96}$$

where the approximation error is less than the last term used.

The function that is frequently used for the area under the tail of the gaussian pdf is denoted by $Q(x)$ and defined as

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2}\, dt, \qquad x \geqslant 0 \tag{2-1-97}$$

By comparing (2-1-95) with (2-1-97), we find

$$Q(x) = \tfrac{1}{2} \operatorname{erfc}\left(\frac{x}{\sqrt{2}}\right) \tag{2-1-98}$$

The characteristic function of a gaussian random variable with mean $m_x$ and variance $\sigma^2$ is

$$\begin{aligned}\psi(jv) &= \int_{-\infty}^{\infty} e^{jvx}\left[\frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-m_x)^2/2\sigma^2}\right] dx \\ &= e^{jvm_x - (1/2)v^2\sigma^2}\end{aligned} \tag{2-1-99}$$

The central moments of a gaussian random variable are

$$E[(X-m_x)^k] \equiv \mu_k = \begin{cases} 1 \cdot 3 \cdots (k-1)\sigma^k & (\text{even } k) \\ 0 & (\text{odd } k) \end{cases} \tag{2-1-100}$$

and the ordinary moments may be expressed in terms of the central moments as

$$E(X^k) = \sum_{i=0}^{k} \binom{k}{i} m_x^i \mu_{k-i} \tag{2-1-101}$$

The sum of $n$ statistically independent gaussian random variables is also a gaussian random variable. To demonstrate this point, let

$$Y = \sum_{i=1}^{n} X_i \tag{2-1-102}$$

where the $X_i$, $i = 1, 2, \ldots, n$, are statistically independent gaussian random variables with means $m_i$ and variances $\sigma_i^2$. Using the result in (2-1-79), we find that the characteristic function of $Y$ is

$$\begin{aligned}\psi_Y(jv) &= \prod_{i=1}^{n} \psi_{X_i}(jv) \\ &= \prod_{i=1}^{n} e^{jvm_i - v^2\sigma_i^2/2} \\ &= e^{jvm_y - v^2\sigma_y^2/2}\end{aligned} \tag{2-1-103}$$

where

$$\begin{aligned} m_y &= \sum_{i=1}^{n} m_i \\ \sigma_y^2 &= \sum_{i=1}^{n} \sigma_i^2 \end{aligned} \tag{2-1-104}$$

Therefore, $Y$ is gaussian-distributed with mean $m_y$ and variance $\sigma_y^2$.

**Chi-Square Distribution** A chi-square-distributed random variable is related to a gaussian-distributed random variable in the sense that the former can be viewed as a transformation of the latter. To be specific, let $Y = X^2$, where $X$ is a gaussian random variable. Then $Y$ has a chi-square distribution. We distinguish between two types of chi-square distributions. The first is called a *central chi-square distribution* and is obtained when $X$ has zero mean. The second is called a *non-central chi-square distribution*, and is obtained when $X$ has a nonzero mean.

First we consider the central chi-square distribution. Let $X$ be gaussian-distributed with zero mean and variance $\sigma^2$. Since $Y = X^2$, the result given in (2-1-47) applies directly with $a = 1$ and $b = 0$. Thus we obtain the pdf of $Y$ in the form

$$p_Y(y) = \frac{1}{\sqrt{2\pi y}\,\sigma} e^{-y/2\sigma^2}, \quad y \geq 0 \tag{2-1-105}$$

The cdf of $Y$ is

$$\begin{aligned} F_Y(y) &= \int_0^y p_Y(u)\,du \\ &= \frac{1}{\sqrt{2\pi}\,\sigma} \int_0^y \frac{1}{\sqrt{u}} e^{-u/2\sigma^2}\,du \end{aligned} \tag{2-1-106}$$

which cannot be expressed in closed form. The characteristic function, however, can be determined in closed form. It is

$$\psi(jv) = \frac{1}{(1 - j2v\sigma^2)^{1/2}} \tag{2-1-107}$$

Now, suppose that the random variable $Y$ is defined as

$$Y = \sum_{i=1}^{n} X_i^2 \tag{2-1-108}$$

where the $X_i$, $i = 1, 2, \ldots, n$, are statistically independent and identically distributed gaussian random variables with zero mean and variance $\sigma^2$. As a consequence of the statistical independence of the $X_i$, the characteristic function of $Y$ is

$$\psi_Y(jv) = \frac{1}{(1 - j2v\sigma^2)^{n/2}} \tag{2-1-109}$$

The inverse transform of this characteristic function yields the pdf

$$p_Y(y) = \frac{1}{\sigma^n 2^{n/2}\Gamma(\frac{1}{2}n)} y^{n/2-1} e^{-y/2\sigma^2}, \quad y \geq 0 \tag{2-1-110}$$

where $\Gamma(p)$ is the gamma function, defined as

$$\begin{aligned} &\Gamma(p) = \int_0^\infty t^{p-1} e^{-t}\,dt, \quad p > 0 \\ &\Gamma(p) = (p-1)!, \quad p \text{ an integer}, p > 0 \\ &\Gamma(\tfrac{1}{2}) = \sqrt{\pi}, \quad \Gamma(\tfrac{3}{2}) = \tfrac{1}{2}\sqrt{\pi} \end{aligned} \tag{2-1-111}$$

This pdf, which is a generalization of (2-1-105), is called a *chi-square* (or

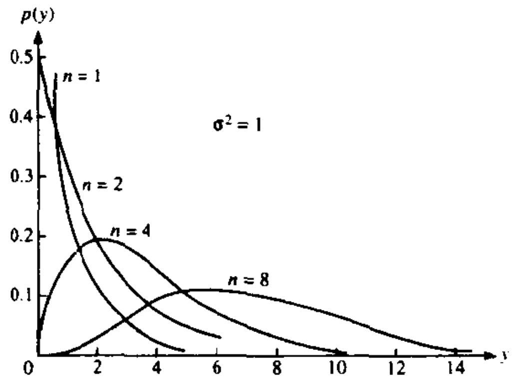



**FIGURE 2-1-9** The pdf of a chi-square-distributed random variable for several degrees of freedom.

*gamma) pdf with n degrees of freedom.* It is illustrated in Fig. 2-1-9. The case $n = 2$ yields the exponential distribution.

The first two moments of $Y$ are

$$\begin{aligned} E(Y) &= n\sigma^2 \\ E(Y^2) &= 2n\sigma^4 + n^2\sigma^4 \\ \sigma_y^2 &= 2n\sigma^4 \end{aligned} \tag{2-1-112}$$

The cdf of $Y$ is

$$F_Y(y) = \int_0^y \frac{1}{\sigma^n 2^{n/2}\Gamma(\frac{1}{2}n)} u^{n/2-1} e^{-u/2\sigma^2}\, du, \qquad y \geqslant 0 \tag{2-1-113}$$

This integral can be easily manipulated into the form of the incomplete gamma function, which is tabulated by Pearson (1965). When $n$ is even, the integral in (2-1-113) can be expressed in closed form. Specifically, let $m = \frac{1}{2}n$, where $m$ is an integer. Then, by repeated integration by parts, we obtain

$$F_Y(y) = 1 - e^{-y/2\sigma^2} \sum_{k=0}^{m-1} \frac{1}{k!}\left(\frac{y}{2\sigma^2}\right)^k, \qquad y \geqslant 0 \tag{2-1-114}$$

Let us now consider a noncentral chi-square distribution, which results from squaring a gaussian random variable having a nonzero mean. If $X$ is gaussian with mean $m_x$ and variance $\sigma^2$, the random variable $Y = X^2$ has the pdf

$$p_Y(y) = \frac{1}{\sqrt{2\pi y}\,\sigma} e^{-(y+m_x^2)/2\sigma^2} \cosh\left(\frac{\sqrt{y}\, m_x}{\sigma^2}\right), \qquad y \geqslant 0 \tag{2-1-115}$$

which is obtained by applying the result in (2-1-47) to the gaussian pdf given by (2-1-92). The characteristic function corresponding to this pdf is

$$\psi_Y(jv) = \frac{1}{(1 - j2v\sigma^2)^{1/2}} e^{jm_x^2 v/(1-j2v\sigma^2)} \tag{2-1-116}$$

To generalize these results, let $Y$ be the sum of squares of gaussian random variables as defined by (2-1-108). The $X_i$, $i = 1, 2, \ldots, n$, are assumed to be statistically independent with means $m_i$, $i = 1, 2, \ldots, n$, and identical variances equal to $\sigma^2$. Then the characteristic function of $Y$, obtained from (2-1-116) by applying the relation in (2-1-79), is

$$\psi_Y(jv) = \frac{1}{(1 - j2v\sigma^2)^{n/2}} \exp\left( \frac{jv \sum_{i=1}^{n} m_i^2}{1 - j2v\sigma^2} \right) \tag{2-1-117}$$

This characteristic function can be inverse-Fourier-transformed to yield the pdf

$$p_Y(y) = \frac{1}{2\sigma^2} \left( \frac{y}{s^2} \right)^{(n-2)/4} e^{-(s^2+y)/2\sigma^2} I_{n/2-1}\left( \sqrt{y}\, \frac{s}{\sigma^2} \right), \quad y \geqslant 0 \tag{2-1-118}$$

where, by definition,

$$s^2 = \sum_{i=1}^{n} m_i^2 \tag{2-1-119}$$

and $I_\alpha(x)$ is the $\alpha$th-order modified Bessel function of the first kind, which may be represented by the infinite series

$$I_\alpha(x) = \sum_{k=0}^{\infty} \frac{(x/2)^{\alpha+2k}}{k!\, \Gamma(\alpha + k + 1)}, \quad x \geqslant 0 \tag{2-1-120}$$

The pdf given by (2-1-118) is called the *noncentral chi-square pdf with n degrees of freedom*. The parameter $s^2$ is called the *noncentrality parameter of the distribution*.

The cdf of the noncentral chi square with $n$ degrees of freedom is

$$F_Y(y) = \int_0^y \frac{1}{2\sigma^2} \left( \frac{u}{s^2} \right)^{(n-2)/4} e^{-(s^2+u)/2\sigma^2} I_{n/2-1}\left( \sqrt{u}\, \frac{s}{\sigma^2} \right) du \tag{2-1-121}$$

There is no closed-form expression for this integral. However, when $m = \frac{1}{2}n$ is an integer, the cdf can be expressed in terms of the generalized Marcum's $Q$ function, which is defined as

$$\begin{aligned} Q_m(a, b) &= \int_b^\infty x \left( \frac{x}{a} \right)^{m-1} e^{-(x^2+a^2)/2} I_{m-1}(ax)\, dx \\ &= Q_1(a, b) + e^{(a^2+b^2)/2} \sum_{k=1}^{m-1} \left( \frac{b}{a} \right)^k I_k(ab) \end{aligned} \tag{2-1-122}$$

where

$$Q_1(a, b) = e^{-(a^2+b^2)/2} \sum_{k=0}^{\infty} \left( \frac{a}{b} \right)^k I_k(ab), \quad b > a > 0 \tag{2-1-123}$$

If we change the variable of integration in (2-1-121) from $u$ to $x$, where

$$x^2 = u/\sigma^2$$

and let $a^2 = s^2/\sigma^2$, then it is easily shown that

$$F_Y(y) = 1 - Q_m\left(\frac{s}{\sigma}, \frac{\sqrt{y}}{\sigma}\right) \tag{2-1-124}$$

Finally, we state that the first two moments of a noncentral chi-square-distributed random variable are

$$\begin{aligned} E(Y) &= n\sigma^2 + s^2 \\ E(Y^2) &= 2n\sigma^4 + 4\sigma^2 s^2 + (n\sigma^2 + s^2)^2 \\ \sigma_y^2 &= 2n\sigma^4 + 4\sigma^2 s^2 \end{aligned} \tag{2-1-125}$$

**Rayleigh Distribution** The Rayleigh distribution is frequently used to model the statistics of signals transmitted through radio channels such as cellular radio. This distribution is closely related to the central chi-square distribution. To illustrate this point, let $Y = X_1^2 + X_2^2$ where $X_1$ and $X_2$ are zero-mean statistically independent gaussian random variables, each having a variance $\sigma^2$. From the discussion above, it follows that $Y$ is chi-square-distributed with two degrees of freedom. Hence, the pdf of $Y$ is

$$p_Y(y) = \frac{1}{2\sigma^2} e^{-y/2\sigma^2}, \quad y \geq 0 \tag{2-1-126}$$

Now, suppose we define a new random variable

$$R = \sqrt{X_1^2 + X_2^2} = \sqrt{Y} \tag{2-1-127}$$

Making a simple change of variable in the pdf of (2-1-126), we obtain the pdf of $R$ in the form

$$p_R(r) = \frac{r}{\sigma^2} e^{-r^2/2\sigma^2}, \quad r \geq 0 \tag{2-1-128}$$

This is the pdf of a Rayleigh-distributed random variable. The corresponding cdf is

$$\begin{aligned} F_R(r) &= \int_0^r \frac{u}{\sigma^2} e^{-u^2/2\sigma^2}\, du \\ &= 1 - e^{-r^2/2\sigma^2}, \quad r \geq 0 \end{aligned} \tag{2-1-129}$$

The moments of $R$ are

$$E(R^k) = (2\sigma^2)^{k/2}\Gamma(1 + \tfrac{1}{2}k) \tag{2-1-130}$$

and the variance is

$$\sigma_r^2 = (2 - \tfrac{1}{2}\pi)\sigma^2 \tag{2-1-131}$$

The characteristic function of the Rayleigh-distributed random variable is

$$\psi_R(jv) = \int_0^\infty \frac{r}{\sigma^2} e^{-r^2/2\sigma^2} e^{jvr}\, dr \tag{2-1-132}$$

This integral may be expressed as

$$\begin{aligned}\psi_R(jv) &= \int_0^\infty \frac{r}{\sigma^2} e^{-r^2/2\sigma^2} \cos vr\, dr + j\int_0^\infty \frac{r}{\sigma^2} e^{-r^2/2\sigma^2} \sin vr\, dr \\ &= {}_1F_1(1, \tfrac{1}{2}; -\tfrac{1}{2}v^2\sigma^2) + j\sqrt{\tfrac{1}{2}\pi}\, v\sigma^2 e^{-v^2\sigma^2/2}\end{aligned} \tag{2-1-133}$$

where ${}_1F_1(1, \frac{1}{2}; -a)$ is the confluent hypergeometric function, which is defined as

$${}_1F_1(\alpha, \beta; x) = \sum_{k=0}^\infty \frac{\Gamma(\alpha + k)\Gamma(\beta)x^k}{\Gamma(\alpha)\Gamma(\beta + k)k!}, \quad \beta \neq 0, -1, -2, \ldots \tag{2-1-134}$$

Beaulieu (1990) has shown that ${}_1F_1(1, \frac{1}{2}; -a)$ may be expressed as

$${}_1F_1(1, \tfrac{1}{2}; -a) = -e^{-a} \sum_{k=0}^\infty \frac{a^k}{(2k-1)k!} \tag{2-1-135}$$

As a generalization of the above expression, consider the random variable

$$R = \sqrt{\sum_{i=1}^n X_i^2} \tag{2-1-136}$$

where the $X_i$, $i = 1, 2, \ldots, n$, are statistically independent, identically distributed zero mean gaussian random variables. The random variable $R$ has a generalized Rayleigh distribution. Clearly, $Y = R^2$ is chi-square-distributed with $n$ degrees of freedom. Its pdf is given by (2-1-110). A simple change in variable in (2-1-110) yields the pdf of $R$ in the form

$$p_R(r) = \frac{r^{n-1}}{2^{(n-2)/2}\sigma^n\Gamma(\frac{1}{2}n)} e^{-r^2/2\sigma^2}, \quad r \geq 0 \tag{2-1-137}$$

As a consequence of the functional relationship between the central chi-square and the Rayleigh distributions, the corresponding cdfs are similar. Thus, for any $n$, the cdf of $R$ can be put in the form of the incomplete gamma function. In the special case when $n$ is even, i.e., $n = 2m$, the cdf of $R$ can be expressed in the closed form

$$F_R(r) = 1 - e^{-r^2/2\sigma^2} \sum_{k=0}^{m-1} \frac{1}{k!}\left(\frac{r^2}{2\sigma^2}\right)^k, \quad r \geq 0 \tag{2-1-138}$$

Finally, we state that the $k$th moment of $R$ is

$$E(R^k) = (2\sigma^2)^{k/2} \frac{\Gamma(\frac{1}{2}(n+k))}{\Gamma(\frac{1}{2}n)}, \quad k \geq 0 \tag{2-1-139}$$

which holds for any integer $n$.

**Rice Distribution** Just as the Rayleigh distribution is related to the central chi-square distribution, the Rice distribution is related to the noncentral chi-square distribution. To illustrate this relation, let $Y = X_1^2 + X_2^2$, where $X_1$ and $X_2$ are statistically independent gaussian random variables with means $m_i$, $i = 1, 2$, and common variance $\sigma^2$. From the previous discussion, we know that $Y$ has a noncentral chi-square distribution with noncentrality parameter $s^2 = m_1^2 + m_2^2$. The pdf of $Y$, obtained from (2-1-118) for $n = 2$, is

$$p_Y(y) = \frac{1}{2\sigma^2} e^{-(s^2+y)/2\sigma^2} I_0\left(\sqrt{y}\frac{s}{\sigma^2}\right), \quad y \geq 0 \tag{2-1-140}$$

Now, we define a new random variable $R = \sqrt{Y}$. The pdf of $R$, obtained from (2-1-140) by a simple change of variable, is

$$p_R(r) = \frac{r}{\sigma^2} e^{-(r^2+s^2)/2\sigma^2} I_0\left(\frac{rs}{\sigma^2}\right), \quad r \geq 0 \tag{2-1-141}$$

This is the pdf of a Ricean-distributed random variable. As will be shown in Chapter 5, this pdf characterizes the statistics of the envelope of a signal corrupted by additive narrowband gaussian noise. It is also used to model the signal statistics of signals transmitted through some radio channels. The cdf of $R$ is easily obtained by specializing the results in (2-1-124) to the case $m = 1$. This yields

$$F_R(r) = 1 - Q_1\left(\frac{s}{\sigma}, \frac{r}{\sigma}\right), \quad r \geq 0 \tag{2-1-142}$$

where $Q_1(a, b)$ is defined by (2-1-123).

As a generalization of the expressions given above, let $R$ be defined as in (2-1-136) where the $X_i$, $i = 1, 2, \ldots, n$ are statistically independent gaussian random variables with means $m_i$, $i = 1, 2, \ldots, n$, and identical variances equal to $\sigma^2$. The random variable $R^2 = Y$ has a noncentral chi-square distribution with $n$ degrees of freedom and noncentrality parameter $s^2$ given by (2-1-119). Its pdf is given by (2-1-118). Hence the pdf of $R$ is

$$p_R(r) = \frac{r^{n/2}}{\sigma^2 s^{(n-2)/2}} e^{-(r^2+s^2)/2\sigma^2} I_{n/2-1}\left(\frac{rs}{\sigma^2}\right), \quad r \geq 0 \tag{2-1-143}$$

and the corresponding cdf is

$$F_R(r) = P(R \leqslant r) = P(\sqrt{Y} \leqslant r) = P(Y \leqslant r^2) = F_Y(r^2) \tag{2-1-144}$$

where $F_Y(r^2)$ is given by (2-1-121). In the special case where $m = \frac{1}{2}n$ is an integer, we have

$$F_R(r) = 1 - Q_m\left(\frac{s}{\sigma}, \frac{r}{\sigma}\right), \quad r \geqslant 0 \tag{2-1-145}$$

which follows from (2-1-124). Finally, we state that the $k$th moment of $R$ is

$$E(R^k) = (2\sigma^2)^{k/2} e^{-s^2/2\sigma^2} \frac{\Gamma(\frac{1}{2}(n+k))}{\Gamma(\frac{1}{2}n)} {}_1F_1\left(\frac{n+k}{2}, \frac{n}{2}; \frac{s^2}{2\sigma^2}\right), \quad k \geqslant 0 \tag{2-1-146}$$

where ${}_1F_1(\alpha, \beta; x)$ is the confluent hypergeometric function.

**Nakagami $m$-Distribution** Both the Rayleigh distribution and the Rice distribution are frequently used to describe the statistical fluctuations of signals received from a multipath fading channel. These channel models are considered in Chapter 14. Another distribution that is frequently used to characterize the statistics of signals transmitted through multipath fading channels is the Nakagami $m$-distribution. The pdf for this distribution is given by Nakagami (1960) as

$$p_R(r) = \frac{2}{\Gamma(m)}\left(\frac{m}{\Omega}\right)^m r^{2m-1} e^{-mr^2/\Omega} \tag{2-1-147}$$

where $\Omega$ is defined as

$$\Omega = E(R^2) \tag{2-1-148}$$

and the parameter $m$ is defined as the ratio of moments, called the *fading figure*,

$$m = \frac{\Omega^2}{E[(R^2 - \Omega)^2]}, \quad m \geqslant \tfrac{1}{2} \tag{2-1-149}$$

A normalized version of (2-1-147) may be obtained by defining another random variable $X = R/\sqrt{\Omega}$ (see Problem 2-15). The $n$th moment of $R$ is

$$E(R^n) = \frac{\Gamma(m + \frac{1}{2}n)}{\Gamma(m)}\left(\frac{\Omega}{m}\right)^{n/2}$$

By setting $m = 1$, we observe that (2-1-147) reduces to a Rayleigh pdf. For

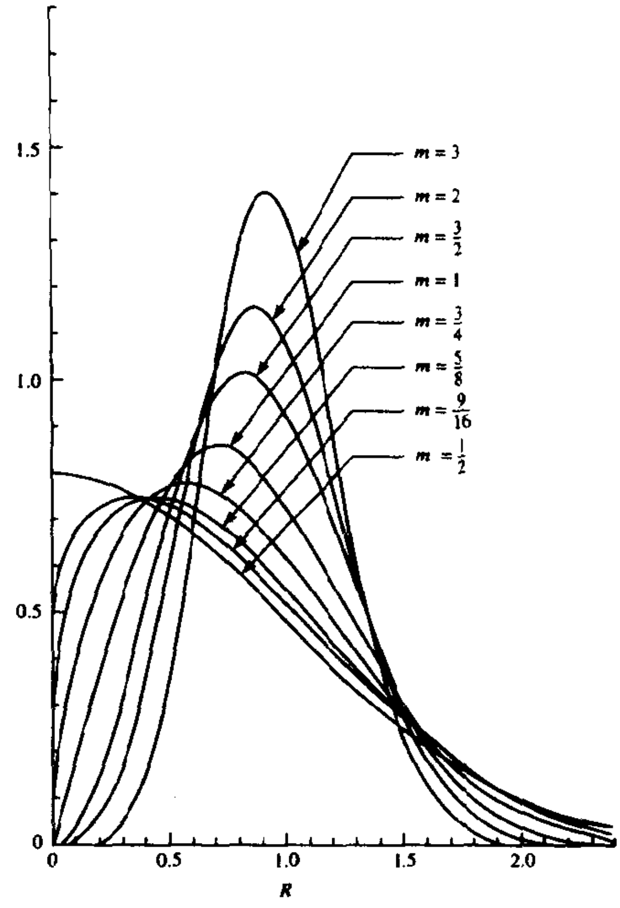



**FIGURE 2-1-10** The $m$-distributed pdf, shown with $\Omega = 1$. $m$ is the fading figure. (*Miyagaki* et al. *1978*.)

values of $m$ in the range $\frac{1}{2} \leq m \leq 1$, we obtain pdfs that have larger tails than a Rayleigh-distributed random variable. For values of $m > 1$, the tail of the pdf decays faster than that of the Rayleigh. Figure 2-1-10 illustrates the pdfs for different values of $m$.

**Multivariate Gaussian Distribution** Of the many multivariate or multidimensional distributions that can be defined, the multivariate gaussian distribution is the most important and the one most likely to be encountered in practice. We shall briefly introduce this distribution and state its basic properties.

Let us assume that $X_i$, $i = 1, 2, \ldots, n$, are gaussian random variables with means $m_i$, $i = 1, 2, \ldots, n$, variances $\sigma_i^2$, $i = 1, 2, \ldots, n$, and covariances $\mu_{ij}$, $i, j = 1, 2, \ldots, n$. Clearly, $\mu_{ii} = \sigma_i^2$, $i = 1, 2, \ldots, n$. Let $\mathbf{M}$ denote the $n \times n$

covariance matrix with elements $\{\mu_{ij}\}$, let $\mathbf{X}$ denote the $n \times 1$ column vector of random variables, and let $\mathbf{m}_x$ denote the $n \times 1$ column vector of mean values $m_i$, $i = 1, 2, \ldots, n$. The joint pdf of the gaussian random variables $X_i$, $i = 1, 2, \ldots, n$, is defined as

$$p(x_1, x_2, \ldots, x_n) = \frac{1}{(2\pi)^{n/2}(\det \mathbf{M})^{1/2}} \exp\left[-\tfrac{1}{2}(\mathbf{x} - \mathbf{m}_x)'\mathbf{M}^{-1}(\mathbf{x} - \mathbf{m}_x)\right] \tag{2-1-150}$$

where $\mathbf{M}^{-1}$ denotes the inverse of $\mathbf{M}$ and $\mathbf{x}'$ denotes the transpose of $\mathbf{x}$.

The characteristic function corresponding to this $n$-dimensional joint pdf is

$$\psi(j\mathbf{v}) = E(e^{j\mathbf{v}'\mathbf{x}})$$

where $\mathbf{v}$ is an $n$-dimensional vector with elements $v_i$, $i = 1, 2, \ldots, n$. Evaluation of this $n$-dimensional Fourier transform yields the result

$$\psi(j\mathbf{v}) = \exp\left(j\mathbf{m}_x'\mathbf{v} - \tfrac{1}{2}\mathbf{v}'\mathbf{M}\mathbf{v}\right) \tag{2-1-151}$$

An important special case of (2-1-150) is the bivariate or two-dimensional gaussian pdf. The mean $\mathbf{m}_x$ and the covariance matrix $\mathbf{M}$ for this case are

$$\mathbf{m}_x = \begin{bmatrix} m_1 \\ m_2 \end{bmatrix}, \qquad \mathbf{M} = \begin{bmatrix} \sigma_1^2 & \mu_{12} \\ \mu_{12} & \sigma_2^2 \end{bmatrix} \tag{2-1-152}$$

where the joint central moment $\mu_{12}$ is defined as

$$\mu_{12} = E[(X_1 - m_1)(X_2 - m_2)]$$

It is convenient to define a normalized covariance

$$\rho_{ij} = \frac{\mu_{ij}}{\sigma_i \sigma_j}, \qquad i \neq j \tag{2-1-153}$$

where $\rho_{ij}$ satisfies the condition $0 \leq |\rho_{ij}| \leq 1$. When dealing with the two-dimensional case, it is customary to drop the subscripts on $\mu_{12}$ and $\rho_{12}$. Hence the covariance matrix is expressed as

$$\mathbf{M} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \tag{2-1-154}$$

Its inverse is

$$\mathbf{M}^{-1} = \frac{1}{\sigma_1^2\sigma_2^2(1 - \rho^2)} \begin{bmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix} \tag{2-1-155}$$

and $\det \mathbf{M} = \sigma_1^2\sigma_2^2(1-\rho^2)$. Substitution for $\mathbf{M}^{-1}$ into (2-1-150) yields the desired bivariate gaussian pdf in the form

$$p(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \times \exp\left[-\frac{\sigma_2^2(x_1-m_1)^2 - 2\rho\sigma_1\sigma_2(x_1-m_1)(x_2-m_2) + \sigma_1^2(x_2-m_2)^2}{2\sigma_1^2\sigma_2^2(1-\rho^2)}\right] \tag{2-1-156}$$

We note that when $\rho = 0$, the joint pdf $p(x_1, x_2)$ in (2-1-156) factors into the product $p(x_1)p(x_2)$, where $p(x_i)$, $i = 1, 2$, are the marginal pdfs. Since $\rho$ is a measure of the correlation between $X_1$ and $X_2$, we have shown that when the gaussian random variables $X_1$ and $X_2$ are uncorrelated, they are also statistically independent. This is an important property of gaussian random variables, which does not hold in general for other distributions. It extends to $n$-dimensional gaussian random variables in a straightforward manner. That is, if $\rho_{ij} = 0$ for $i \neq j$ then the random variables $X_i$, $i = 1, 2, \ldots, n$ are uncorrelated and, hence, statistically independent.

Now, let us consider a linear transformation of $n$ gaussian random variables $X_i$, $i = 1, 2, \ldots, n$, with mean vector $\mathbf{m}_x$ and covariance matrix $\mathbf{M}$. Let

$$\mathbf{Y} = \mathbf{AX} \tag{2-1-157}$$

where $\mathbf{A}$ is a nonsingular matrix. As shown previously, the jacobian of this transformation is $J = 1/\det \mathbf{A}$. Since $\mathbf{X} = \mathbf{A}^{-1}\mathbf{Y}$, we may substitute for $\mathbf{X}$ in (2-1-150) and, thus, we obtain the joint pdf of $\mathbf{Y}$ in the form

$$\begin{aligned} p(\mathbf{y}) &= \frac{1}{(2\pi)^{n/2}(\det \mathbf{M})^{1/2}\det \mathbf{A}} \exp\left[-\tfrac{1}{2}(\mathbf{A}^{-1}\mathbf{y} - \mathbf{m}_x)'\mathbf{M}^{-1}(\mathbf{A}^{-1}\mathbf{y} - \mathbf{m}_x)\right] \\ &= \frac{1}{(2\pi)^{n/2}(\det \mathbf{Q})^{1/2}} \exp\left[-\tfrac{1}{2}(\mathbf{y} - \mathbf{m}_y)'\mathbf{Q}^{-1}(\mathbf{y} - \mathbf{m}_y)\right] \end{aligned} \tag{2-1-158}$$

where the vector $\mathbf{m}_y$ and the matrix $\mathbf{Q}$ are defined as

$$\begin{aligned} \mathbf{m}_y &= \mathbf{A}\mathbf{m}_x \\ \mathbf{Q} &= \mathbf{AMA} \end{aligned} \tag{2-1-159}$$

Thus we have shown that a linear transformation of a set of jointly gaussian random variables results in another set of jointly gaussian random variables.

Suppose that we wish to perform a linear transformation that results in $n$ statistically independent gaussian random variables. How should the matrix $\mathbf{A}$ be selected? From our previous discussion, we know that the gaussian random variables are statistically independent if they are pairwise-uncorrelated, i.e., if the covariance matrix $\mathbf{Q}$ is diagonal. Therefore, we must have

$$\mathbf{AMA}' = \mathbf{D} \tag{2-1-160}$$

where $\mathbf{D}$ is a diagonal matrix. The matrix $\mathbf{M}$ is a covariance matrix; hence, it is positive definite. One solution is to select $\mathbf{A}$ to be an orthogonal matrix ($\mathbf{A}' = \mathbf{A}^{-1}$) consisting of columns that are the eigenvectors of the covariance matrix $\mathbf{M}$. Then $\mathbf{D}$ is a diagonal matrix with diagonal elements equal to the eigenvalues of $\mathbf{M}$.

**Example 2-1-5**

Consider the bivariate gaussian pdf with covariance matrix

$$\mathbf{M} = \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{bmatrix}$$

Let us determine the transformation $\mathbf{A}$ that will result in uncorrelated random variables. First, we solve for the eigenvalues of $\mathbf{M}$. The characteristic equation is

$$\det(\mathbf{M} - \lambda\mathbf{I}) = 0$$

$$(1-\lambda)^2 - \tfrac{1}{4} = 0$$

$$\lambda = \tfrac{3}{2}, \tfrac{1}{2}$$

Next we determine the two eigenvectors. If $\mathbf{a}$ denotes an eigenvector, we have

$$(\mathbf{M} - \lambda\mathbf{I})\mathbf{a} = 0$$

With $\lambda_1 = \frac{3}{2}$ and $\lambda_2 = \frac{1}{2}$, we obtain the eigenvectors

$$\mathbf{a}_1 = \begin{bmatrix} \sqrt{\frac{1}{2}} \\ \sqrt{\frac{1}{2}} \end{bmatrix}, \quad \mathbf{a}_2 = \begin{bmatrix} \sqrt{\frac{1}{2}} \\ -\sqrt{\frac{1}{2}} \end{bmatrix}$$

Therefore,

$$\mathbf{A} = \sqrt{\tfrac{1}{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

It is easily verified that $\mathbf{A}^{-1} = \mathbf{A}'$ and that

$$\mathbf{AMA}' = \mathbf{D}$$

where the diagonal elements of $\mathbf{D}$ are $\frac{3}{2}$ and $\frac{1}{2}$.

### 2-1-5 Upper Bounds on the Tail Probability

In evaluating the performance of a digital communication system, it is often necessary to determine the area under the tail of the pdf. We refer to this area as the *tail probability*. In this section, we present two upper bounds on the tail probability. The first, obtained from the Chebyshev inequality, is rather loose. The second, called the *Chernoff bound*, is much tighter.

**Chebyshev Inequality** Suppose that $X$ is an arbitrary random variable with finite mean $m_x$ and finite variance $\sigma_x^2$. For any positive number $\delta$,

$$P(|X - m_x| \geqslant \delta) \leqslant \frac{\sigma_x^2}{\delta^2} \tag{2-1-161}$$

This relation is called the *Chebyshev inequality*. The proof of this bound is relatively simple. We have

$$\begin{aligned}\sigma_x^2 = \int_{-\infty}^{\infty} (x - m_x)^2 p(x)\, dx &\geqslant \int_{|x - m_x| \geqslant \delta} (x - m_x)^2 p(x)\, dx \\ &\geqslant \delta^2 \int_{|x - m_x| \geqslant \delta} p(x)\, dx = \delta^2 P(|X - m_x| \geqslant \delta)\end{aligned}$$

Thus the validity of the inequality is established.

It is apparent that the Chebyshev inequality is simply an upper bound on the area under the tails of the pdf $p(y)$, where $Y = X - m_x$, i.e., the area of $p(y)$ in the intervals $(-\infty, -\delta)$ and $(\delta, \infty)$. Hence, the Chebyshev inequality may be expressed as

$$1 - [F_Y(\delta) - F_Y(-\delta)] \leqslant \frac{\sigma_x^2}{\delta^2} \tag{2-1-162}$$

or, equivalently, as

$$1 - [F_X(m_x + \delta) - F_X(m_x - \delta)] \leqslant \frac{\sigma_x^2}{\delta^2} \tag{2-1-163}$$

There is another way to view the Chebyshev bound. Working with the zero mean random variable $Y = X - m_x$, for convenience, suppose we define a function $g(Y)$ as

$$g(Y) = \begin{cases} 1 & (|Y| \geqslant \delta) \\ 0 & (|Y| < \delta) \end{cases} \tag{2-1-164}$$

Since $g(Y)$ is either 0 or 1 with probabilities $P(|Y| < \delta)$ and $P(|Y| \geqslant \delta)$, respectively, its mean value is

$$E[g(Y)] = P(|Y| \geqslant \delta) \tag{2-1-165}$$

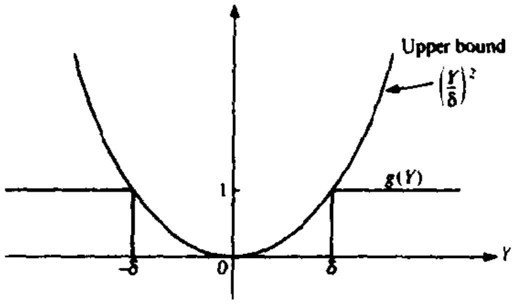



**FIGURE 2-1-11** A quadratic upper bound on $g(Y)$ used in obtaining the tail probability (Chebyshev bound).

Now suppose that we upper-bound $g(Y)$ by the quadratic $(Y/\delta)^2$, i.e.,

$$g(Y) \leq \left(\frac{Y}{\delta}\right)^2 \tag{2-1-166}$$

The graph of $g(Y)$ and the upper bound are shown in Fig. 2-1-11. It follows that

$$E[g(Y)] \leq E\left(\frac{Y^2}{\delta^2}\right) = \frac{E(Y^2)}{\delta^2} = \frac{\sigma_y^2}{\delta^2} = \frac{\sigma_x^2}{\delta^2}$$

Since $E[g(Y)]$ is the tail probability, as seen from (2-1-165), we have obtained the Chebyshev bound.

For many practical applications, the Chebyshev bound is extremely loose. The reason for this may be attributed to the looseness of the quadratic $(Y/\delta)^2$ in overbounding $g(Y)$. There are certainly many other functions that can be used to overbound $g(Y)$. Below, we use an exponential bound to derive an upper bound on the tail probability that is extremely tight.

**Chernoff Bound** The Chebyshev bound given above involves the area under the two tails of the pdf. In some applications we are interested only in the area under one tail, either in the interval $(\delta, \infty)$ or in the interval $(-\infty, \delta)$. In such a case we can obtain an extremely tight upper bound by overbounding the function $g(Y)$ by an exponential having a parameter that can be optimized to yield as tight an upper bound as possible. Specifically, we consider the tail probability in the interval $(\delta, \infty)$. The function $g(Y)$ is overbounded as

$$g(Y) \leq e^{v(Y-\delta)} \tag{2-1-167}$$

where $g(Y)$ is now defined as

$$g(Y) = \begin{cases} 1 & (Y \geq \delta) \\ 0 & (Y < \delta) \end{cases} \tag{2-1-168}$$

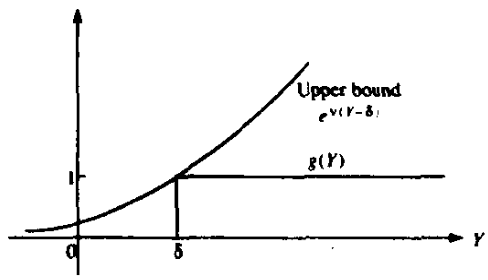



**FIGURE 2-1-12** An exponential upper bound on $g(Y)$ used in obtaining the tail probability (Chernoff bound).

and $v \geq 0$ is the parameter to be optimized. The graph of $g(Y)$ and the exponential upper bound are shown in Fig. 2-1-12.

The expected value of $g(Y)$ is

$$E[g(Y)] = P(Y \geq \delta) \leq E(e^{v(Y-\delta)}) \tag{2-1-169}$$

This bound is valid for any $v \geq 0$. The tightest upper bound is obtained by selecting the value of $v$ that minimizes $E(e^{v(Y-\delta)})$. A necessary condition for a minimum is

$$\frac{d}{dv}E(e^{v(Y-\delta)}) = 0 \tag{2-1-170}$$

But the order of differentiation and expectation can be interchanged, so that

$$\begin{aligned}\frac{d}{dv}E(e^{v(Y-\delta)}) &= E\left(\frac{d}{dv}e^{v(Y-\delta)}\right)\\ &= E[(Y-\delta)e^{v(Y-\delta)}]\\ &= e^{-v\delta}[E(Ye^{vY}) - \delta E(e^{vY})] = 0\end{aligned}$$

Therefore the value of $v$ that gives the tightest upper bound is the solution to the equation

$$E(Ye^{vY}) - \delta E(e^{vY}) = 0 \tag{2-1-171}$$

Let $\hat{v}$ be the solution of (2-1-171). Then, from (2-1-169), the upper bound on the one-sided tail probability is

$$P(Y \geq \delta) \leq e^{-\hat{v}\delta}E(e^{\hat{v}Y}) \tag{2-1-172}$$

This is the Chernoff bound for the upper tail probability for a discrete or a continuous random variable having a zero mean.† This bound may be used to show that $Q(x) \leq e^{-x^2/2}$, where $Q(x)$ is the area in the tail of the gaussian pdf (see Problem 2-18).

† Note that $E(e^{vY})$ for real $v$ is not the characteristic function of $Y$. It is called the *moment generating function of Y*.

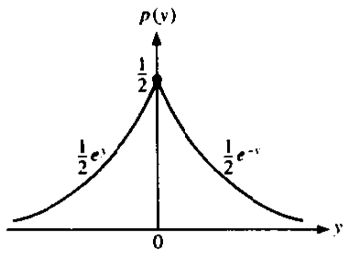



**FIGURE 2-1-13** The pdf of a Laplace-distributed random variable.

An upper bound on the lower tail probability can be obtained in a similar manner, with the result that

$$P(Y \leq \delta) \leq e^{-\hat{v}\delta} E(e^{\hat{v}Y}) \qquad (2\text{-}1\text{-}173)$$

where $\hat{v}$ is the solution to (2-1-171) and $\delta < 0$.

**Example 2-1-6**

Consider the (Laplace) pdf

$$p(y) = \tfrac{1}{2} e^{-|y|} \qquad (2\text{-}1\text{-}174)$$

which is illustrated in Fig. 2-1-13. Let us evaluate the upper tail probability from the Chernoff bound and compare it with the true tail probability, which is

$$P(Y \geq \delta) = \int_{\delta}^{\infty} \tfrac{1}{2} e^{-y}\, dy = \tfrac{1}{2} e^{-\delta} \qquad (2\text{-}1\text{-}175)$$

To solve (2-1-171) for $\hat{v}$, we must determine the moments $E(Ye^{vY})$ and $E(e^{vY})$. For the pdf in (2-1-174), we find that

$$\begin{aligned} E(Ye^{vY}) &= \frac{2v}{(v+1)^2(v-1)^2} \\ E(e^{vY}) &= \frac{1}{(1+v)(1-v)} \end{aligned} \qquad (2\text{-}1\text{-}176)$$

Substituting these moments into (2-1-171), we obtain the quadratic equation

$$v^2\delta + 2v - \delta = 0$$

which has the solutions

$$\hat{v} = \frac{-1 \pm \sqrt{1+\delta^2}}{\delta} \qquad (2\text{-}1\text{-}177)$$

Since $\hat{v}$ must be positive, one of the two solutions is discarded. Thus

$$\hat{v} = \frac{-1 + \sqrt{1+\delta^2}}{\delta} \qquad (2\text{-}1\text{-}178)$$

Finally, we evaluate the upper bound in (2-1-172) by eliminating $E(e^{\hat{\nu}Y})$ using the second relation in (2-1-176) and by substituting for $\hat{\nu}$ from (2-1-178). The result is

$$P(Y \geq \delta) \leq \frac{\delta^2}{2(-1+\sqrt{1+\delta^2})} e^{1-\sqrt{1+\delta^2}} \tag{2-1-179}$$

For $\delta \gg 1$, (2-1-179) reduces to

$$P(Y \geq \delta) \leq \frac{\delta}{2} e^{-\delta} \tag{2-1-180}$$

We note that the Chernoff bound decreases exponentially as $\delta$ increases. Consequently, it approximates closely the exact tail probability given by (2-1-175). In contrast, the Chebyshev upper bound for the upper tail probability obtained by taking one-half of the probability in the two tails (due to symmetry in the pdf) is

$$P(Y \geq \delta) \leq \frac{1}{\delta^2}$$

Hence, this bound is extremely loose.

When the random variable has a nonzero mean, the Chernoff bound can be extended as we now demonstrate. If $Y = X - m_x$, we have

$$P(Y \geq \delta) = P(X - m_x \geq \delta) = P(X \geq m_x + \delta) = P(X \geq \delta_m)$$

where, by definition, $\delta_m = m_x + \delta$. Since $\delta > 0$, it follows that $\delta_m > m_x$. Let $g(X)$ be defined as

$$g(X) = \begin{cases} 1 & (X \geq \delta_m) \\ 0 & (X < \delta_m) \end{cases} \tag{2-1-181}$$

and upper-bounded as

$$g(X) \leq e^{\nu(X-\delta_m)} \tag{2-1-182}$$

From this point, the derivation parallels the steps contained in (2-1-169)–(2-1-172). The final result is

$$P(X \geq \delta_m) \leq e^{-\hat{\nu}\delta_m} E(e^{\hat{\nu}X}) \tag{2-1-183}$$

where $\delta_m > m_x$ and $\hat{\nu}$ is the solution to the equation

$$E(Xe^{\nu X}) - \delta_m E(e^{\nu X}) = 0 \tag{2-1-184}$$

In a similar manner, we can obtain the Chernoff bound for the lower tail probability. For $\delta < 0$, we have

$$P(X - m_x \leq \delta) = P(X \leq m_x + \delta) = P(X \leq \delta_m) \leq E(e^{\nu(X-\delta_m)}) \tag{2-1-185}$$

From our previous development, it is apparent that (2-1-185) results in the bound

$$P(X \leq \delta_m) \leq e^{-\hat{\nu}\delta_m} E(e^{\hat{\nu}X}) \tag{2-1-186}$$

where $\delta_m < m_x$ and $\hat{\nu}$ is the solution to (2-1-184).

### 2-1-6 Sums of Random Variables and the Central Limit Theorem

We have previously considered the problem of determining the pdf of a sum of $n$ statistically independent random variables. In this section, we again consider the sum of statistically independent random variables, but our approach is different and is independent of the particular pdf of the random variables in the sum. To be specific, suppose that $X_i$, $i = 1, 2, \ldots, n$, are statistically independent and identically distributed random variables, each having a finite mean $m_x$ and a finite variance $\sigma_x^2$. Let $Y$ be defined as the normalized sum, called the *sample mean*:

$$Y = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad (2\text{-}1\text{-}187)$$

First we shall determine upper bounds on the tail probabilities of $Y$ and then we shall prove a very important theorem regarding the pdf of $Y$ in the limit as $n \to \infty$.

The random variable $Y$ defined in (2-1-187) is frequently encountered in estimating the mean of a random variable $X$ from a number of observations $X_i$, $i = 1, 2, \ldots, n$. In other words, the $X_i$, $i = 1, 2, \ldots, n$, may be considered as independent samples drawn from a distribution $F_X(x)$, and $Y$ is the estimate of the mean $m_x$.

The mean of $Y$ is

$$\begin{aligned} E(Y) = m_y &= \frac{1}{n}\sum_{i=1}^{n} E(X_i) \\ &= m_x \end{aligned}$$

The variance of $Y$ is

$$\begin{aligned} \sigma_y^2 &= E(Y^2) - m_y^2 = E(Y^2) - m_x^2 \\ &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} E(X_iX_j) - m_x^2 \\ &= \frac{1}{n^2}\sum_{i=1}^{n} E(X_i^2) + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{\substack{j=1 \\ i\neq j}}^{n} E(X_i)E(X_j) - m_x^2 \\ &= \frac{1}{n}(\sigma_x^2 + m_x^2) + \frac{1}{n^2}n(n-1)m_x^2 - m_x^2 \\ &= \frac{\sigma_x^2}{n} \end{aligned}$$

When $Y$ is viewed as an estimate for the mean $m_x$, we note that its expected value is equal to $m_x$ and its variance decreases inversely with the number of samples $n$. As $n$ approaches infinity, the variance $\sigma_y^2$ approaches zero. An estimate of a parameter (in this case the mean $m_x$) that satisfies the conditions that its expected value converges to the true value of the parameter and the variance converges to zero as $n \to \infty$ is said to be a *consistent estimate*.

The tail probability of the random variable $Y$ can be upper-bounded by use of the bounds presented in Section 2-1-5. The Chebyshev inequality applied to $Y$ is

$$P(|Y - m_y| \geq \delta) \leq \frac{\sigma_y^2}{\delta^2}$$

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - m_x\right| \geq \delta\right) \leq \frac{\sigma_x^2}{n\delta^2} \tag{2-1-188}$$

In the limit as $n \to \infty$, (2-1-188) becomes

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - m_x\right| \geq \delta\right) = 0 \tag{2-1-189}$$

Therefore, the probability that the estimate of the mean differs from the true mean $m_x$ by more than $\delta$ ($\delta > 0$) approaches zero as $n$ approaches infinity. This statement is a form of the law of large numbers. Since the upper bound converges to zero relatively slowly, i.e., inversely with $n$, the expression in (2-1-188) is called the *weak law of large numbers*.

The Chernoff bound applied to the random variable $Y$ yields an exponential dependence of $n$, and thus provides a tighter upper bound on the one-sided tail probability. Following the procedure developed in Section 2-1-5, we can determine that the tail probability for $y$ is

$$\begin{aligned} P(Y - m_y \geq \delta) &= P\left(\frac{1}{n}\sum_{i=1}^{n} X_i - m_x \geq \delta\right) \\ &= P\left(\sum_{i=1}^{n} X_i \geq n\delta_m\right) \leq E\left\{\exp\left[v\left(\sum_{i=1}^{n} X_i - n\delta_m\right)\right]\right\} \end{aligned} \tag{2-1-190}$$

where $\delta_m = m_x + \delta$ and $\delta > 0$. But the $X_i$, $i = 1, 2, \ldots, n$, are statistically independent and identically distributed. Hence,

$$\begin{aligned} E\left\{\exp\left[v\left(\sum_{i=1}^{n} X_i - n\delta_m\right)\right]\right\} &= e^{-vn\delta_m} E\left[\exp\left(v\sum_{i=1}^{n} X_i\right)\right] \\ &= e^{-vn\delta_m}\prod_{i=1}^{n} E(e^{vX_i}) \\ &= [e^{-v\delta_m}E(e^{vX})]^n \end{aligned} \tag{2-1-191}$$

where $X$ denotes any one of the $X_i$. The parameter $v$ that yields the tightest upper bound is obtained by differentiating (2-1-191) and setting the derivative equal to zero. This yields the equation

$$E(Xe^{vX}) - \delta_m E(e^{vX}) = 0 \tag{2-1-192}$$

Let the solution of (2-1-192) be denoted by $\hat{\nu}$. Then, the bound on the upper tail probability is

$$P\left(\frac{1}{n}\sum_{i=1}^{n} X_i \geq \delta_m\right) \leq [e^{-\hat{\nu}\delta_m} E(e^{\hat{\nu}X})]^n, \qquad \delta_m > m_x \tag{2-1-193}$$

In a similar manner, we find that the lower tail probability is upper-bounded as

$$P(Y \leq \delta_m) \leq [e^{-\hat{\nu}\delta_m} E(e^{\hat{\nu}X})]^n, \qquad \delta_m < m_x \tag{2-1-194}$$

where $\hat{\nu}$ is the solution to (2-1-192).

**Example 2-1-7**

Let $X_i$, $i = 1, 2, \ldots, n$, be a set of statistically independent random variables defined as

$$X_i = \begin{cases} 1 & \text{with probability } p < \frac{1}{2} \\ -1 & \text{with probability } 1 - p \end{cases}$$

We wish to determine a tight upper bound on the probability that the sum of the $X_i$ is greater than zero. Since $p < \frac{1}{2}$, we note that the sum will have a negative value for the mean; hence we seek the upper tail probability. With $\delta_m = 0$ in (2-1-193), we have

$$P\left(\sum_{i=1}^{n} X_i \geq 0\right) \leq [E(e^{\hat{\nu}X})]^n \tag{2-1-195}$$

where $\hat{\nu}$ is the solution to the equation

$$E(Xe^{\nu X}) = 0 \tag{2-1-196}$$

Now

$$E(Xe^{\nu X}) = -(1-p)e^{-\nu} + pe^{\nu} = 0$$

Hence

$$\hat{\nu} = \ln\left(\sqrt{\frac{1-p}{p}}\right) \tag{2-1-197}$$

Furthermore,

$$E(e^{\hat{\nu}X}) = pe^{\hat{\nu}} + (1-p)e^{-\hat{\nu}}$$

Therefore the bound in (2-1-195) becomes

$$\begin{aligned} P\left(\sum_{i=1}^{n} X_i \geq 0\right) &\leq [pe^{\hat{\nu}} + (1-p)e^{-\hat{\nu}}]^n \\ &\leq \left[p\sqrt{\frac{1-p}{p}} + (1-p)\sqrt{\frac{p}{1-p}}\right]^n \\ &\leq [4p(1-p)]^{n/2} \end{aligned} \tag{2-1-198}$$

We observe that the upper bound decays exponentially with $n$, as expected.

In contrast, if the Chebyshev bound were evaluated, the tail probability would decrease inversely with $n$.

**Central Limit Theorem** We conclude this section with an extremely useful theorem concerning the cdf of a sum of random variables in the limit as the number of terms in the sum approaches infinity. There are several versions of this theorem. We shall prove the theorem for the case in which the random variables $X_i$, $i = 1, 2, \ldots, n$, being summed are statistically independent and identically distributed, each having a finite mean $m_x$ and a finite variance $\sigma_x^2$. For convenience, we define the normalized random variable

$$U_i = \frac{X_i - m_x}{\sigma_x}, \qquad i = 1, 2, \ldots, n$$

Thus $U_i$ has a zero mean and unit variance. Now, let

$$Y = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} U_i \tag{2-1-199}$$

Since each term in the sum has a zero mean and unit variance, it follows that the normalized (by $1/\sqrt{n}$) random variable $Y$ has zero mean and unit variance. We wish to determine the cdf of $Y$ in the limit as $n \to \infty$.

The characteristic function of $Y$ is

$$\begin{aligned}
\psi_Y(jv) = E(e^{jvY}) &= E\left[\exp\left(\frac{jv \sum_{i=1}^{n} U_i}{\sqrt{n}}\right)\right] \\
&= \prod_{i=1}^{n} \psi_{U_i}\left(\frac{jv}{\sqrt{n}}\right) \\
&= \left[\psi_U\left(\frac{jv}{\sqrt{n}}\right)\right]^n
\end{aligned} \tag{2-1-200}$$

where $U$ denotes any of the $U_i$, which are identically distributed. Now, let us expand the characteristic function of $U$ in a Taylor series. The expansion yields

$$\psi_U\left(j\frac{v}{\sqrt{n}}\right) = 1 + j\frac{v}{\sqrt{n}}E(U) - \frac{v^2}{n2!}E(U^2) + \frac{(jv)^3}{(\sqrt{n})^3 3!}E(U^3) - \cdots \tag{2-1-201}$$

Since $E(U) = 0$ and $E(U^2) = 1$, (2-1-201) simplifies to

$$\psi_U\left(\frac{jv}{\sqrt{n}}\right) = 1 - \frac{v^2}{2n} + \frac{1}{n}R(v, n) \tag{2-1-202}$$

where $R(v, n)/n$ denotes the remainder. We note that $R(v, n)$ approaches zero as $n \to \infty$. Substitution of (2-1-202) into (2-1-200) yields the characteristic function of $Y$ in the form

$$\psi_Y(jv) = \left[1 - \frac{v^2}{2n} + \frac{R(v, n)}{n}\right]^n \tag{2-1-203}$$

Taking the natural logarithm of (2-1-203), we obtain

$$\ln \psi_Y(jv) = n \ln \left[1 - \frac{v^2}{2n} + \frac{R(v, n)}{n}\right] \tag{2-1-204}$$

For small values of $x$, $\ln(1 + x)$ can be expanded in the power series

$$\ln(1 + x) = x - \tfrac{1}{2}x^2 + \tfrac{1}{3}x^3 - \dots$$

This expansion applied to (2-1-204) yields

$$\ln \psi_Y(jv) = n\left[-\frac{v^2}{2n} + \frac{R(v, n)}{n} - \frac{1}{2}\left(-\frac{v^2}{2n} + \frac{R(v, n)}{n}\right)^2 + \dots\right] \tag{2-1-205}$$

Finally, when we take the limit as $n \to \infty$, (2-1-205) reduces to $\lim_{n \to \infty} \ln \psi_Y(jv) = -\frac{1}{2}v^2$, or, equivalently,

$$\lim_{n \to \infty} \psi_Y(jv) = e^{-v^2/2} \tag{2-1-206}$$

But, this is just the characteristic function of a gaussian random variable with zero mean and unit variance. Thus we have the important result that the sum of statistically independent and identically distributed random variables with finite mean and variance approaches a gaussian cdf as $n \to \infty$. This result is known as the *central limit theorem*.

Although we assumed that the random variables in the sum are identically distributed, the assumption can be relaxed provided that additional restrictions are imposed on the properties of the random variables. There is one variation of the theorem, for example, in which the assumption of identically distributed random variables is abandoned in favor of a condition on the third absolute moment of the random variables in the sum. For a discussion of this and other variations of the central limit theorem, the reader is referred to the book by Cramer (1946).

## 2-2 STOCHASTIC PROCESSES

Many of the random phenomena that occur in nature are functions of time. For example, the meteorological phenomena such as the random fluctuations in air temperature and air pressure are functions of time. The thermal noise voltages generated in the resistors of an electronic device such as a radio receiver are also a function of time. Similarly, the signal at the output of a source that generates information is characterized as a random signal that varies with time. An audio signal that is transmitted over a telephone channel is an example of such a signal. All these are examples of stochastic (random) processes. In our study of digital communications, we encounter stochastic processes in the characterization and modeling of signals generated by information sources, in the characterization of communication channels used to transmit the information, in the characterization of noise generated in a receiver, and in the design of the optimum receiver for processing the received random signal.

At any given time instant, the value of a stochastic process, whether it is the value of the noise voltage generated by a resistor or the amplitude of the signal generated by an audio source, is a random variable. Thus, we may view a stochastic process as a random variable indexed by the parameter $t$. We shall denote such a process by $X(t)$. In general, the parameter $t$ is continuous, whereas $X$ may be either continuous or discrete, depending on the characteristics of the source that generates the stochastic process.

The noise voltage generated by a single resistor or a single information source represents a single realization of the stochastic process. Hence, it is called a *sample function* of the stochastic process. The set of all possible sample functions, e.g., the set of all noise voltage waveforms generated by resistors, constitute an ensemble of sample functions or, equivalently, the stochastic process $X(t)$. In general, the number of sample functions in the ensemble is assumed to be extremely large; often it is infinite.

Having defined a stochastic process $X(t)$ as an ensemble of sample functions, we may consider the values of the process at any set of time instants $t_1 > t_2 > t_3 > \ldots > t_n$ where $n$ is any positive integer. In general, the random variables $X_{t_i} \equiv X(t_i)$, $i = 1, 2, \ldots, n$, are characterized statistically by their joint pdf $p(x_{t_1}, x_{t_2}, \ldots, x_{t_n})$. Furthermore, all the probabilistic relations defined in Section 2-1 for multidimensional random variables carry over to the random variables $X_{t_i}$, $i = 1, 2, \ldots, n$.

**Stationary Stochastic Processes** As indicated above, the random variables $X_{t_i}$, $i = 1, 2, \ldots, n$, obtained from the stochastic process $X(t)$ for any set of time instants $t_1 > t_2 > t_3 > \ldots > t_n$ and any $n$ are characterized statistically by the joint pdf $p(x_{t_1}, x_{t_2}, \ldots, x_{t_n})$. Let us consider another set of $n$ random variables $X_{t_i+t} \equiv X(t_i + t)$, $i = 1, 2, \ldots, n$, where $t$ is an arbitrary time shift. These random variables are characterized by the joint pdf $p(x_{t_1+t}, x_{t_2+t}, \ldots, x_{t_n+t})$. The joint pdfs of the random variables $X_{t_i}$ and $X_{t_i+t}$, $i = 1, 2, \ldots, n$, may or may not be identical. When they are identical, i.e., when

$$p(x_{t_1}, x_{t_2}, \ldots, x_{t_n}) = p(x_{t_1+t}, x_{t_2+t}, \ldots, x_{t_n+t}) \tag{2-2-1}$$

for all $t$ and all $n$, the stochastic process is said to be *stationary in the strict sense*. That is, the statistics of a stationary stochastic process are invariant to any translation of the time axis. On the other hand, when the joint pdfs are different, the stochastic process is *nonstationary*.

### 2-2-1 Statistical Averages

Just as we have defined statistical averages for random variables, we may similarly define statistical averages for a stochastic process. Such averages are also called *ensemble averages*. Let $X(t)$ denote a random process and let $X_{t_i} \equiv X(t_i)$. The *nth moment* of the random variable $X_{t_i}$ is defined as

$$E(X_{t_i}^n) = \int_{-\infty}^{\infty} x_{t_i}^n p(x_{t_i})\, dx_{t_i} \tag{2-2-2}$$

In general, the value of the $n$th moment will depend on the time instant $t_i$ if the pdf of $X_{t_i}$ depends on $t_i$. When the process is stationary, however, $p(x_{t_i+t}) = p(x_{t_i})$ for all $t$. Hence, the pdf is independent of time, and, as a consequence, the $n$th moment is independent of time.

Next we consider the two random variables $X_{t_i} \equiv X(t_i)$, $i = 1, 2$. The correlation between $X_{t_1}$ and $X_{t_2}$ is measured by the joint moment

$$E(X_{t_1}X_{t_2}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_{t_1}x_{t_2}p(x_{t_1}, x_{t_2})\, dx_{t_1}\, dx_{t_2} \tag{2-2-3}$$

Since this joint moment depends on the time instants $t_1$ and $t_2$, it is denoted by $\phi(t_1, t_2)$. The function $\phi(t_1, t_2)$ is called the *autocorrelation function* of the stochastic process. When the process $X(t)$ is stationary, the joint pdf of the pair $(X_{t_1}, X_{t_2})$ is identical to the joint pdf of the pair $(X_{t_1+t}, X_{t_2+t})$ for any arbitrary $t$. This implies that the autocorrelation function of $X(t)$ does not depend on the specific time instants $t_1$ and $t_2$, but, instead, it depends on the time difference $t_1 - t_2$. Thus, for a stationary stochastic process, the joint moment in (2-2-3) is

$$E(X_{t_1}X_{t_2}) = \phi(t_1, t_2) = \phi(t_1 - t_2) = \phi(\tau) \tag{2-2-4}$$

where $\tau = t_1 - t_2$ or, equivalently, $t_2 = t_1 - \tau$. If we let $t_2 = t_1 + \tau$, we have

$$\phi(-\tau) = E(X_{t_1}X_{t_1+\tau}) = E(X_{t_1'}X_{t_1'-\tau}) = \phi(\tau)$$

Therefore, $\phi(\tau)$ is an even function. We also note that $\phi(0) = E(X_t^2)$ denotes the average power in the process $X(t)$.

There exist nonstationary processes with the property that the mean value of the process is independent of time (a constant) and where the autocorrelation function satisfies the condition that $\phi(t_1, t_2) = \phi(t_1 - t_2)$. Such a process is called *wide-sense stationary*. Consequently, wide-sense stationarity is a less stringent condition than strict-sense stationarity. When reference is made to a stationary stochastic process in any subsequent discussion in which correlation functions are involved, the less stringent condition (wide-sense stationarity) is implied.

Related to the autocorrelation function is the autocovariance function of a stochastic process, which is defined as

$$\begin{aligned}\mu(t_1, t_2) &= E\{[X_{t_1} - m(t_1)][X_{t_2} - m(t_2)]\} \\ &= \phi(t_1, t_2) - m(t_1)m(t_2)\end{aligned} \tag{2-2-5}$$

where $m(t_1)$ and $m(t_2)$ are the means of $X_{t_1}$ and $X_{t_2}$, respectively. When the process is stationary, the autocovariance function simplifies to

$$\mu(t_1, t_2) = \mu(t_1 - t_2) = \mu(\tau) = \phi(\tau) - m^2 \tag{2-2-6}$$

where $\tau = t_1 - t_2$.

Higher-order joint moments of two or more random variables derived from a stochastic process $X(t)$ are defined in an obvious manner. With the possible exception of the gaussian random process, for which higher-order moments can be expressed in terms of first and second moments, high-order moments are encountered very infrequently in practice.

**Averages for a Gaussian Process** Suppose that $X(t)$ is a gaussian random process. Hence, at time instants $t = t_i$, $i = 1, 2, \ldots, n$, the random variables $X_{t_i}$, $i = 1, 2, \ldots, n$, are jointly gaussian with mean values $m(t_i)$, $i = 1, 2, \ldots, n$, and autocovariances

$$\mu(t_i, t_j) = E[(X_{t_i} - m(t_i))(X_{t_j} - m(t_j))], \quad i, j = 1, 2, \ldots, n \tag{2-2-7}$$

If we denote the $n \times n$ covariance matrix with elements $\mu(t_i, t_j)$ by $\mathbf{M}$ and the vector of mean values by $\mathbf{m}_x$, then the joint pdf of the random variables $X_{t_i}$, $i = 1, 2, \ldots, n$ is given by (2-1-150).

If the gaussian process is stationary then $m(t_i) = m$ for all $t_i$ and $\mu(t_i, t_j) = \mu(t_i - t_j)$. We observe that the gaussian random process is completely specified by the mean and autocovariance functions. Since the joint gaussian pdf depends only on these two moments, it follows that if the gaussian process is wide-sense stationary, it is also strict-sense stationary. Of course, the converse is always true for any stochastic process.

**Averages for Joint Stochastic Processes** Let $X(t)$ and $Y(t)$ denote two stochastic processes and let $X_{t_i} \equiv X(t_i)$, $i = 1, 2, \ldots, n$, and $Y_{t'_j} \equiv Y(t'_j)$, $j = 1, 2, \ldots, m$, represent the random variables at times $t_1 > t_2 > t_3 > \ldots > t_n$ and $t'_1 > t'_2 > \ldots > t'_m$, respectively. The two processes are characterized statistically by their joint pdf

$$p(x_{t_1}, x_{t_2}, \ldots, x_{t_n}, y_{t'_1}, y_{t'_2}, \ldots, y_{t'_m})$$

for any set of time instants $t_1, t_2, \ldots, t_n$, $t'_1, t'_2, \ldots, t'_m$ and for any positive integer values of $n$ and $m$.

The *cross-correlation function* of $X(t)$ and $Y(t)$, denoted by $\phi_{xy}(t_1, t_2)$, is defined as the joint moment

$$\phi_{xy}(t_1, t_2) = E(X_{t_1} Y_{t_2}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_{t_1} y_{t_2} p(x_{t_1}, y_{t_2})\, dx_{t_1}\, dy_{t_2} \tag{2-2-8}$$

and the *cross-covariance* is

$$\mu_{xy}(t_1, t_2) = \phi_{xy}(t_1, t_2) - m_x(t_1) m_y(t_2) \tag{2-2-9}$$

When the processes are jointly and individually stationary, we have $\phi_{xy}(t_1, t_2) = \phi_{xy}(t_1 - t_2)$ and $\mu_{xy}(t_1, t_2) = \mu_{xy}(t_1 - t_2)$. In this case, we note that

$$\phi_{xy}(-\tau) = E(X_{t_1} Y_{t_1+\tau}) = E(X_{t_1'-\tau} Y_{t_1'}) = \phi_{yx}(\tau) \tag{2-2-10}$$

The stochastic processes $X(t)$ and $Y(t)$ are said to be *statistically independent* if and only if

$$p(x_{t_1}, x_{t_2}, \ldots, x_{t_n}, y_{t_1'}, y_{t_2'}, \ldots, y_{t_m'}) = p(x_{t_1}, x_{t_2}, \ldots, x_{t_n}) p(y_{t_1'}, y_{t_2'}, \ldots, y_{t_m'})$$

for all choices of $t_i$ and $t_i'$ and for all positive integers $n$ and $m$. The processes are said to be *uncorrelated* if

$$\phi_{xy}(t_1, t_2) = E(X_{t_1}) E(Y_{t_2})$$

Hence,

$$\mu_{xy}(t_1, t_2) = 0$$

A *complex-valued stochastic process* $Z(t)$ is defined as

$$Z(t) = X(t) + jY(t) \tag{2-2-11}$$

where $X(t)$ and $Y(t)$ are stochastic processes. The joint pdf of the random variables $Z_{t_i} \equiv Z(t_i)$, $i = 1, 2, \ldots$, is given by the joint pdf of the components $(X_{t_i}, Y_{t_i})$, $i = 1, 2, \ldots, n$. Thus, the pdf that characterizes $Z_{t_i}$, $i = 1, 2, \ldots, n$, is

$$p(x_{t_1}, x_{t_2}, \ldots, x_{t_n}, y_{t_1}, y_{t_2}, \ldots, y_{t_n})$$

The complex-valued stochastic process $Z(t)$ is encountered in the representation of narrowband bandpass noise in terms of its equivalent lowpass components. An important characteristic of such a process is its autocorrelation function. The function is defined as

$$\begin{aligned}
\phi_{zz}(t_1, t_2) &= \tfrac{1}{2} E(Z_{t_1} Z_{t_2}^*) \\
&= \tfrac{1}{2} E[(X_{t_1} + jY_{t_1})(X_{t_2} - jY_{t_2})] \\
&= \tfrac{1}{2}\{\phi_{xx}(t_1, t_2) + \phi_{yy}(t_1, t_2) + j[\phi_{yx}(t_1, t_2) - \phi_{xy}(t_1, t_2)]\}
\end{aligned} \tag{2-2-12}$$

where $\phi_{xx}(t_1, t_2)$ and $\phi_{yy}(t_1, t_2)$ are the autocorrelation functions of $X(t)$ and $Y(t)$, respectively, and $\phi_{yx}(t_1, t_2)$ and $\phi_{xy}(t_1, t_2)$ are the cross-correlation functions. The factor of $\frac{1}{2}$ in the definition of the autocorrelation function of a complex-valued stochastic process is an arbitrary but mathematically convenient normalization factor, as we will demonstrate in our treatment of such processes in Chapter 4.

When the processes $X(t)$ and $Y(t)$ are jointly and individually stationary, the autocorrelation function of $Z(t)$ becomes

$$\phi_{zz}(t_1, t_2) = \phi_{zz}(t_1 - t_2) = \phi_{zz}(\tau)$$

where $t_2 = t_1 - \tau$. Also, the complex conjugate of (2-2-12) is

$$\phi_{zz}^*(\tau) = \tfrac{1}{2} E(Z_{t_1}^* Z_{t_1-\tau}) = \tfrac{1}{2} E(Z_{t_1'+\tau}^* Z_{t_1'}) = \phi_{zz}(-\tau) \tag{2-2-13}$$

Hence, $\phi_{zz}(\tau) = \phi_{zz}^*(-\tau)$.

Now, suppose that $Z(t) = X(t) + jY(t)$ and $W(t) = U(t) + jV(t)$ are two complex-valued stochastic processes. The cross-correlation function of $Z(t)$ and $W(t)$ is defined as

$$\begin{aligned}\phi_{zw}(t_1, t_2) &= \tfrac{1}{2}E(Z_{t_1} W_{t_2}^*) \\ &= \tfrac{1}{2}E[(X_{t_1} + jY_{t_1})(U_{t_2} - jV_{t_2})] \\ &= \tfrac{1}{2}\{\phi_{xu}(t_1, t_2) + \phi_{yv}(t_1, t_2) + j[\phi_{yu}(t_1, t_2) - \phi_{xv}(t_1, t_2)]\}\end{aligned} \tag{2-2-14}$$

When $X(t)$, $Y(t)$, $U(t)$, and $V(t)$ are pairwise-stationary, the cross-correlation functions in (2-2-14) become functions of the time difference $\tau = t_1 - t_2$. Furthermore,

$$\phi_{zw}^*(\tau) = \tfrac{1}{2}E(Z_{t_1}^* W_{t_1 - \tau}) = \tfrac{1}{2}E(Z_{t_1' + \tau}^* W_{t_1'}) = \phi_{wz}(-\tau) \tag{2-2-15}$$

## 2-2-2 Power Density Spectrum

The frequency content of a signal is a very basic characteristic that distinguishes one signal from another. In general, a signal can be classified as having either a finite (nonzero) average power (infinite energy) or finite energy. The frequency content of a finite energy signal is obtained as the Fourier transform of the corresponding time function. If the signal is periodic, its energy is infinite and, consequently, its Fourier transform does not exist. The mechanism for dealing with periodic signals is to represent them in a Fourier series. With such a representation, the Fourier coefficients determine the distribution of power at the various discrete frequency components.

A stationary stochastic process is an infinite energy signal, and, hence, its Fourier transform does not exist. The spectral characteristic of a stochastic signal is obtained by computing the Fourier transform of the autocorrelation function. That is, the distribution of power with frequency is given by the function

$$\Phi(f) = \int_{-\infty}^{\infty} \phi(\tau) e^{-j2\pi f \tau}\, d\tau \tag{2-2-16}$$

The inverse Fourier transform relationship is

$$\phi(\tau) = \int_{-\infty}^{\infty} \Phi(f) e^{j2\pi f \tau}\, df \tag{2-2-17}$$

We observe that

$$\begin{aligned}\phi(0) &= \int_{-\infty}^{\infty} \Phi(f)\, df \\ &= E(|X_t|^2) \geq 0\end{aligned} \tag{2-2-18}$$

Since $\phi(0)$ represents the average power of the stochastic signal, which is the area under $\Phi(f)$, $\Phi(f)$ is the distribution of power as a function of frequency. Therefore, $\Phi(f)$ is called the *power density spectrum* of the stochastic process.

If the stochastic process is real, $\phi(\tau)$ is real and even, and, hence $\Phi(f)$ is real and even. On the other hand, if the process is complex, $\phi(\tau) = \phi^*(-\tau)$ and, hence

$$\Phi^*(f) = \int_{-\infty}^{\infty} \phi^*(\tau) e^{j2\pi f\tau}\, d\tau = \int_{-\infty}^{\infty} \phi^*(-\tau) e^{-j2\pi f\tau}\, d\tau$$
$$= \int_{-\infty}^{\infty} \phi(\tau) e^{-j2\pi f\tau}\, d\tau = \Phi(f) \tag{2-2-19}$$

Therefore, $\Phi(f)$ is real.

The definition of a power density spectrum can be extended to two jointly stationary stochastic processes $X(t)$ and $Y(t)$, which have a cross-correlation function $\phi_{xy}(\tau)$. The Fourier transform of $\phi_{xy}(\tau)$, i.e.,

$$\Phi_{xy}(f) = \int_{-\infty}^{\infty} \phi_{xy}(\tau) e^{-j2\pi f\tau}\, d\tau \tag{2-2-20}$$

is called the *cross-power density spectrum*. If we conjugate both sides of (2-2-20), we have

$$\Phi_{xy}^*(f) = \int_{-\infty}^{\infty} \phi_{xy}^*(\tau) e^{j2\pi f\tau}\, d\tau = \int_{-\infty}^{\infty} \phi_{xy}^*(-\tau) e^{-j2\pi f\tau}\, d\tau$$
$$= \int_{-\infty}^{\infty} \phi_{yx}(\tau) e^{-j2\pi f\tau}\, d\tau = \Phi_{yx}(f) \tag{2-2-21}$$

This relation holds in general. However, if $X(t)$ and $Y(t)$ are real stochastic processes,

$$\Phi_{xy}^*(f) = \int_{-\infty}^{\infty} \phi_{xy}(\tau) e^{j2\pi f\tau}\, d\tau = \Phi_{xy}(-f) \tag{2-2-22}$$

By combining the result in (2-2-21) with the result in (2-2-22), we find that the cross-power density spectrum of two real processes satisfies the condition

$$\Phi_{yx}(f) = \Phi_{xy}(-f) \tag{2-2-23}$$

### 2-2-3 Response of a Linear Time-Invariant System to a Random Input Signal

Consider a linear time-invariant system (filter) that is characterized by its impulse response $h(t)$ or, equivalently, by its frequency response $H(f)$, where $h(t)$ and $H(f)$ are a Fourier transform pair. Let $x(t)$ be the input signal to the system and let $y(t)$ denote the output signal. The output of the system may be expressed in terms of the convolution integral as

$$y(t) = \int_{-\infty}^{\infty} h(\tau) x(t-\tau)\, d\tau \tag{2-2-24}$$

Now, suppose that $x(t)$ is a sample function of a stationary stochastic process $X(t)$. Then, the output $y(t)$ is a sample function of a stochastic process $Y(t)$. We wish to determine the mean and autocorrelation functions of the output.

Since convolution is a linear operation performed on the input signal $x(t)$, the expected value of the integral is equal to the integral of the expected value. Thus, the mean value of $Y(t)$ is

$$\begin{aligned} m_y = E[Y(t)] &= \int_{-\infty}^{\infty} h(\tau)E[X(t-\tau)]\,d\tau \\ &= m_x \int_{-\infty}^{\infty} h(\tau)\,d\tau = m_x H(0) \end{aligned} \tag{2-2-25}$$

where $H(0)$ is the frequency response of the linear system at $f = 0$. Hence, the mean value of the output process is a constant.

The autocorrelation function of the output is

$$\begin{aligned} \phi_{yy}(t_1, t_2) &= \tfrac{1}{2}E(Y_{t_1}Y_{t_2}^*) \\ &= \frac{1}{2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(\beta)h^*(\alpha)E[X(t_1-\beta)X^*(t_2-\alpha)]\,d\alpha\,d\beta \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(\beta)h^*(\alpha)\phi_{xx}(t_1-t_2+\alpha-\beta)\,d\alpha\,d\beta \end{aligned}$$

The last step indicates that the double integral is a function of the time difference $t_1 - t_2$. In other words, if the input process is stationary, the output is also stationary. Hence

$$\phi_{yy}(\tau) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h^*(\alpha)h(\beta)\phi_{xx}(\tau+\alpha-\beta)\,d\alpha\,d\beta \tag{2-2-26}$$

By evaluating the Fourier transform of both sides of (2-2-26), we obtain the power density spectrum of the output process in the form

$$\begin{aligned} \Phi_{yy}(f) &= \int_{-\infty}^{\infty} \phi_{yy}(\tau)e^{-j2\pi f\tau}\,d\tau \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h^*(\alpha)h(\beta)\phi_{xx}(\tau+\alpha-\beta)e^{-j2\pi f\tau}\,d\tau\,d\alpha\,d\beta \\ &= \Phi_{xx}(f)\,|H(f)|^2 \end{aligned} \tag{2-2-27}$$

Thus, we have the important result that the power density spectrum of the output signal is the product of the power density spectrum of the input multiplied by the magnitude squared of the frequency response of the system.

When the autocorrelation function $\phi_{yy}(\tau)$ is desired, it is usually easier to determine the power density spectrum $\Phi_{yy}(f)$ and then to compute the inverse transform. Thus, we have

$$\begin{aligned}\phi_{yy}(\tau) &= \int_{-\infty}^{\infty} \Phi_{yy}(f) e^{j2\pi f\tau}\, df \\ &= \int_{-\infty}^{\infty} \Phi_{xx}(f)\, |H(f)|^2\, e^{j2\pi f\tau}\, df \end{aligned} \tag{2-2-28}$$

We observe that the average power in the output signal is

$$\phi_{yy}(0) = \int_{-\infty}^{\infty} \Phi_{xx}(f)\, |H(f)|^2\, df \tag{2-2-29}$$

Since $\phi_{yy}(0) = E(|Y_t|^2)$, it follows that

$$\int_{-\infty}^{\infty} \Phi_{xx}(f)\, |H(f)|^2\, df \geq 0$$

Suppose we let $|H(f)|^2 = 1$ for any arbitrarily small interval $f_1 \leq f \leq f_2$, and $H(f) = 0$ outside this interval. Then,

$$\int_{f_1}^{f_2} \Phi_{xx}(f)\, df \geq 0$$

But this is possible if and only if $\Phi_{xx}(f) \geq 0$ for all $f$.

**Example 2-2-1**

Suppose that the lowpass filter illustrated in Fig. 2-2-1 is excited by a stochastic process $x(t)$ having a power density spectrum

$$\Phi_{xx}(f) = \tfrac{1}{2}N_0 \quad \text{for all } f$$

A stochastic process having a flat power density spectrum is called *white*

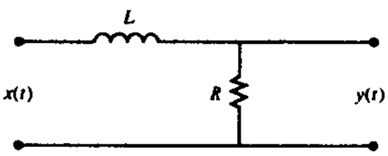



**FIGURE 2-2-1** An example of a lowpass filter.

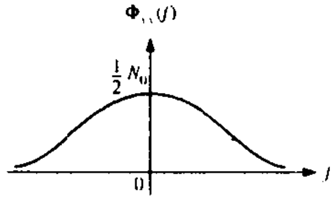



**FIGURE 2-2-2** The power density spectrum of the lowpass filter output when the input is white noise.

*noise*. Let us determine the power density spectrum of the output process. The transfer function of the lowpass filter is

$$H(f) = \frac{R}{R + j2\pi fL} = \frac{1}{1 + j2\pi fL/R}$$

and, hence,

$$|H(f)|^2 = \frac{1}{1 + (2\pi L/R)^2 f^2} \tag{2-2-30}$$

The power density spectrum of the output process is

$$\Phi_{yy}(f) = \frac{N_0}{2} \frac{1}{1 + (2\pi L/R)^2 f^2} \tag{2-2-31}$$

This power density spectrum is illustrated in Fig. 2-2-2. Its inverse Fourier transform yields the autocorrelation function

$$\begin{aligned}\phi_{yy}(\tau) &= \int_{-\infty}^{\infty} \frac{N_0}{2} \frac{1}{1 + (2\pi L/R)^2 f^2} e^{j2\pi f\tau}\, df \\ &= \frac{RN_0}{4L} e^{-(R/L)|\tau|} \end{aligned} \tag{2-2-32}$$

The autocorrelation function $\phi_{yy}(\tau)$ is shown in Fig. 2-2-3. We observe that the second moment of the process $Y(t)$ is $\phi_{yy}(0) = RN_0/4L$.

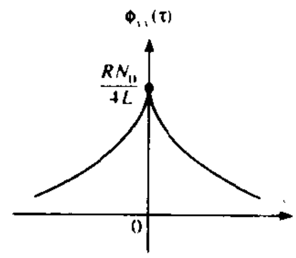



**FIGURE 2-2-3** The autocorrelation function of the output of the lowpass filter for a white-noise input.

As a final exercise, we determine the cross-correlation function between $y(t)$ and $x(t)$, where $x(t)$ denotes the input and $y(t)$ denotes the output of the linear system. We have

$$\begin{aligned}\phi_{yx}(t_1, t_2) = \tfrac{1}{2}E(Y_{t_1}X_{t_2}^*) &= \frac{1}{2}\int_{-\infty}^{\infty} h(\alpha)E[X(t_1-\alpha)X^*(t_2)]\,d\alpha \\ &= \int_{-\infty}^{\infty} h(\alpha)\phi_{xx}(t_1-t_2-\alpha)\,d\alpha = \phi_{yx}(t_1-t_2)\end{aligned}$$

Hence, the stochastic processes $X(t)$ and $Y(t)$ are jointly stationary. With $t_1 - t_2 = \tau$, we have

$$\phi_{yx}(\tau) = \int_{-\infty}^{\infty} h(\alpha)\phi_{xx}(\tau-\alpha)\,d\alpha \tag{2-2-33}$$

Note that the integral in (2-2-33) is a convolution integral. Hence in the frequency domain the relation (2-2-33) becomes

$$\Phi_{yx}(f) = \Phi_{xx}(f)H(f) \tag{2-2-34}$$

We observe that if the input process is white noise, the cross correlation of the input with the output of the system yields the impulse response $h(t)$ to within a scale factor.

### 2-2-4 Sampling Theorem for Band-Limited Stochastic Processes

Recall that a deterministic signal $s(t)$ that has a Fourier transform $S(f)$ is called band-limited if $S(f) = 0$ for $|f| > W$, where $W$ is the highest frequency contained in $s(t)$. Such a signal is uniquely represented by samples of $s(t)$ taken at a rate of $f_s \geqslant 2W$ samples/s. The minimum rate $f_N = 2W$ samples/s is called the *Nyquist rate*. Sampling below the Nyquist rate results in frequency aliasing.

The band-limited signal sampled at the Nyquist rate can be reconstructed from its samples by use of the interpolation formula

$$s(t) = \sum_{n=-\infty}^{\infty} s\left(\frac{n}{2W}\right)\frac{\sin\left[2\pi W\left(t-\dfrac{n}{2W}\right)\right]}{2\pi W\left(t-\dfrac{n}{2W}\right)} \tag{2-2-35}$$

where $\{s(n/2W)\}$ are the samples of $s(t)$ taken at $t = n/2W$, $n = 0, \pm 1, \pm 2, \ldots$. Equivalently, $s(t)$ can be reconstructed by passing the sampled signal through

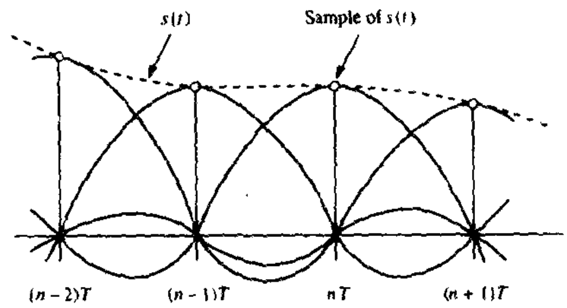



**FIGURE 2-2-4** Signal reconstruction based on ideal interpolation.

an ideal low-pass filter with impulse response $h(t) = (\sin 2\pi Wt)/2\pi Wt$. Figure 2-2-4 illustrates the signal reconstruction process based on ideal interpolation.

A stationary stochastic process $X(t)$ is said to be *band-limited* if its power density spectrum $\Phi(f) = 0$ for $|f| > W$. Since $\Phi(f)$ is the Fourier transform of the autocorrelation function $\phi(\tau)$, it follows that $\phi(\tau)$ can be represented as

$$\phi(\tau) = \sum_{n=-\infty}^{\infty} \phi\left(\frac{n}{2W}\right) \frac{\sin\left[2\pi W\left(\tau - \dfrac{n}{2W}\right)\right]}{2\pi W\left(\tau - \dfrac{n}{2W}\right)} \tag{2-2-36}$$

where $\{\phi(n/2W)\}$ are samples of $\phi(\tau)$ taken at $\tau = n/2W$, $n = 0, \pm 1, \pm 2, \ldots$.

Now, if $X(t)$ is a band-limited stationary stochastic process then $X(t)$ can be represented as

$$X(t) = \sum_{n=-\infty}^{\infty} X\left(\frac{n}{2W}\right) \frac{\sin\left[2\pi W\left(t - \dfrac{n}{2W}\right)\right]}{2\pi W\left(t - \dfrac{n}{2W}\right)} \tag{2-2-37}$$

where $\{X(n/2W)\}$ are samples of $X(t)$ taken at $t = n/2W$, $n = 0, \pm 1, \pm 2, \ldots$. This is the sampling representation for a stationary stochastic process. The samples are random variables that are described statistically by appropriate joint probability density functions. The signal representation in (2-2-37) is easily established by showing that (Problem 2-17)

$$E\left\{\left|X(t) - \sum_{n=-\infty}^{\infty} X\left(\frac{n}{2W}\right) \frac{\sin\left[2\pi W\left(t - \dfrac{n}{2W}\right)\right]}{2\pi W\left(t - \dfrac{n}{2W}\right)}\right|^2\right\} = 0 \tag{2-2-38}$$

Hence, equality between the sampling representation and the stochastic process $X(t)$ holds in the sense that the mean square error is zero.

### 2-2-5 Discrete-Time Stochastic Signals and Systems

The characterization of continuous-time stochastic signals given above can be easily carried over to discrete-time stochastic signals. Such signals are usually obtained by uniformly sampling a continuous-time stochastic process.

A discrete-time stochastic process $X(n)$ consists of an ensemble of sample sequences $\{x(n)\}$. The statistical properties of $X(n)$ are similar to the characterization of $X(t)$ with the restriction that $n$ is now an integer (time) variable. Hence, the *mth moment* of $X(n)$ is defined as

$$E[X_n^m] = \int_{-\infty}^{\infty} X_n^m p(X_n)\, dX_n \tag{2-2-39}$$

and the *autocorrelation sequence* is

$$\phi(n, k) = \tfrac{1}{2}E(X_n X_k^*) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} X_n X_k^* p(X_n, X_k)\, dX_n\, dX_k \tag{2-2-40}$$

Similarly, the autocovariance sequence is

$$\mu(n, k) = \phi(n, k) - E(X_n)E(X_k^*) \tag{2-2-41}$$

For a stationary process, we have $\phi(n, k) \equiv \phi(n - k)$, $\mu(n, k) \equiv \mu(n - k)$, and

$$\mu(n - k) = \phi(n - k) - |m_x|^2 \tag{2-2-42}$$

where $m_x = E(X_n)$ is the mean value.

As in the case of continuous-time stochastic processes, a discrete-time stationary process has infinite energy but a finite average power, which is given as

$$E(|X_n|^2) = \phi(0) \tag{2-2-43}$$

The power density spectrum for the discrete-time process is obtained by computing the Fourier transform of $\phi(n)$. Since $\phi(n)$ is a discrete-time sequence, the Fourier transform is defined as

$$\Phi(f) = \sum_{n=-\infty}^{\infty} \phi(n) e^{-j2\pi fn} \tag{2-2-44}$$

and the inverse transform relationship is

$$\phi(n) = \int_{-1/2}^{1/2} \Phi(f) e^{j2\pi fn}\, df \tag{2-2-45}$$

We make the observation that the power density spectrum $\Phi(f)$ is periodic with a period $f_p = 1$. In other words, $\Phi(f + k) = \Phi(f)$ for $k = \pm 1, \pm 2, \ldots$. This is a characteristic of the Fourier transform of any discrete-time sequence such as $\phi(n)$.

Finally, let us consider the response of a discrete-time, linear time-invariant system to a stationary stochastic input signal. The system is characterized in the time domain by its unit sample response $h(n)$ and in the frequency domain by the frequency response $H(f)$, where

$$H(f) = \sum_{n=-\infty}^{\infty} h(n)e^{-j2\pi fn} \tag{2-2-46}$$

The response of the system to the stationary stochastic input signal $X(n)$ is given by the convolution sum

$$y(n) = \sum_{k=-\infty}^{\infty} h(k)x(n-k) \tag{2-2-47}$$

The mean value of the output of the system is

$$\begin{aligned} m_y &= E[y(n)] = \sum_{k=-\infty}^{\infty} h(k)E[x(n-k)] \\ m_y &= m_x \sum_{k=-\infty}^{\infty} h(k) = m_x H(0) \end{aligned} \tag{2-2-48}$$

where $H(0)$ is the zero frequency (dc) gain of the system.

The autocorrelation sequence for the output process is

$$\begin{aligned} \phi_{yy}(k) &= \tfrac{1}{2}E[y^*(n)y(n+k)] \\ &= \tfrac{1}{2}\sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty} h^*(i)h(j)E[x^*(n-i)x(n+k-j)] \\ &= \sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty} h^*(i)h(j)\phi_{xx}(k-j+i) \end{aligned} \tag{2-2-49}$$

This is the general form for the autocorrelation sequence of the system output in terms of the autocorrelation of the system input and the unit sample response of the system. By taking the Fourier transform of $\phi_{yy}(k)$ and substituting the relation in (2-2-49), we obtain the corresponding frequency domain relationship

$$\Phi_{yy}(f) = \Phi_{xx}(f)\,|H(f)|^2 \tag{2-2-50}$$

which is identical to (2-2-27) except that in (2-2-50) the power density spectra $\Phi_{yy}(f)$ and $\Phi_{xx}(f)$ and the frequency response $H(f)$ are periodic functions of frequency with period $f_p = 1$.

### 2-2-6 Cyclostationary Processes

In dealing with signals that carry digital information we encounter stochastic processes that have statistical averages that are periodic. To be specific, let us consider a stochastic process of the form

$$X(t) = \sum_{n=-\infty}^{\infty} a_n g(t-nT) \tag{2-2-51}$$

where $\{a_n\}$ is a (discrete-time) sequence of random variables with mean $m_a = E(a_n)$ for all $n$ and autocorrelation sequence $\phi_{aa}(k) = \frac{1}{2}E(a_n^* a_{n+k})$. The signal $g(t)$ is deterministic. The stochastic process $X(t)$ represents the signal for several different types of linear modulation techniques which are introduced in Chapter 4. The sequence $\{a_n\}$ represents the digital information sequence (of symbols) that is transmitted over the communication channel and $1/T$ represents the rate of transmission of the information symbols.

Let us determine the mean and autocorrelation function of $X(t)$. First, the mean value is

$$\begin{aligned} E[X(t)] &= \sum_{n=-\infty}^{\infty} E(a_n) g(t - nT) \\ &= m_a \sum_{n=-\infty}^{\infty} g(t - nT) \end{aligned} \tag{2-5-52}$$

We observe that the mean is time-varying. In fact, it is periodic with period $T$.

The autocorrelation function of $X(t)$ is

$$\begin{aligned} \phi_{xx}(t+\tau, t) &= \tfrac{1}{2}E[X(t+\tau)X^*(t)] \\ &= \tfrac{1}{2} \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} E(a_n^* a_m) g^*(t - nT) g(t + \tau - mT) \\ &= \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \phi_{aa}(m-n) g^*(t - nT) g(t + \tau - mT) \end{aligned} \tag{2-2-53}$$

Again, we observe that

$$\phi_{xx}(t + \tau + kT, t + kT) = \phi_{xx}(t + \tau, t) \tag{2-2-54}$$

for $k = \pm 1, \pm 2, \ldots$. Hence, the autocorrelation function of $X(t)$ is also periodic with period $T$.

Such a stochastic process is called *cyclostationary* or *periodically stationary*. Since the autocorrelation function depends on both the variables $t$ and $\tau$, its frequency domain representation requires the use of a two-dimensional Fourier transform.

Since it is highly desirable to characterize such signals by their power density spectrum, an alternative approach is to compute the *time-average autocorrelation function* over a single period, defined as

$$\bar{\phi}_{xx}(\tau) = \frac{1}{T} \int_{-T/2}^{T/2} \phi_{xx}(t + \tau, t)\, dt \tag{2-2-55}$$

Thus, we eliminate the time dependence by dealing with the average autocorrelation function. Now, the fourier transform of $\bar{\phi}_{xx}(\tau)$ yields the *average power density spectrum* of the cyclostationary stochastic process. This approach allows us to simply characterize cyclostationary processes in the frequency domain in terms of the power spectrum. That is, the power density spectrum is

$$\Phi_{xx}(f) = \int_{-\infty}^{\infty} \bar{\phi}_{xx}(\tau) e^{-j2\pi f\tau}\, d\tau \tag{2-2-56}$$

## 2-3 BIBLIOGRAPHICAL NOTES AND REFERENCES

In this chapter we have provided a review of basic concepts and definitions in the theory of probability and stochastic processes. As stated in the opening paragraph, this theory is an important mathematical tool in the statistical modeling of information sources, communication channels, and in the design of digital communication systems. Of particular importance in the evaluation of communication system performance is the Chernoff bound. This bound is frequently used in bounding the probability of error of digital communication systems that employ coding in the transmission of information. Our coverage also highlighted a number of probability distributions and their properties, which are frequently encountered in the design of digital communication systems.

The texts by Davenport and Root (1958), Davenport (1970), Papoulis (1984) Pebbles (1987), Helstrom (1991) and Leon-Garcia (1994) provide engineering-oriented treatments of probability and stochastic processes. A more mathematical treatment of probability theory may be found in the text by Loève (1955). Finally, we cite the book by Miller (1964), which treats multidimensional gaussian distributions.

## PROBLEMS

**2-1** One experiment has four mutually exclusive outcomes $A_i$, $i = 1, 2, 3, 4$, and a second experiment has three mutually exclusive outcomes $B_j$, $j = 1, 2, 3$. The joint probabilities $P(A_i, B_j)$ are

$$\begin{aligned}
&P(A_1, B_1) = 0.10, \quad &&P(A_1, B_2) = 0.08, \quad &&P(A_1, B_3) = 0.13\\
&P(A_2, B_1) = 0.05, &&P(A_2, B_2) = 0.03, &&P(A_2, B_3) = 0.09\\
&P(A_3, B_1) = 0.05, &&P(A_3, B_2) = 0.12, &&P(A_3, B_3) = 0.14\\
&P(A_4, B_1) = 0.11, &&P(A_4, B_2) = 0.04, &&P(A_4, B_3) = 0.06
\end{aligned}$$

Determine the probabilities $P(A_i)$, $i = 1, 2, 3, 4$, and $P(B_j)$, $j = 1, 2, 3$.

**2-2** The random variables $X_i$, $i = 1, 2, \ldots, n$, have the joint pdf $p(x_1, x_2, \ldots, x_n)$. Prove that

$$\begin{aligned}
&p(x_1, x_2, x_3, \ldots, x_n)\\
&\quad = p(x_n \mid x_{n-1}, \ldots, x_1)\, p(x_{n-1} \mid x_{n-2}, \ldots, x_1) \cdots p(x_3 \mid x_2, x_1)\, p(x_2 \mid x_1)\, p(x_1)
\end{aligned}$$

**2-3** The pdf of a random variable $X$ is $p(x)$. A random variable $Y$ is defined as

$$Y = aX + b$$

where $a < 0$. Determine the pdf of $Y$ in terms of the pdf of $X$.

**2-4** Suppose that $X$ is a gaussian random variable with zero mean and unit variance. Let

$$Y = aX^3 + b, \qquad a > 0$$

Determine and plot the pdf of $Y$.

**2-5** **a** Let $X_r$ and $X_i$ be statistically independent zero-mean gaussian random variables with identical variance. Show that a (rotational) transformation of the form

$$Y_r + jY_i = (X_r + jX_i)e^{j\phi}$$

results in another pair $(Y_r, Y_i)$ of gaussian random variables that have the same joint pdf as the pair $(X_r, X_i)$.

**b** Note that

$$\begin{bmatrix} Y_r \\ Y_i \end{bmatrix} = \mathbf{A} \begin{bmatrix} X_r \\ X_i \end{bmatrix}$$

where $\mathbf{A}$ is a $2 \times 2$ matrix. As a generalization of the two-dimensional transformation of the gaussian random variables considered in (a), what property must the linear transformation $\mathbf{A}$ satisfy if the pdfs for $\mathbf{X}$ and $\mathbf{Y}$, where $\mathbf{Y} = \mathbf{AX}$, $\mathbf{X} = (X_1 X_2 \cdots X_n)$ and $\mathbf{Y} = (Y_1 Y_2 \cdots Y_n)$, are identical?

**2-6** The random variable $Y$ is defined as

$$Y = \sum_{i=1}^{n} X_i$$

where the $X_i$, $i = 1, 2, \ldots, n$, are statistically independent random variables with

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

**a** Determine the characteristic function of $Y$.
**b** From the characteristic function, determine the moments $E(Y)$ and $E(Y^2)$.

**2-7** The four random variables $X_1$, $X_2$, $X_3$, $X_4$ are zero-mean jointly gaussian random variables with covariance $\mu_{ij} = E(X_i X_j)$ and characteristic function $\psi(jv_1, jv_2, jv_3, jv_4)$. Show that

$$E(X_1 X_2 X_3 X_4) = \mu_{12}\mu_{34} + \mu_{13}\mu_{24} + \mu_{14}\mu_{23}$$

**2-8** From the characteristic functions for the central chi-square and noncentral chi-square random variables given by (2-1-109) and (2-1-117), respectively, determine the corresponding first and second moments given by (2-1-112) and (2-1-125)

**2-9** The pdf of a Cauchy distributed random variable $X$ is

$$p(x) = \frac{a/\pi}{x^2 + a^2}, \qquad -\infty < x < \infty$$

**a** Determine the mean and variance of $X$.
**b** Determine the characteristic function of $X$.

**2-10** The random variable $Y$ is defined as

$$Y = \frac{1}{n} \sum_{i=1}^{n} X_i$$

where $X_i$, $i = 1, 2, \ldots, n$, are statistically independent and identically distributed random variables each of which has the Cauchy pdf given in Problem 2-9
**a** Determine the characteristic function of $Y$.
**b** Determine the pdf of $Y$.
**c** Consider the pdf of $Y$ in the limit as $n \to \infty$. Does the central limit hold? Explain your answer.

**2-11** Assume that random processes $x(t)$ and $y(t)$ are individually and jointly stationary.
**a** Determine the autocorrelation function of $z(t) = x(t) + y(t)$.
**b** Determine the autocorrelation function of $z(t)$ when $x(t)$ and $y(t)$ are uncorrelated.
**c** Determine the autocorrelation function of $z(t)$ when $x(t)$ and $y(t)$ are uncorrelated and have zero means.

**2-12** The autocorrelation function of a stochastic process $X(t)$ is

$$\phi_{xx}(\tau) = \tfrac{1}{2} N_0 \delta(\tau)$$

Such a process is called *white noise*. Suppose $x(t)$ is the input to an ideal bandpass filter having the frequency response characteristic shown in Fig. P2-12. Determine the total noise power at the output of the filter.

**2-13** The covariance matrix of three random variables $X_1$, $X_2$ and $X_3$ is

$$\begin{bmatrix} \mu_{11} & 0 & \mu_{13} \\ 0 & \mu_{22} & 0 \\ \mu_{31} & 0 & \mu_{33} \end{bmatrix}$$

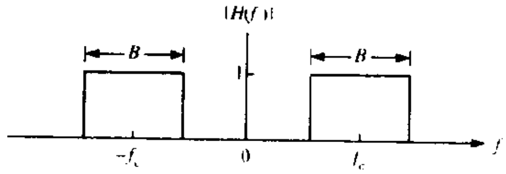



**FIGURE P.2-12**

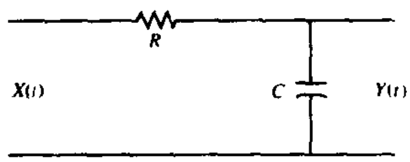



**FIGURE P2-16**

The linear transformation $\mathbf{Y} = \mathbf{AX}$ is made where

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

Determine the covariance matrix of $\mathbf{Y}$.

**2-14** Let $X(t)$ be a stationary real normal process with zero mean. Let a new process $Y(t)$ be defined by

$$Y(t) = X^2(t)$$

Determine the autocorrelation function of $Y(t)$ in terms of the autocorrelation function of $X(t)$. *Hint*: Use the result on gaussian variables derived in Problem 2-7.

**2-15** For the Nakagami pdf, given by (2-1-147), define the normalized random variable $X = R/\sqrt{\Omega}$. Determine the pdf of $X$.

**2-16** The input $X(t)$ in the circuit shown in Fig. P2-16 is a stochastic process with $E[X(t)] = 0$ and $\phi_{xx}(\tau) = \sigma^2\delta(\tau)$, i.e., $X(t)$ is a white noise process.

**a** Determine the spectral density $\Phi_{yy}(f)$.

**b** Determine $\phi_{yy}(\tau)$ and $E[Y^2(t)]$.

**2-17** Demonstrate the validity of (2-2-38).

**2-18** Use the Chernoff bound to show that $Q(x) \leq e^{-x^2/2}$ where $Q(x)$ is defined by (2-1-97).

**2-19** Determine the mean, the autocorrelation sequence, and the power density spectrum of the output of a system with unit sample response

$$h(n) = \begin{cases} 1 & (n = 0) \\ -2 & (n = 1) \\ 1 & (n = 2) \\ 0 & (\text{otherwise}) \end{cases}$$

when the input $x(n)$ is a white-noise process with variance $\sigma_x^2$.

**2-20** The autocorrelation sequence of a discrete-time stochastic process is $\phi(k) = (\frac{1}{2})^{|k|}$. Determine its power density spectrum.

**2-21** A discrete-time stochastic process $X(n) \equiv X(nT)$ is obtained by periodic sampling of a continuous-time zero-mean stationary process $X(t)$ where $T$ is the sampling interval, i.e., $f_s = 1/T$ is the sampling rate.

**a** Determine the relationship between the autocorrelation function of $X(t)$ and the autocorrelation sequence of $X(n)$.

**b** Express the power density spectrum of $X(n)$ in terms of the power density spectrum of the process $X(t)$.

c Determine the conditions under which the power density spectrum of $X(n)$ is equal to the power density spectrum of $X(t)$.

**2-22** Consider a band-limited zero-mean stationary stochastic $X(t)$ with power density spectrum

$$\Phi(f) = \begin{cases} 1 & (|f| \leq W) \\ 0 & (|f| > W) \end{cases}$$

$X(t)$ is sampled at a rate $f_s = 1/T$ to yield a discrete-time process $X(n) \equiv X(nT)$.

a Determine the expression for the autocorrelation sequence of $X(n)$.

b Determine the minimum value of $T$ that results in a white (spectrally flat) sequence.

c Repeat (b) if the power density spectrum of $X(t)$ is

$$\Phi(f) = \begin{cases} 1 - |f|/W & (|f| \leq W) \\ 0 & (|f| > W) \end{cases}$$

**2-23** Show that the functions

$$f_k(t) = \frac{\sin\left[2\pi W\left(t - \dfrac{k}{2W}\right)\right]}{2\pi W\left(t - \dfrac{k}{2W}\right)}, \quad k = 0, \pm 1, \pm 2, \ldots$$

are orthogonal over the interval $[-\infty, \infty]$, i.e.,

$$\int_{-\infty}^{\infty} f_k(t) f_j(t)\, dt = \begin{cases} 1/2W & (k = j) \\ 0 & (k \neq j) \end{cases}$$

Therefore, the sampling theorem reconstruction formula may be viewed as a series expansion of the band-limited signal $s(t)$, where the weights are samples of $s(t)$ and the $\{f_k(t)\}$ are the set of orthogonal functions used in the series expansion.

**2-24** The noise equivalent bandwidth of a system is defined as

$$B_{eq} = \frac{1}{G} \int_0^{\infty} |H(f)|^2\, df$$

where $G = \max |H(f)|^2$. Using this definition, determine the noise equivalent bandwidth of the ideal bandpass filter shown in Fig. P2-12 and the lowpass system shown in Fig. P2-16.