8 FUNDAMENTALS OF PROBABILITY THEORY

Thus far, we have been studying signals whose values at any instant t are determined by their analytical or graphical description. These are called **deterministic** signals, implying complete certainty about their values at any moment t. Such signals, which can be specified with certainty, cannot convey information. It will be seen in Chapter 13 that information is inherently related to uncertainty. The higher the uncertainty about a signal (or message) to be received, the higher its information content. If a message to be received is specified (i.e., if it is known beforehand), then it contains no uncertainty and conveys no new information to the receiver. Hence, signals that convey information must be unpredictable. In addition to information-bearing signals, noise signals that perturb information signals in a system are also unpredictable (otherwise they can simply be subtracted). These unpredictable message signals and noise waveforms are examples of **random processes** that play key roles in communication systems and their analysis.

Random phenomena arise either because of our partial ignorance of the generating mechanism (as in message or noise signals) or because the laws governing the phenomena may be fundamentally random (as in quantum mechanics). Yet in another situation, such as the outcome of rolling a die, it is possible to predict the outcome provided we know exactly all the conditions: the angle of the throw, the nature of the surface on which it is thrown, the force imparted by the player, and so on. The exact analysis, however, is so complex and so sensitive to all the conditions that it is impractical to carry it out, and we are content to accept the outcome prediction on an average basis. Here the random phenomenon arises from our unwillingness to carry out the exact and full analysis because it is impractical to amass all the conditions precisely or not worth the effort.

We shall begin with a review of the basic concepts of the theory of probability, which forms the basis for describing random processes.

8.1 CONCEPT OF PROBABILITY

To begin the discussion of probability, we must define some basic elements and important terms. The term **experiment** is used in probability theory to describe a process whose outcome cannot be fully predicted because the conditions under which it is performed cannot be predetermined with sufficient accuracy and completeness. Tossing a coin, rolling a die, and drawing a card

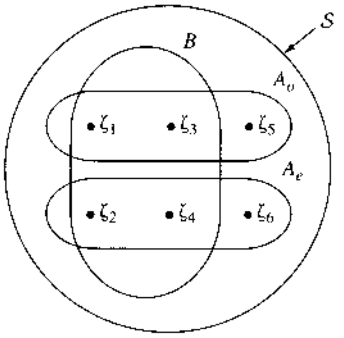


Figure 8.1 Sample space for a throw of a die.

from a deck are some examples of such experiments. An experiment may have several separately identifiable **outcomes**. For example, rolling a die has six possible identifiable outcomes (1, 2, 3, 4, 5, and 6). **An event** is a subset of outcomes that share some common characteristics. An event occurs if the outcome of the experiment belongs to the specific subset of outcomes defining the event. In the experiment of rolling a die, for example, the event "odd number on a throw" can result from any one of three outcomes (viz., 1, 3, and 5). Hence, this event is a set consisting of three outcomes (1, 3, and 5). Thus, events are groupings of outcomes into classes among which we choose to distinguish. The ideas of **experiment**, **outcomes**, and **events** form the basic foundation of probability theory. These ideas can be better understood by using the concepts of set theory.

We define the **sample space** $\mathcal{S}$ as a collection of all possible and separately identifiable outcomes of an experiment. In other words, the **sample space** $\mathcal{S}$ specifies the **experiment**. Each outcome is an **element**, or **sample point**, of this space $\mathcal{S}$ and can be conveniently represented by a point in the sample space. In the experiment of rolling a die, for example, the sample space consists of six elements represented by six sample points ζ_1, ζ_2, ζ_3, ζ_4, ζ_5, and ζ_6, where ζ_i represents the outcome "a number i is thrown" (Fig. 8.1). The event, on the other hand, is a subset of $\mathcal{S}$. The event "an odd number is thrown," denoted by A_o, is a subset of $\mathcal{S}$ (or a set of sample points ζ_1, ζ_3, and ζ_5). Similarly, the event A_e, "an even number is thrown," is another subset of $\mathcal{S}$ (or a set of sample points ζ_2, ζ_4, and ζ_6):

$$A_o = (\zeta_1,\ \zeta_3,\ \zeta_5) \qquad\qquad A_e = (\zeta_2,\ \zeta_4,\ \zeta_6)$$

Let us denote the event "a number equal to or less than 4 is thrown" as B. Thus,

$$B = (\zeta_1,\ \zeta_2,\ \zeta_3,\ \zeta_4)$$

These events are clearly marked in Fig. 8.1. Note that an outcome can also be an event, because an outcome is a subset of $\mathcal{S}$ with only one element.

The **complement** of any event A, denoted by A^c, is the event containing all points not in A. Thus, for the event B in Fig. 8.1, $B^c = (\zeta_5,\ \zeta_6)$, $A_o^c = A_e$, and $A_e^c = A_o$. An event that has no sample points is a **null event**, which is denoted by $\emptyset$ and is equal to $\mathcal{S}^c$.

The **union** of events A and B, denoted by $A \cup B$, is the event that contains all points in A and B. This is the event stated as having "an outcome of either A or B." For the events in Fig. 8.1,

$$A_o \cup B = (\zeta_1,\ \zeta_3,\ \zeta_5,\ \zeta_2,\ \zeta_4)$$
$$A_e \cup B = (\zeta_2,\ \zeta_4,\ \zeta_6,\ \zeta_1,\ \zeta_3)$$

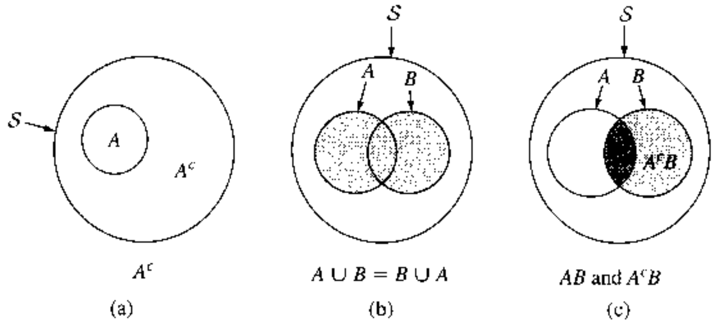


Figure 8.2 Representation of (a) complement, (b) union, and (c) intersection of events.

Observe that the union operation commutes:

$$A \cup B = B \cup A \tag{8.1}$$

The **intersection** of events A and B, denoted by $A \cap B$ or simply by AB, is the event that contains points common to A and B. This is the event that "outcome is both A and B," also known as the **joint event** $A \cap B$. Thus, the event A_eB, "a number that is even and equal to or less than 4 is thrown," is a set (ζ_2, ζ_4), and similarly for A_oB,

$$A_eB = (\zeta_2, \zeta_4) \qquad A_oB = (\zeta_1, \zeta_3)$$

Observe that the intersection also commutes

$$A \cap B = B \cap A \tag{8.2}$$

All these concepts can be demonstrated on a Venn diagram (Fig. 8.2). If the events A and B are such that

$$A \cap B = \emptyset \tag{8.3}$$

then A and B are said to be **disjoint**, or **mutually exclusive**, events. This means events A and B cannot occur simultaneously. In Fig. 8.1 events A_e and A_o are mutually exclusive, meaning that in any trial of the experiment if A_e occurs, A_o cannot occur at the same time, and vice versa.

Relative Frequency and Probability

Although the outcome of an experiment is unpredictable, there is a *statistical regularity* about the outcomes. For example, if a coin is tossed a large number of times, about half the times the outcome will be "heads," and the remaining half of the times it will be "tails." We may say that the relative frequency of the two outcomes "heads" or "tails" is one-half. This relative frequency represents the likelihood of a particular event.

Let A be one of the events of interest in an experiment. If we conduct a sequence of N independent trials* of this experiment, and if the event A occurs in $N(A)$ out of these N trials, then the fraction

$$f(A) = \lim_{N\to\infty} \frac{N(A)}{N} \tag{8.4}$$

* Trials conducted under similar discernible conditions.

is called the **relative frequency** of the event A. Observe that for small N, the fraction $N(A)/N$ may vary widely with N. As N increases, the fraction will approach a limit because of statistical regularity.

The probability of an event has the same connotations as the relative frequency of that event. Hence, we estimate the probability of each event, as the relative frequency of that event.* Therefore, to an event A, we assign the probability $P(A)$ as

$$P(A) = \lim_{N \to \infty} \frac{N(A)}{N} \tag{8.5}$$

From Eq. (8.5), it follows that

$$0 \leq P(A) \leq 1 \tag{8.6}$$

Example 8.1 Assign probabilities to each of the six outcomes in Fig. 8.1.

Because each of the six outcomes is equally likely in a large number of independent trials, each outcome will appear in one-sixth of the trials. Hence,

$$P(\zeta_i) = \frac{1}{6} \qquad i = 1, 2, 3, 4, 5, 6 \tag{8.7}$$

Consider now the two events A and B of an experiment. Suppose we conduct N independent trials of this experiment and events A and B occur in $N(A)$ and $N(B)$ trials, respectively. If A and B are mutually exclusive (or disjoint), then if A occurs, B cannot occur, and vice versa. Hence, the event $A \cup B$ occurs in $N(A) + N(B)$ trials and

$$\begin{aligned} P(A \cup B) &= \lim_{N \to \infty} \frac{N(A) + N(B)}{N} \\ &= P(A) + P(B) \qquad \text{if } A \cap B = \emptyset \end{aligned} \tag{8.8}$$

This result can be extended to more than two mutually exclusive events. In other words, if events $\{A_i\}$ are mutually exclusive such that

$$A_i \cap A_j = \emptyset \qquad i \neq j$$

then

$$P\left(\bigcup_i A_i\right) = \sum_i P(A_i)$$

* Observe that we are not defining the probability by the relative frequency. To a given event, a probability is closely estimated by the relative frequency of the event when this experiment is repeated many times. Modern theory of probability, being a branch of mathematics, starts with certain axioms about probability [Eqs. (8.6), (8.8), and (8.11)]. It assumes that somehow these probabilities are assigned by nature. We use relative frequency to estimate probability because it is reasonable in the sense that it closely approximates our experience and expectation of "probability."

Example 8.2 Assign probabilities to the events A_e, A_o, B, A_eB, and A_oB in Fig. 8.1.

Because $A_e = (\zeta_2 \cup \zeta_4 \cup \zeta_6)$ where ζ_2, ζ_4, and ζ_6 are mutually exclusive,

$$P(A_e) = P(\zeta_2) + P(\zeta_4) + P\zeta_6)$$

From Eq. (8.7) it follows that

$$P(A_e) = \frac{1}{2} \tag{8.9a}$$

Similarly,

$$P(A_o) = \frac{1}{2} \tag{8.9b}$$

$$P(B) = \frac{2}{3} \tag{8.9c}$$

From Fig. 8.1 we also observe that

$$A_eB = \zeta_2 \cup \zeta_4$$

and

$$P(A_eB) = P(\zeta_2) + P(\zeta_4) = \frac{1}{3} \tag{8.10a}$$

Similarly,

$$P(A_oB) = \frac{1}{3} \tag{8.10b}$$

We can also show that

$$P(\mathcal{S}) = 1 \tag{8.11}$$

This result can be proved by using the relative frequency. Let an experiment be repeated N times (N large). Because $\mathcal{S}$ is the union of all possible outcomes, $\mathcal{S}$ occurs in every trial. Hence, N out of N trials lead to event $\mathcal{S}$, and the result follows.

Example 8.3 Two dice are thrown. Determine the probability that the sum on the dice is seven.

For this experiment, the sample space contains 36 sample points because 36 possible outcomes exist. All the outcomes are equally likely. Hence, the probability of each outcome is 1/36.

A sum of seven can be obtained by the six combinations: (1, 6), (2, 5), (3, 4), (4, 3) (5, 2), and (6, 1). Hence, the event "a seven is thrown" is the union of six outcomes, each with probability 1/36. Therefore,

$$P(\text{"a seven is thrown"}) = \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} = \frac{1}{6}$$

Example 8.4 A coin is tossed four times in succession. Determine the probability of obtaining exactly two heads.

A total of $2^4 = 16$ distinct outcomes are possible, all of which are equally likely because of the symmetry of the situation. Hence, the sample space consists of 16 points, each with probability 1/16. The 16 outcomes are as follows:

HHHH	*TTTT*
HHHT	*TTTH*
HHTH	*TTHT*
⟶ *HHTT*	⟶ *TTHH*
HTHH	*THTT*
⟶ *HTHT*	⟶ *THTH*
⟶ *HTTH*	⟶ *THHT*
HTTT	*THHH*

Six out of these 16 outcomes lead to the event "obtaining exactly two heads" (arrows). Because all of the six outcomes are disjoint (mutually exclusive),

$$P(\text{obtaining exactly two heads}) = \frac{6}{16} = \frac{3}{8}$$

In Example 8.4, the method of listing all possible outcomes quickly becomes unwieldy as the number of tosses increases. For example, if a coin is tossed just 10 times, the total number of outcomes is 1024. A more convenient approach would be to apply the results of combinatorial analysis used in Bernoulli trials, to be discussed shortly.

Conditional Probability and Independent Events

Conditional Probability: It often happens that the probability of one event is influenced by the outcome of another event. As an example, consider drawing two cards in succession from a deck. Let A denote the event that the first card drawn is an ace. We do not replace the card drawn in the first trial. Let B denote the event that the second card drawn is an ace. It is evident that the probability of drawing an ace in the second trial will be influenced by the outcome of the first draw. If the first draw does not result in an ace, then the probability of obtaining an ace in the second trial is 4/51. The probability of event B thus depends on whether event A occurs. We now introduce the **conditional probability** $P(B|A)$ to denote the probability of event B when it is known that event A has occurred. $P(B|A)$ is read as "probability of B given A."

Let there be N trials of an experiment, in which the event A occurs n_1 times. Of *these* n_1 trials, event B occurs n_2 times. It is clear that n_2 is the number of times that the joint event $A \cap B$ (Fig. 8.2c) occurs. That is,

$$P(A \cap B) = \lim_{N \to \infty} \left(\frac{n_2}{N}\right) = \lim_{N \to \infty} \left(\frac{n_1}{N}\right)\left(\frac{n_2}{n_1}\right)$$

Note that $\lim_{N \to \infty}(n_1/N) = P(A)$. Also, $\lim_{N \to \infty}(n_2/n_1) = P(B|A)$,* because B occurs n_2 of the n_1 times that A occurred. This represents the conditional probability of B given A. Therefore,

$$P(A \cap B) = P(A)P(B|A) \tag{8.12}$$

and

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \qquad \text{provided } P(A) > 0 \tag{8.13a}$$

Using a similar argument, we obtain

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \qquad \text{provided } P(B) > 0 \tag{8.13b}$$

It follows from Eqs. (8.13) that

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)} \tag{8.14a}$$

$$P(B|A) = \frac{P(B)P(A|B)}{P(A)} \tag{8.14b}$$

Equations (8.14) are called **Bayes' rule**. In Bayes' rule, one conditional probability is expressed in terms of the reversed conditional probability.

Example 8.5 An experiment consists of drawing two cards from a deck in succession (without replacing the first card drawn). Assign a value to the probability of obtaining two red aces in two draws.

Let A and B be the events "red ace in the first draw" and "red ace in the second draw," respectively. We wish to determine $P(A \cap B)$,

$$P(A \cap B) = P(A)P(B|A)$$

and the relative frequency of A is 2/52 = 1/26. Hence,

$$P(A) = \frac{1}{26}$$

---

* Here we are implicitly using the fact that $n_1 \to \infty$ as $N \to \infty$. This is true provided the ratio $\lim_{N \to \infty}(n_1/N) \neq 0$, that is, if $P(A) \neq 0$.

Also, $P(B|A)$ is the probability of drawing a red ace in the second draw given that the first draw was a red ace. The relative frequency of this event is 1/51, so

$$P(B|A) = \frac{1}{51}$$

Hence,

$$P(A \cap B) = \left(\frac{1}{26}\right)\left(\frac{1}{51}\right) = \frac{1}{1326}$$

Independent Events: Under conditional probability, we presented an example where the occurrence of one event was influenced by the occurrence of another. There are, of course, many examples in which two or more events are entirely independent; that is, the occurrence of one event in no way influences the occurrence of the other event. As an example, we again consider the drawing of two cards in succession, but in this case we replace the card obtained in the first draw and shuffle the deck before the second draw. In this case, the outcome of the second draw is in no way influenced by the outcome of the first draw. Thus $P(B)$, the probability of drawing an ace in the second draw, is independent of whether the event A (drawing an ace in the first trial) occurs. Thus, the events A and B are independent. The conditional probability $P(B|A)$ is given by $P(B)$.

The event B is said to be **independent** of the event A if and only if

$$P(A \cap B) = P(A)P(B) \tag{8.15a}$$

Note that if the events A and B are independent, it follows from Eqs. (8.13a) and (8.15b) that

$$P(B|A) = P(B) \tag{8.15b}$$

This relationship states that if B is independent of A, then its probability is not affected by the event A. Naturally, if event B is independent of event A, then event A is also independent of B. It can been seen from Eqs. (8.14) that

$$P(A|B) = P(A) \tag{8.15c}$$

Note that there is a huge difference between **independent events** and **mutually exclusive events**. If A and B are mutually exclusive, then $A \cap B$ is empty and $P(A \cap B) = 0$. If A and B are mutually exclusive, then A and B cannot occur at the same time. This clearly means that they are NOT independent events.

Bernoulli Trials

In Bernoulli trials, if a certain event A occurs, we call it a "success." If $P(A) = p$, then the probability of success is p. If q is the probability of failure, then $q = 1 - p$. We shall find the probability of k successes in n (Bernoulli) trials. The outcome of each trial is independent of the outcomes of the other trials. It is clear that in n trials, if success occurs in k trials, failure occurs in $n - k$ trials. Since the outcomes of the trials are independent, the probability of this event is clearly $p^n(1-p)^{n-k}$, that is,

$$P(k \text{ successes in a specific order in } n \text{ trials}) = p^k(1-p)^{n-k}$$

But the event of "k successes in n trials" can occur in many different ways (different orders). It is well known from combinatorial analysis that there are

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \tag{8.16}$$

ways in which k positions can be taken from n positions (which is the same as the number of ways of achieving k successes in n trials).

This can be proved as follows. Consider an urn containing n distinguishable balls marked $1, 2, \ldots, n$. Suppose we draw k balls from this urn without replacing them. The first ball could be any one of the n balls, the second ball could be any one of the remaining $(n-1)$ balls, and so on. Hence, the total number of ways in which k balls can be drawn is

$$n(n-1)(n-2)\ldots(n-k+1) = \frac{n!}{(n-k)!}$$

Next, consider any one set of the k balls drawn. These balls can be ordered in different ways. We could label any one of the k balls as number 1, and any one of the remaining $(k-1)$ balls as number 2, and so on. This will give a total of $k(k-1)(k-2)\cdots 1 = k!$ distinguishable patterns formed from the k balls. The total number of ways in which k things can be taken from n things is $n!/(n-k)!$ But many of these ways will use the same k things, arranged in different order. The ways in which k things can be taken from n things without regard to order (unordered subset k taken from n things) is $n!/(n-k)!$ divided by $k!$ This is precisely defined by Eq. (8.16).

This means the probability of k successes in n trials is

$$\begin{aligned} P(k \text{ successes in } n \text{ trials}) &= \binom{n}{k} p^k (1-p)^{n-k} \\ &= \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \end{aligned} \tag{8.17}$$

Tossing a coin and observing the number of heads is a Bernoulli trial with $p = 0.5$. Hence, the probability of observing k heads in n tosses is

$$P(k \text{ heads in } n \text{ tosses}) = \binom{n}{k} (0.5)^k (0.5)^{n-k} = \frac{n!}{k!(n-k)!}(0.5)^n$$

Example 8.6 A binary symmetric channel (BSC) has an error probability P_e (i.e., the probability of receiving **0** when **1** is transmitted, or vice versa, is P_e). Note that the channel behavior is symmetrical with respect to **0** and **1**. Thus,

$$P(\mathbf{0}|\mathbf{1}) = P(\mathbf{1}|\mathbf{0}) = P_e$$

and

$$P(\mathbf{0}|\mathbf{0}) = P(\mathbf{1}|\mathbf{1}) = 1 - P_e$$

where $P(y|x)$ denotes the probability of receiving y when x is transmitted. A sequence of n binary digits is transmitted over this channel. Determine the probability of receiving exactly k digits in error.

The reception of each digit is independent of the other digits. This is an example of a Bernoulli trial with the probability of success $p = P_e$ ("success" here is receiving a digit in error). Clearly, the probability of k successes in n trials (k errors in n digits) is

$$P(\text{receiving } k \text{ out of } n \text{ digits in error}) = \binom{n}{k} P_e^k (1 - P_e)^{n-k}$$

For example, if $P_e = 10^{-5}$, the probability of receiving two digits wrong in a sequence of eight digits is

$$\binom{8}{2} (10^{-5})^2 (1 - 10^{-5})^6 \simeq \frac{8!}{2!\,6!} 10^{-10} = (2.8)10^{-9}$$

Example 8.7 PCM Repeater Error Probability

In pulse code modulation, regenerative repeaters are used to detect pulses (before they are lost in noise) and retransmit new, clean pulses. This combats the accumulation of noise and pulse distortion.

A certain PCM channel consists of n identical links in tandem (Fig. 8.3). The pulses are detected at the end of each link and clean new pulses are transmitted over the next link. If P_e is the probability of error in detecting a pulse over any one link, show that P_E, the probability of error in detecting a pulse over the entire channel (over the n links in tandem), is

$$P_E \simeq nP_e \qquad nP_e \ll 1$$

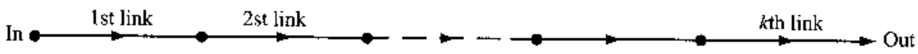


Figure 8.3 A PCM repeater.

The probabilities of detecting a pulse correctly over one link and over the entire channel (n links in tandem) are $1 - P_e$ and $1 - P_E$, respectively. A pulse can be detected correctly over the entire channel if either the pulse is detected correctly over every link or errors are made over an even number of links only.

$$\begin{aligned} 1 - P_E = {} & P(\text{correct detection over all links}) \\ & + P(\text{error over two links only}) \\ & + P(\text{error over four links only}) + \cdots \\ & + P\left(\text{error over } 2\left\lfloor \frac{n}{2} \right\rfloor \text{ links only}\right) \end{aligned}$$

where $\lfloor a \rfloor$ denotes the largest integer less than or equal to a.

Because pulse detection over each link is independent of the other links (see Example 8.6),

$$P(\text{correct detection over all } n \text{ links}) = (1 - P_e)^n$$

and

$$P(\text{error over } k \text{ links only}) = \frac{n!}{k!(n-k)!} P_e^k (1 - P_e)^{n-k}$$

Hence,

$$1 - P_E = (1 - P_e)^n + \sum_{k=2,4,6,\ldots}^{\alpha} \frac{n!}{k!(n-k)!} P_e^k (1 - P_e)^{n-k}$$

In practice, $P_e \ll 1$, so only the first two terms on the right-hand side of this equation are of significance. Also, $(1 - P_e)^{n-k} \simeq 1$, and

$$\begin{aligned} 1 - P_E &\simeq (1 - P_e)^n + \frac{n!}{2!(n-2)!} P_e^2 \\ &= (1 - P_e)^n + \frac{n(n-1)}{2} P_e^2 \end{aligned}$$

If $nP_e \ll 1$, then the second term can also be neglected, and

$$\begin{aligned} 1 - P_E &\simeq (1 - P_e)^n \\ &\simeq 1 - nP_e \qquad nP_e \ll 1 \end{aligned}$$

and

$$P_E \simeq nP_e$$

We can explain this result heuristically by considering the transmission of N $(N \to \infty)$ pulses. Each link makes NP_e errors, and the total number of errors is approximately nNP_e (approximately, because some of the erroneous pulses over a link will be erroneous over other links). Thus the overall error probability is nP_e.

Example 8.8

In binary communication, one of the techniques used to increase the reliability of a channel is to repeat a message several times. For example, we can send each message (**0** or **1**) three times. Hence, the transmitted digits are **000** (for message **0**) or **111** (for message **1**). Because of channel noise, we may receive any one of the eight possible combinations of three binary digits. The decision as to which message is transmitted is made by the majority rule; that is, if at least two of the three detected digits are **0**, the decision is **0**, and so on. This scheme permits correct reception of data even if one out of three digits is in error. Detection error occurs only if at least two out of three digits are received in error. If P_e is the error probability of one digit, and $P(\epsilon)$ is the probability of making a wrong decision in this scheme, then

$$\begin{aligned} P(\epsilon) &= \sum_{k=2}^{3} \binom{3}{k} P_e^k (1 - P_e)^{3-k} \\ &= 3P_e^2(1 - P_e) + P_e^3 \end{aligned}$$

In practice, $P_e \ll 1$, and

$$P(\epsilon) \simeq 3P_e^2$$

For instance, if $P_e = 10^{-4}$, $P(\epsilon) \simeq 3 \times 10^{-8}$. Thus, the error probability is reduced from 10^{-4} to 3×10^{-8}. We can use any odd number of repetitions for this scheme to function.

In this example, higher reliability is achieved at the cost of a reduction in the rate of information transmission by a factor of 3. We shall see in Chapter 14 that more efficient ways exist to effect a trade-off between reliability and the rate of transmission through the use of error correction codes.

Multiplication Rule for Conditional Probabilities

As shown in Eq. (8.12), we can write the joint event

$$P(A \cap B) = P(A)P(B/A)$$

This rule on joint events can be generalized for multiple events $A_1, A_2, \ldots, A_n$ via iterations. If $A_1A_2 \cdots A_n \neq \emptyset$, then we have

$$P(A_1A_2 \cdots A_n) = \frac{P(A_1A_2 \cdots A_n)}{P(A_1A_2 \cdots A_{n-1})} \cdot \frac{P(A_1A_2 \cdots A_{n-1})}{P(A_1A_2 \cdots A_{n-2})} \cdots \frac{P(A_1A_2)}{P(A_1)} \cdot P(A_1) \tag{8.18a}$$

$$= P(A_n|A_1A_2 \cdots A_{n-1}) \cdot P(A_{n-1}|A_1A_2 \cdots A_{n-2}) \cdots P(A_2|A_1) \cdot P(A_1) \tag{8.18b}$$

Note that since $A_1A_2 \cdots A_n \neq \emptyset$, every denominator in Eq. (8.18a) is positive and well defined.

Example 8.9 Suppose a box of diodes consist of N_g good diodes and N_b bad diodes. If five diodes are randomly selected, one at a time, without replacement, determine the probability of obtaining the sequence of diodes in the order of *good, bad, good, good, bad.*

We can denote G_i as the event that the ith draw is a good diode. We are interested in the event of $G_1G_2^cG_3G_4G_5^c$.

$$\begin{aligned} P(G_1G_2^cG_3G_4G_5^c) &= P(G_1)P(G_2^c|G_1)P(G_3|G_1G_2^c)P(G_4|G_1G_2^cG_3)P(G_5^c|G_1G_2^cG_3G_4) \\ &= \frac{N_g}{N_g+N_b} \cdot \frac{N_b}{N_g+N_b-1} \cdot \frac{N_g-1}{N_b+N_g-2} \cdot \frac{N_g-2}{N_g+N_b-3} \\ &\quad \cdot \frac{N_b-1}{N_g+N_b-4} \end{aligned}$$

To Divide and Conquer: The Total Probability Theorem

In analyzing a particular event of interest, sometimes a direct approach to evaluating its probability can be difficult because there can be so many different outcomes to enumerate. When dealing with such problems, it is often advantageous to adopt the *divide-and-conquer* approach by separating all the possible causes leading to the particular event of interest B. The total probability theorem provides a perfect tool for analyzing the probability of such problems.

We define $\mathcal{S}$ as the sample space of the experiment of interest. As shown in Fig. 8.4, the entire sample space can be partitioned into n disjoint events $A_1, \ldots, A_n$. We can now state the theorem:

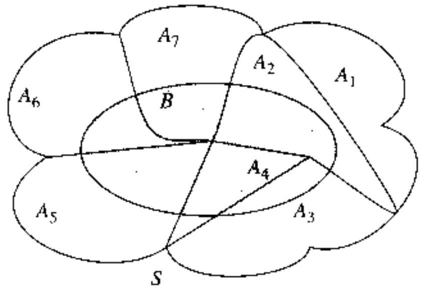


Figure 8.4 The event of interest B and the partition of S by $\{A_i\}$.

Total Probability Theorem: Let n disjoint events $A_1, \ldots, A_n$ form a partition of the sample space S such that

$$\bigcup_{i=1}^{n} A_i = S \quad \text{and} \quad A_i \cap A_j = \emptyset, \quad \text{if} \quad i \neq j$$

Then the probability of an event B can be written as

$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$$

Proof: The proof of this theorem is quite simple based on Fig. 8.4. Since $\{A_i\}$ form a partition of S, then

$$\begin{aligned} B &= B \cap S = B \cap (A_1 \cup A_2 \cup \cdots \cup A_n) \\ &= (A_1B) \cup (A_2B) \cup \cdots \cup (A_nB) \end{aligned}$$

Because $\{A_i\}$ are disjoint, so are $\{A_iB\}$. Thus,

$$P(B) = \sum_{i=1}^{n} P(A_iB) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$$

This theorem can simplify the analysis of the more complex event of interest B by identifying all different causes A_i for B. By quantifying the effect of A_i on B through $P(B|A_i)$, the theorem allows us to "divide-and-conquer" a complex problem (of event B).

Example 8.10 The decoding of a data packet may be in error because of N distinct error patterns $E_1, E_2, \ldots, E_N$ it encounters. These error patterns are mutually exclusive, each with probability $P(E_i) = p_i$. When the error pattern E_i occurs, the data packet would be incorrectly decoded with probability q_i. Find the probability that the data packet is incorrectly decoded.

We apply total probability theorem to tackle this problem. First, define B as the event that the data packet is incorrectly decoded. Based on the problem, we know that

$$P(B|E_i) = q_i \quad \text{and} \quad P(E_i) = p_i$$

Furthermore, the data packet has been incorrectly decoded. Therefore

$$\sum_{i=1}^{n} p_i = 1$$

Applying the total probability theorem, we find that

$$P(B) = \sum_{i=1}^{n} P(B|E_i)P(E_i) = \sum_{i=1}^{n} q_i p_i$$

Isolating a Particular Cause: Bayes' Theorem

The total probability theorem facilitates the probabilistic analysis of a complex event by using a *divide-and-conquer* approach. In practice, it may also be of interest to determine the likelihood of a particular cause of an event among many disjoint possible causes. Bayes' theorem provides the solution to this problem.

Bayes' Theorem: Let n disjoint events $A_1, \ldots, A_n$ form a partition of the sample space $\mathcal{S}$. Let B be an event with $P(B) > 0$. Then for $j = 1, \ldots, n$,

$$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{P(B)} = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)}$$

The proof is already given by the theorem itself.

Bayes' theorem provides a simple method for computing the conditional probability of A_j given that B has occurred. The probability $P(A_j|B)$ is often known as the *posterior probability* of event A_j. It describes, among n possible causes of B, the probability that B may be caused by A_j. In other words, Bayes' theorem isolates and finds the relative likelihood of each possible cause to an event of interest.

Example 8.11 A communication system always encounters one of three possible interference waveforms: F_1, F_2, or F_3. The probability of each interference is 0.8, 0.16, and 0.04, respectively. The communication system fails with probabilities 0.01, 0.1, and 0.4 when it encounters F_1, F_2, and F_3, respectively. Given that the system has failed, find the probability that the failure is a result of F_1, F_2, or F_3, respectively.

Denote B as the event of system failure. We know from the description that

$$P(F_1) = 0.8 \qquad P(F_2) = 0.16 \qquad P(F_3) = 0.04$$

Furthermore, the effect of each interference on the system is given by

$$P(B|F_1) = 0.01 \qquad P(B|F_2) = 0.1 \qquad P(B|F_3) = 0.4$$

Now following Bayes' theorem, we find that

$$P(F_1|B) = \frac{P(B|F_1)P(F_1)}{\sum_{i=1}^{3} P(B|F_i)P(F_i)} = \frac{(0.01)(0.8)}{(0.01)(0.8) + (0.1)(0.16) + (0.4)(0.04)} = 0.2$$

$$P(F_2|B) = \frac{P(B|F_2)P(F_2)}{\sum_{i=1}^{3} P(B|F_i)P(F_i)} = 0.4$$

$$P(F_3|B) = \frac{P(B|F_3)P(F_3)}{\sum_{i=1}^{3} P(B|F_i)P(F_i)} = 0.4$$

Example 8.11 illustrates the major difference between the *posterior probability* $P(F_i|B)$ and the *prior probability* $P(F_i)$. Although the prior probability $P(F_3) = 0.04$ is the lowest among the three possible interferences, once the failure event B has occurred, $P(F_3|B) = 0.4$ is actually one of the most likely events. Bayes' theorem is an important tool in communications for determining the relative likelihood of a particular cause to an event.

Axiomatic Theory of Probability

The relative frequency definition of probability is intuitively appealing. Unfortunately, it has some serious mathematical objections. Logically there is no reason why we should get the same estimate of the relative frequency whether we base it on 10,000 trials or on 20. Moreover, in the relative frequency definition, it is not clear when and in what mathematical sense the limit in Eq. (8.5) exists. If we consider a set of an infinite number of trials, we can partition such a set into several subsets, such as odd and even numbered trials. Each of these subsets (of infinite trials each) would have its own relative frequency. So far, all the attempts to prove that the relative frequencies of all the subsets are equal have been futile.[1] There are some other difficulties also. For instance, in some cases, such as Julius Caesar having visited Great Britain, it is an experiment for which we cannot repeat the event an infinite number of trials. Thus, we can never know the probability of such an event. We, therefore, need to develop a theory of probability that is not tied down to any particular definition of probability. In other words, we must separate the empirical and the formal problems of probability. Assigning probabilities to events is an empirical aspect, and setting up purely formal calculus to deal with probabilities (assigned by whatever empirical method) is the formal aspect.

It is instructive to consider here the basic difference between physical sciences and mathematics. Physical sciences are based on **inductive logic**, while mathematics is strictly a **deductive logic**. Inductive logic consists of making a large number of observations and then generalizing, from these observations, laws that will explain these observations. For instance, history and experience tell us that every human being must die someday. This leads to a law that *humans are mortals*. This is inductive logic. Based on a law (or laws) obtained by inductive logic, we can make further deductions. The statement "John is a human being, so he must die some day" is an example of deductive logic. Deriving the laws of the physical sciences is basically an exercise in inductive logic, whereas mathematics is pure deductive logic. In a physical science we make observations in a certain field and generalize these observations into laws such as Ohm's law, Maxwell's equations, and quantum mechanics. There are no other proofs for these inductively obtained laws; they are found to be true by observation. But once we have such inductively formulated laws (axioms or hypotheses), by using thought process, we can deduce additional results based on these *basic laws or axioms* alone. This is the proper domain of mathematics. All these deduced results have to be proved rigorously based on a set of axioms. Thus, based on Maxwell's equations alone, we can derive the laws of the propagation of electromagnetic waves.

This discussion shows that the discipline of mathematics can be summed up in one aphorism: "This implies that." In other words, if we are given a certain set of axioms (hypotheses), then, based upon these axioms alone, what else is true? As Bertrand Russell puts it: "Pure mathematics consists entirely of such asseverations as that, if such and such proposition is true of anything, then such and such another proposition is true of that thing." Seen in this light, it may appear that assigning probability to an event may not necessarily be the responsibility of the mathematical discipline of probability. Under mathematical discipline, we need to start with a set of axioms about probability and then investigate what else can be said about probability based on this set of axioms alone. We start with a concept (as yet undefined) of probability and postulate axioms. The axioms must be internally consistent and should conform to the observed relationships and behavior of probability in the practical and the intuitive sense. It is beyond the scope of this book to discuss how these axioms are formulated. The modern theory of probability starts with Eqs. (8.6), (8.8), and (8.11) as its axioms. Based on these three axioms alone, what else is true is the essence of modern theory of probability. The relative frequency approach uses Eq. (8.5) to define probability, and Eqs. (8.5), (8.8), and (8.11) follow as a consequence of this definition. In the axiomatic approach, on the other hand, we do not say anything about how we assign probability $P(A)$ to an event A; rather, we postulate that the probability function must obey the three postulates or axioms in Eqs. (8.6), (8.8), and (8.11). The modern theory of probability does not concern itself with the problem of assigning probabilities to events. It assumes that somehow the probabilities were assigned to these events a priori.

If a mathematical model is to conform to the real phenomenon, we must assign these probabilities in away that is consistent with an empirical and an intuitive understanding of probability. The concept of relative frequency is admirably suited for this. Thus, although we use relative frequency to assign (not define) probabilities, it is all under the table, not a part of the mathematical discipline of probability.

8.2 RANDOM VARIABLES

The outcome of an experiment may be a real number (as in the case of rolling a die), or it may be nonnumerical and describable by a phrase (such as "heads" or "tail" in tossing a coin). From a mathematical point of view, it is simpler to have numerical values for all outcomes. For this reason, we assign a real number to each sample point according to some rule. If there are m sample points $\zeta_1, \zeta_2, \ldots, \zeta_m$, then using some convenient rule, we assign a real number $\mathrm{x}(\zeta_i)$ to sample point ζ_i ($i = 1, 2, \ldots, m$). In the case of tossing a coin, for example, we may assign the number 1 for the outcome heads and the number -1 for the outcome tails (Fig. 8.5).

Thus, x(.) is a function that maps sample points $\zeta_1, \zeta_2, \ldots, \zeta_m$ into real numbers $x_1, x_2, \ldots, x_n$.* We now have a **random variable** x that takes on values $x_1, x_2, \ldots, x_n$. We shall use roman type (x) to denote a random variable (RV) and italic type (e.g., $x_1, x_2, \ldots, x_n$) to denote the value it takes. The probability of an RV x taking a value x_i is $P_{\mathrm{x}}(x_i)$ = Probability of "$\mathrm{x} = x_i$."

Discrete Random Variables

A random variable is discrete if there exists a denumerable sequence of distinct numbers x_i such that

$$\sum_i P_{\mathrm{x}}(x_i) = 1 \tag{8.19}$$

* The number m is not necessarily equal to n. More than one sample point can map into one value of x.

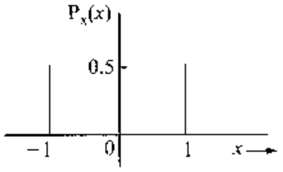


Figure 8.5 Probabilities in a coin-tossing experiment.

Thus, a discrete RV can assume only certain discrete values. An RV that can assume any value over a continuous set is called a **continuous** random variable.

Example 8.12 Two dice are thrown. The sum of the points appearing on the two dice is an RV x. Find the values taken by x, and the corresponding probabilities.

We see that x can take on all integral values from 2 through 12. Various probabilities can be determined by the method outlined in Example 8.3.

There are 36 sample points in all, each with probability 1/36. Dice outcomes for various values of x are shown in Table 8.1. Note that although there are 36 sample points, they all map into 11 values of x. This is because more than one sample point maps into the same value of x. For example, six sample points map into x = 7.

The reader can verify that $\sum_{i=2}^{12} P_{\mathrm{x}}(x_i) = 1$.

TABLE 8.1

Value of x_i	Dice Outcomes	$P_{\mathrm{x}}(x_i)$
2	(1, 1)	1/36
3	(1, 2), (2, 1)	2/36 = 1/18
4	(1, 3), (2, 2), (3, 1)	3/36 = 1/12
5	(1, 4), (2, 3), (3, 2), (4, 1)	4/36 = 1/9
6	(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)	5/36
7	(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)	6/36 =1/6
8	(2, 6), (3, 5), (4, 4), (5, 3), (6, 2)	5/36
9	(3, 6), (4, 5), (5, 4), (6, 3)	4/36 = 1/9
10	(4, 6), (5, 5), (6, 4)	3/36 = 1/12
11	(5, 6), (6, 5)	2/36 = 1/18
12	(6, 6)	1/36

The preceding discussion can be extended to two RVs, x and y. The joint probability $P_{\mathrm{xy}}(x_i, y_j)$ is the probability that "x $= x_i$ and y $= y_j$." Consider, for example, the case of a coin tossed twice in succession. If the outcomes of the first and second tosses are mapped into RVs x and y, then x and y each takes values 1 and −1. Because the outcomes of the two tosses are independent, x and y are independent, and

$$P_{\mathrm{xy}}(x_i, y_j) = P_{\mathrm{x}}(x_i)\, P_{\mathrm{y}}(y_j)$$

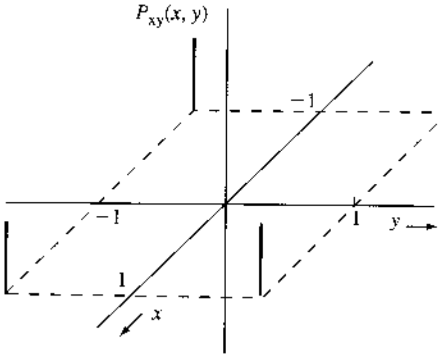


Figure 8.6 Representation of joint probabilities of two random variables.

and

$$P_{xy}(1,\ 1) = P_{xy}(1,\ -1) = P_{xy}(-1,\ 1) = P_{xy}(-1,\ -1) = \tfrac{1}{4}$$

These probabilities are plotted in Fig. 8.6.

For a general case where the variable x can take values $x_1,\ x_2,\ \ldots,\ x_n$ and the variable y can take values $y_1,\ y_2,\ \ldots,\ y_m$, we have

$$\sum_i \sum_j P_{xy}(x_i,\ y_j) = 1 \tag{8.20}$$

This follows from the fact that the summation on the left is the probability of the union of all possible outcomes and must be unity (a certain event).

Conditional Probabilities

If x and y are two RVs, then the conditional probability of $x = x_i$ given $y = y_j$ is denoted by $P_{x|y}(x_i|y_j)$. Moreover,

$$\sum_i P_{x|y}(x_i|y_j) = \sum_j P_{y|x}(y_j|x_i) = 1 \tag{8.21}$$

This can be proved by observing that probabilities $P_{x|y}(x_i|y_j)$ are specified over the sample space corresponding to the condition $y = y_j$. Hence, $\sum_i P_{x|y}(x_i|y_j)$ is the probability of the union of all possible outcomes of x (under the condition $y = y_j$) and must be unity (a certain event). A similar argument applies to $\sum_j P_{y|x}(y_j|x_i)$. Also from Eq. (8.12), we have

$$P_{xy}(x_i,\ y_j) = P_{x|y}(x_i|y_j)P_y(y_j) = P_{y|x}(y_j|x_i)P_x(x_i) \tag{8.22}$$

Bayes' rule follows from Eq. (8.22). Also from Eq. (8.22), we have

$$\begin{aligned}\sum_i P_{xy}(x_i,\ y_j) &= \sum_i P_{x|y}(x_1|y_j)P_y(y_j)\\ &= P_y(y_j)\sum_i P_{x|y}(x_i|y_j)\\ &= P_y(y_j)\end{aligned} \tag{8.23a}$$

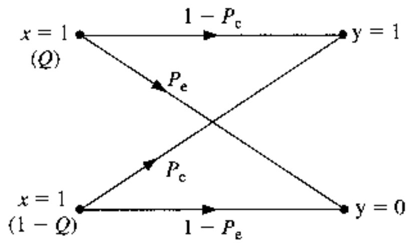
Similarly,

$$P_x(x_i) = \sum_j P_{xy}(x_i, y_j) \tag{8.23b}$$

The probabilities $P_x(x_i)$ and $P_y(y_j)$ are called **marginal probabilities**. Equations (8.23) show how to determine marginal probabilities from joint probabilities. Results of Eqs. (8.20) through (8.23) can be extended to more than two RVs.

Example 8.13 A binary symmetric channel (BSC) error probability is P_e. The probability of transmitting **1** is Q, and that of transmitting **0** is $1 - Q$ (Fig. 8.7). Determine the probabilities of receiving **1** and **0** at the receiver.

Figure 8.7 Binary symmetric channel (BSC).



If x and y are the transmitted digit and the received digit, respectively, then for a BSC,

$$P_{y|x}(\mathbf{0}|\mathbf{1}) = P_{y|x}(\mathbf{1}|\mathbf{0}) = P_e$$
$$P_{y|x}(\mathbf{0}|\mathbf{0}) = P_{y|x}(\mathbf{1}|\mathbf{1}) = 1 - P_e$$

Also,

$$P_x(\mathbf{1}) = Q \qquad \text{and} \qquad P_x(\mathbf{0}) = 1 - Q$$

We need to find $P_y(\mathbf{1})$ and $P_y(\mathbf{0})$. From the total probability theorem,

$$P_y(y_j) = \sum_i P_x(x_i) P_{y|x}(y_j|x_i)$$

we find

$$P_y(\mathbf{1}) = P_x(\mathbf{0}) P_{y|x}(\mathbf{1}|\mathbf{0}) + P_x(\mathbf{1}) P_{y|x}(\mathbf{1}|\mathbf{1})$$
$$= (1 - Q)P_e + Q(1 - P_e)$$

Similarly,

$$P_y(\mathbf{0}) = (1 - Q)(1 - P_e) + QP_e$$

These answers seem almost obvious from Fig. 8.7.

Note that because of channel errors, the probability of receiving a digit **1** is not the same as that of transmitting **1**. The same is true of **0**.

Example 8.14 Over a certain binary communication channel, the symbol **0** is transmitted with probability 0.4 and **1** is transmitted with probability 0.6. It is given that $P(\epsilon|\mathbf{0}) = 10^{-6}$ and $P(\epsilon|\mathbf{1}) = 10^{-4}$, where $P(\epsilon|x_i)$ is the probability of detecting the error given that x_i is transmitted. Determine $P(\epsilon)$, the error probability of the channel.

If $P(\epsilon, x_i)$ is the joint probability that x_i is transmitted and it is detected wrongly, then the total probability theorem yields

$$\begin{aligned} P(\epsilon) &= \sum_i P(\epsilon|x_i)P(x_i) \\ &= P_{\mathrm{x}}(\mathbf{0})P(\epsilon|\mathbf{0}) + P_{\mathrm{x}}(\mathbf{1})P(\epsilon|\mathbf{1}) \\ &= 0.4(10^{-6}) + 0.6(10^{-4}) \\ &= 0.604(10^{-4}) \end{aligned}$$

Note that $P(\epsilon|\mathbf{0}) = 10^{-6}$ means that on the average, one out of 1 million received **0**s will be detected erroneously. Similarly, $P(\epsilon|\mathbf{1}) = 10^{-4}$ means that on the average, one out of 10,000 received **1**s will be in error. But $P(\epsilon) = 0.604(10^{-4})$ indicates that on the average, one out of $1/0.604(10^{-4}) \simeq 16{,}556$ digits (regardless of whether they are **1**s or **0**s) will be received in error.

Cumulative Distribution Function

The **cumulative distribution function (CDF)** $F_{\mathrm{x}}(x)$ of an RV x is the probability that x takes a value less than or equal to x; that is,

$$F_{\mathrm{x}}(x) = P(\mathrm{x} \leq x) \tag{8.24}$$

We can show that a CDF $F_{\mathrm{x}}(x)$ has the following four properties:

1. $F_{\mathrm{x}}(x) \geq 0$ (8.25a)
2. $F_{\mathrm{x}}(\infty) = 1$ (8.25b)
3. $F_{\mathrm{x}}(-\infty) = 0$ (8.25c)
4. $F_{\mathrm{x}}(x)$ is a nondecreasing function, that is, (8.25d)

   $F_{\mathrm{x}}(x_1) \leq F_{\mathrm{x}}(x_2)$ for $x_1 \leq x_2$ (8.25e)

The first property is obvious. The second and third properties are proved by observing that $F_{\mathrm{x}}(\infty) = P(\mathrm{x} \leq \infty)$ and $F_{\mathrm{x}}(-\infty) = P(\mathrm{x} \leq -\infty)$. To prove the fourth property, we have, from Eq. (8.24),

$$\begin{aligned} F_{\mathrm{x}}(x_2) &= P(\mathrm{x} \leq x_2) \\ &= P[(\mathrm{x} \leq x_1) \cup (x_1 < \mathrm{x} \leq x_2)] \end{aligned}$$

Because $x \le x_1$ and $x_1 < x \le x_2$ are disjoint, we have

$$\begin{aligned} F_x(x_2) &= P(x \le x_1) + P(x_1 < x \le x_2) \\ &= F_x(x_1) + P(x_1 < x \le x_2) \end{aligned} \tag{8.26}$$

Because $P(x_1 < x \le x_2)$ is nonnegative, the result follows.

Example 8.15 In an experiment, a trial consists of four successive tosses of a coin. If we define an RV x as the number of heads appearing in a trial, determine $P_x(x)$ and $F_x(x)$.

A total of 16 distinct equiprobable outcomes are listed in Example 8.4. Various probabilities can be readily determined by counting the outcomes pertaining to a given value of x. For example, only one outcome maps into x=0, whereas six outcomes map into x=2. Hence, $P_x(0) = 1/16$ and $P_x(2) = 6/16$. In the same way, we find

$$\begin{aligned} P_x(0) &= P_x(4) = 1/16 \\ P_x(1) &= P_x(3) = 4/16 = 1/4 \\ P_x(2) &= 6/16 = 3/8 \end{aligned}$$

The probabilities $P_x(x_i)$ and the corresponding CDF $F_x(x_i)$ are shown in Fig. 8.8.

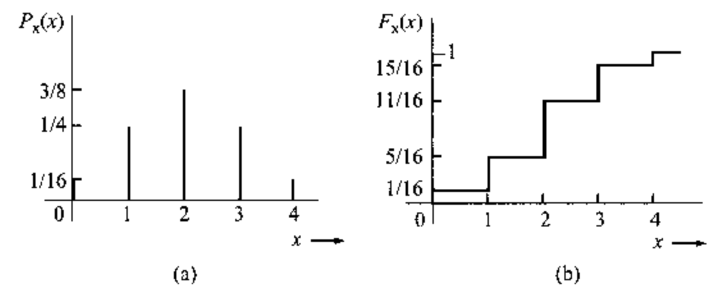


Figure 8.8 (a) Probabilities $P_x(x_i)$ and (b) the cumulative distribution function (CDF).

Continuous Random Variables

A continuous RV x can assume any value in a certain interval. In a continuum of any range, an uncountably infinite number of possible values exist, and $P_x(x_i)$, the probability that $x = x_i$, as one of the uncountably infinite values, is generally zero. Consider the case of a temperature **T** at a certain location. We may suppose that this temperature can assume any of a range of values. Thus, an infinite number of possible temperature values may prevail, and the probability that the random variable **T** will assume a certain value T_i is zero. The situation is somewhat similar to that described in Sec. 3.1 in connection with a continuously loaded beam (Fig. 3.5b). There is a loading along the beam at every point, but at any one point the load is zero. The meaningful measure in that case was the loading (or weight) not at a point, but over a finite interval. Similarly, for a continuous RV, the meaningful quantity is not the probability that $x = x_i$ but the probability that $x < x \le x + \Delta x$. For such a measure, the CDF is eminently suited because the latter probability is simply $F_x(x + \Delta x) - F_x(x)$ [see Eq. (8.26)]. Hence, we begin our study of continuous RVs with the CDF.

Properties of the CDF [Eqs. (8.25) and (8.26)] derived earlier are general and are valid for continuous as well as discrete RVs.

Probability Density Function: From Eq. (8.26), we have

$$F_x(x + \Delta x) = F_x(x) + P(x < \mathrm{x} \leq x + \Delta x) \tag{8.27a}$$

If $\Delta x \to 0$, then we can also express $F_x(x + \Delta x)$ via Taylor expansion as

$$F_x(x + \Delta x) \simeq F_x(x) + \frac{dF_x(x)}{dx}\Delta x \tag{8.27b}$$

From Eqs. (8.27), it follows that as $\Delta x \to 0$,

$$\frac{dF_x(x)}{dx}\Delta x = P(x < \mathrm{x} \leq x + \Delta x) \tag{8.28}$$

We designated the derivative of $F_x(x)$ with respect to x by $p_x(x)$ (Fig. 8.9),

$$\frac{dF_x(x)}{dx} = p_x(x) \tag{8.29}$$

The function $p_x(x)$ is called the **probability density function (PDF)** of the RV x. It follows from Eq. (8.28) that the probability of observing the RV x in the interval $(x,\ x + \Delta x)$ is $p_x(x)\Delta x$ $(\Delta x \to 0)$. This is the area under the PDF $p_x(x)$ over the interval Δx, as shown in Fig. 8.9b.

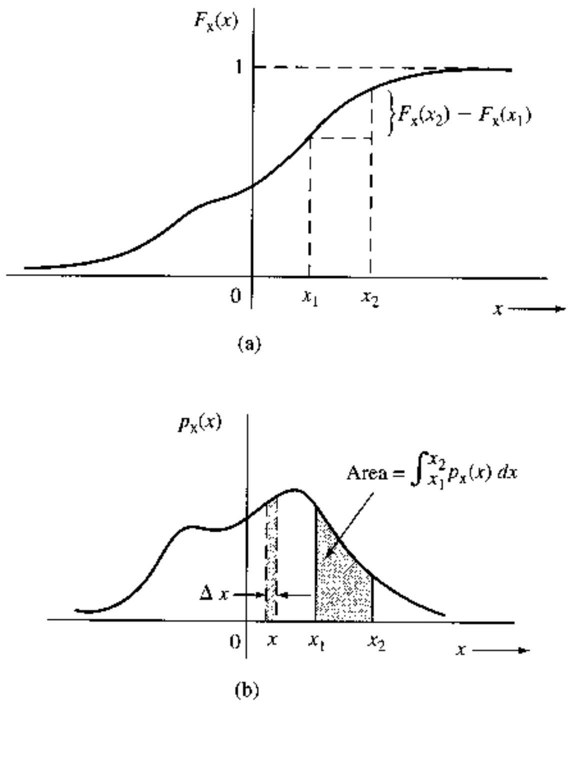


Figure 8.9 (a) Cumulative distribution function (CDF). (b) Probability density function (PDF).

From Eq. (8.29), we can see that

$$F_{\mathrm{x}}(x) = \int_{-\infty}^{x} p_{\mathrm{x}}(u)\, du \tag{8.30}$$

Here we use the fact that $F_{\mathrm{x}}(-\infty) = 0$. We also have from Eq. (8.26)

$$\begin{aligned} P(x_1 < \mathrm{x} \leq x_2) &= F_{\mathrm{x}}(x_2) - F_{\mathrm{x}}(x_1) \\ &= \int_{-\infty}^{x_2} p_{\mathrm{x}}(x)\, dx - \int_{-\infty}^{x_1} p_{\mathrm{x}}(x)\, dx \\ &= \int_{x_1}^{x_2} p_{\mathrm{x}}(x)\, dx \end{aligned} \tag{8.31}$$

Thus, the probability of observing x in any interval (x_1, x_2) is given by the area under the PDF $p_{\mathrm{x}}(x)$ over the interval (x_1, x_2), as shown in Fig. 8.9b. Compare this with a continuously loaded beam (Fig. 3.5b), where the weight over any interval was given by an integral of the loading density over the interval.

Because $F_{\mathrm{x}}(\infty) = 1$, we have

$$\int_{-\infty}^{\infty} p_{\mathrm{x}}(x)\, dx = 1 \tag{8.32}$$

This also follows from the fact that the integral in Eq. (8.32) represents the probability of observing x in the interval $(-\infty, \infty)$. Every PDF must satisfy the condition in Eq. (8.32). It is also evident that the PDF must not be negative, that is,

$$p_{\mathrm{x}}(x) \geq 0$$

Although it is true that the probability of an impossible event is **0** and that of a certain event is **1**, the converse is not true. An event whose probability is **0** is not necessarily an impossible event, and an event with a probability of **1** is not necessarily a certain event. This may be illustrated by the following example. The temperature T of a certain city on a summer day is an RV taking on any value in the range of 5 to 50°C. Because the PDF $p_{\mathrm{T}}(T)$ is continuous, the probability that T = 34.56, for example, is zero. But this is not an impossible event. Similarly, the probability that T takes on any value but 34.56 is **1**, although this is not a certain event. In fact, a continuous RV x takes every value in a certain range. Yet $p_{\mathrm{x}}(x)$, the probability that $\mathrm{x} = x$, is zero for every x in that range.

We can also determine the PDF $p_{\mathrm{x}}(x)$ for a discrete random variable. Because the CDF $F_{\mathrm{x}}(x)$ for the discrete case is always a sequence of step functions (Fig. 8.8), the PDF (the derivative of the CDF) will consist of a train of positive impulses. If an RV x takes values $x_1, x_2, \ldots, x_n$ with probabilities $a_1, a_2, \ldots, a_n$, respectively, then

$$F_{\mathrm{x}}(x) = a_1 u(x - x_1) + a_2 u(x - x_2) + \cdots + a_n u(x - x_n) \tag{8.33a}$$

This can be easily verified from Example 8.15 (Fig. 8.8). Hence,

$$\begin{aligned} p_{\mathrm{x}}(x) &= a_1\, \delta(x - x_1) + a_2\, \delta(x - x_2) + \cdots + a_n\, \delta(x - x_n) \\ &= \sum_{r=1}^{n} a_r\, \delta(x - x_r) \end{aligned} \tag{8.33b}$$

It is, of course, possible to have a mixed case, where a PDF may have a continuous part and an impulsive part (see Prob. 8.2-4).

The Gaussian Random Variable

Consider a PDF (Fig. 8.10)

$$p_x(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \tag{8.34}$$

This is a case of the well-known standard **Gaussian**, or **normal**, probability density. It has zero mean and unit variance. This function was named after the famous mathematician Carl Friedrich Gauss.

The CDF $F_x(x)$ in this case is

$$F_x(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-x^2/2}\, dx$$

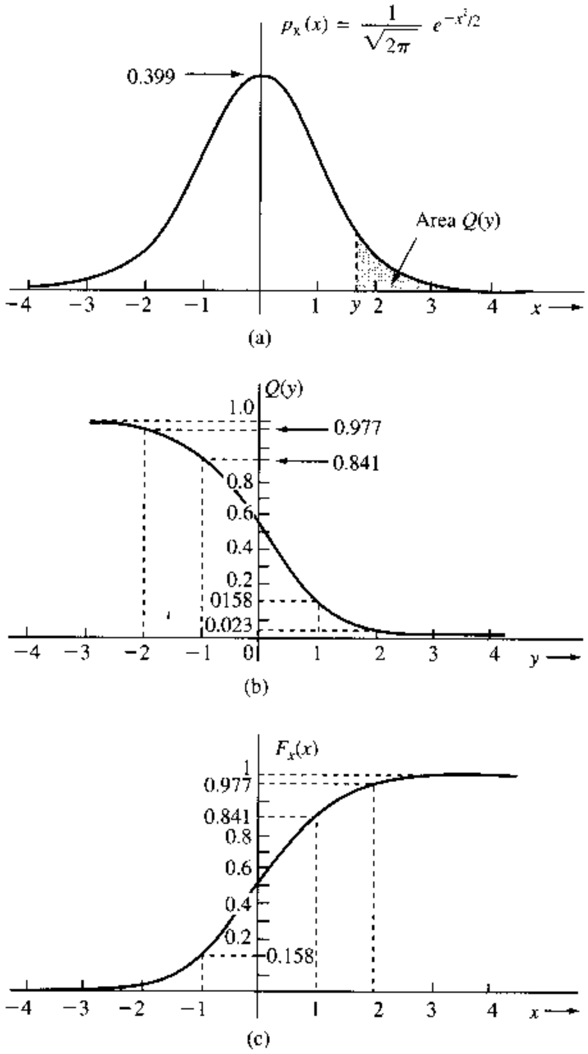


Figure 8.10 (a) Gaussian PDF. (b) Function $Q(y)$. (c) CDF of the Gaussian PDF.

This integral cannot be evaluated in a closed form and must be computed numerically. It is convenient to use the function $Q(.)$, defined as[2]

$$Q(y) \triangleq \frac{1}{\sqrt{2\pi}} \int_y^{\infty} e^{-x^2/2}\, dx \tag{8.35}$$

The area under $p_x(x)$ from y to ∞ (shaded in Fig. 8.10a) is* $Q(y)$. From the symmetry of $p_x(x)$ about the origin, and the fact that the total area under $p_x(x) = 1$, it follows that

$$Q(-y) = 1 - Q(y) \tag{8.36}$$

Observe that for the PDF in Fig. 8.10a, the CDF is given by (Fig. 8.10c)

$$F_x(x) = 1 - Q(x) \tag{8.37}$$

The function $Q(x)$ is tabulated in Table 8.2 (see also later: Fig. 8.12d). This function is widely tabulated and can be found in most of the standard mathematical tables.[2, 3] It can be shown that,[4]

$$Q(x) \simeq \frac{1}{x\sqrt{2\pi}} e^{-x^2/2} \qquad \text{for } x \gg 1 \tag{8.38a}$$

For example, when $x = 2$, the error in this approximation is 18.7%. But for $x = 4$ it is 10.4% and for $x = 6$ it is 2.3%.

A much better approximation to $Q(x)$ is

$$Q(x) \simeq \frac{1}{x\sqrt{2\pi}} \left(1 - \frac{0.7}{x^2}\right) e^{-x^2/2} \qquad x > 2 \tag{8.38b}$$

The error in this approximation is just within 1% for $x > 2.15$. For larger values of x the error approaches 0.

A more general Gaussian density function has two parameters (m, σ) and is (Fig. 8.11)

$$p_x(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-m)^2/2\sigma^2} \tag{8.39}$$

For this case,

$$F_x(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-(x-m)^2/2\sigma^2}\, dx$$

---

* The function $Q(x)$ is closely related to functions erf (x) and erfc (x),

$$\text{erfc}\,(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-y^2}\, dy = 2Q(x\sqrt{2})$$

Therefore,

$$Q(x) = \frac{1}{2}\,\text{erfc}\left(\frac{x}{\sqrt{2}}\right) = \frac{1}{2}\left[1 - \text{erf}\left(\frac{x}{\sqrt{2}}\right)\right]$$

TABLE 8.2[3]
$Q(x)$

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0000	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
.1000	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
.2000	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
.3000	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
.4000	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
.5000	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
.6000	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
.7000	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
.8000	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
.9000	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.000	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.100	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.200	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.9853E-01
1.300	.9680E-01	.9510E-01	.9342E-01	.9176E-01	.9012E-01	.8851E-01	.8691E-01	.8534E-01	.8379E-01	.8226E-01
1.400	.8076E-01	.7927E-01	.7780E-01	.7636E-01	.7493E-01	.7353E-01	.7215E-01	.7078E-01	.6944E-01	.6811E-01
1.500	.6681E-01	.6552E-01	.6426E-01	.6301E-01	.6178E-01	.6057E-01	.5938E-01	.5821E-01	.5705E-01	.5592E-01
1.600	.5480E-01	.5370E-01	.5262E-01	.5155E-01	.5050E-01	.4947E-01	.4846E-01	.4746E-01	.4648E-01	.4551E-01
1.700	.4457E-01	.4363E-01	.4272E-01	.4182E-01	.4093E-01	.4006E-01	.3920E-01	.3836E-01	.3754E-01	.3673E-01
1.800	.3593E-01	.3515E-01	.3438E-01	.3362E-01	.3288E-01	.3216E-01	.3144E-01	.3074E-01	.3005E-01	.2938E-01
1.900	.2872E-01	.2807E-01	.2743E-01	.2680E-01	.2619E-01	.2559E-01	.2500E-01	.2442E-01	.2385E-01	.2330E-01
2.000	.2275E-01	.2222E-01	.2169E-01	.2118E-01	.2068E-01	.2018E-01	.1970E-01	.1923E-01	.1876E-01	.1831E-01
2.100	.1786E-01	.1743E-01	.1700E-01	.1659E-01	.1618E-01	.1578E-01	.1539E-01	.1500E-01	.1463E-01	.1426E-01
2.200	.1390E-01	.1355E-01	.1321E-01	.1287E-01	.1255E-01	.1222E-01	.1191E-01	.1160E-01	.1130E-01	.1101E-01
2.300	.1072E-01	.1044E-01	.1017E-01	.9903E-02	.9642E-02	.9387E-02	.9137E-02	.8894E-02	.8656E-02	.8424E-02
2.400	.8198E-02	.7976E-02	.7760E-02	.7549E-02	.7344E-02	.7143E-02	.6947E-02	.6756E-02	.6569E-02	.6387E-02
2.500	.6210E-02	.6037E-02	.5868E-02	.5703E-02	.5543E-02	.5386E-02	.5234E-02	.5085E-02	.4940E-02	.4799E-02
2.600	.4661E-02	.4527E-02	.4396E-02	.4269E-02	.4145E-02	.4025E-02	.3907E-02	.3793E-02	.3681E-02	.3573E-02
2.700	.3467E-02	.3364E-02	.3264E-02	.3167E-02	.3072E-02	.2980E-02	.2890E-02	.2803E-02	.2718E-02	.2635E-02
2.800	.2555E-02	.2477E-02	.2401E-02	.2327E-02	.2256E-02	.2186E-02	.2118E-02	.2052E-02	.1988E-02	.1926E-02
2.900	.1866E-02	.1807E-02	.1750E-02	.1695E-02	.1641E-02	.1589E-02	.1538E-02	.1489E-02	.1441E-02	.1395E-02
3.000	.1350E-02	.1306E-02	.1264E-02	.1223E-02	.1183E-02	.1144E-02	.1107E-02	.1070E-02	.1035E-02	.1001E-02
3.100	.9676E-03	.9354E-03	.9043E-03	.8740E-03	.8447E-03	.8164E-03	.7888E-03	.7622E-03	.7364E-03	.7114E-03
3.200	.6871E-03	.6637E-03	.6410E-03	.6190E-03	.5976E-03	.5770E-03	.5571E-03	.5377E-03	.5190E-03	.5009E-03
3.300	.4834E-03	.4665E-03	.4501E-03	.4342E-03	.4189E-03	.4041E-03	.3897E-03	.3758E-03	.3624E-03	.3495E-03
3.400	.3369E-03	.3248E-03	.3131E-03	.3018E-03	.2909E-03	.2802E-03	.2701E-03	.2602E-03	.2507E-03	.2415E-03
3.500	.2326E-03	.2241E-03	.2158E-03	.2078E-03	.2001E-03	.1926E-03	.1854E-03	.1785E-03	.1718E-03	.1653E-03
3.600	.1591E-03	.1531E-03	.1473E-03	.1417E-03	.1363E-03	.1311E-03	.1261E-03	.1213E-03	.1166E-03	.1121E-03
3.700	.1078E-03	.1036E-03	.9961E-04	.9574E-04	.9201E-04	.8842E-04	.8496E-04	.8162E-04	.7841E-04	.7532E-04
3.800	.7235E-04	.6948E-04	.6673E-04	.6407E-04	.6152E-04	.5906E-04	.5669E-04	.5442E-04	.5223E-04	.5012E-04
3.900	.4810E-04	.4615E-04	.4427E-04	.4247E-04	.4074E-04	.3908E-04	.3747E-04	.3594E-04	.3446E-04	.3304E-04
4.000	.3167E-04	.3036E-04	.2910E-04	.2789E-04	.2673E-04	.2561E-04	.2454E-04	.2351E-04	.2252E-04	.2157E-04
4.100	.2066E-04	.1978E-04	.1894E-04	.1814E-04	.1737E-04	.1662E-04	.1591E-04	.1523E-04	.1458E-04	.1395E-04
4.200	.1335E-04	.1277E-04	.1222E-04	.1168E-04	.1118E-04	.1069E-04	.1022E-04	.9774E-05	.9345E-05	.8934E-05
4.300	.8540E-05	.8163E-05	.7801E-05	.7455E-05	.7124E-05	.8807E-05	.6503E-05	.6212E-05	.5934E-05	.5668E-05
4.400	.5413E-05	.5169E-05	.4935E-05	.4712E-05	.4498E-05	.4294E-05	.4098E-05	.3911E-05	.3732E-05	.3561E-05
4.500	.3398E-05	.3241E-05	.3092E-05	.2949E-05	.2813E-05	.2682E-05	.2558E-05	.2439E-05	.2325E-05	.2216E-05
4.600	.2112E-05	.2013E-05	.1919E-05	.1828E-05	.1742E-05	.1660E-05	.1581E-05	.1506E-05	.1434E-05	.1366E-05
4.700	.1301E-05	.1239E-05	.1179E-05	.1123E-05	.1069E-05	.1017E-05	.9680E-06	.9211E-06	.8765E-06	.8339E-06
4.800	.7933E-06	.7547E-06	.7178E-06	.6827E-06	.6492E-06	.6173E-06	.5869E-06	.5580E-06	.5304E-06	.5042E-06
4.900	.4792E-06	.4554E-06	.4327E-06	.4111E-06	.3906E-06	.3711E-06	.3525E-06	.3448E-06	.3179E-06	.3019E-06
5.000	.2867E-06	.2722E-06	.2584E-06	.2452E-06	.2328E-06	.2209E-06	.2096E-06	.1989E-06	.1887E-06	.1790E-06
5.100	.1698E-06	.1611E-06	.1528E-06	.1449E-06	.1374E-06	.1302E-06	.1235E-06	.1170E-06	.1109E-06	.1051E-06

(continued)

TABLE 8.2
Continued

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
5.200	.9964E-07	.9442E-07	.8946E-07	.8476E-07	.8029E-07	.7605E-07	.7203E-07	.6821E-07	.6459E-07	.6116E-07
5.300	.5790E-07	.5481E-07	.5188E-07	.4911E-07	.4647E-07	.4398E-07	.4161E-07	.3937E-07	.3724E-07	.3523E-07
5.400	.3332E-07	.3151E-07	.2980E-07	.2818E-07	.2664E-07	.2518E-07	.2381E-07	.2250E-07	.2127E-07	.2010E-07
5.500	.1899E-07	.1794E-07	.1695E-07	.1601E-07	.1512E-07	.1428E-07	.1349E-07	.1274E-07	.1203E-07	.1135E-07
5.600	.1072E-07	.1012E-07	.9548E-08	.9010E-08	.8503E-08	.8022E-08	.7569E-08	.7140E-08	.6735E-08	.6352E-08
5.700	.5990E-08	.5649E-08	.5326E-08	.5022E-08	.4734E-08	.4462E-08	.4206E-08	.3964E-08	.3735E-08	.3519E-08
5.800	.3316E-08	.3124E-08	.2942E-08	.2771E-08	.2610E-08	.2458E-08	.2314E-08	.2179E-08	.2051E-08	.1931E-08
5.900	.1818E-08	.1711E-08	.1610E-08	.1515E-08	.1425E-08	.1341E-08	.1261E-08	.1186E-08	.1116E-08	.1049E-08
6.000	.9866E-09	.9276E-09	.8721E-09	.8198E-09	.7706E-09	.7242E-09	.6806E-09	.6396E-09	.6009E-09	.5646E-09
6.100	.5303E-09	.4982E-09	.4679E-09	.4394E-09	.4126E-09	.3874E-09	.3637E-09	.3414E-09	.3205E-09	.3008E-09
6.200	.2823E-09	.2649E-09	.2486E-09	.2332E-09	.2188E-09	.2052E-09	.1925E-09	.1805E-09	.1692E-09	.1587E-09
6.300	.1488E-09	.1395E-09	.1308E-09	.1226E-09	.1149E-09	.1077E-09	.1009E-09	.9451E-10	.8854E-10	.8294E-10
6.400	.7769E-10	.7276E-10	.6814E-10	.6380E-10	.5974E-10	.5593E-10	.5235E-10	.4900E-10	.4586E-10	.4292E-10
6.500	.4016E-10	.3758E-10	.3515E-10	.3288E-10	.3077E-10	.2877E-10	.2690E-10	.2516E-10	.2352E-10	.2199E-10
6.600	.2056E-10	.1922E-10	.1796E-10	.1678E-10	.1568E-10	.1465E-10	.1369E-10	.1279E-10	.1195E-10	.1116E-10
6.700	.1042E-10	.9731E-11	.9086E-11	.8483E-11	.7919E-11	.7392E-11	.6900E-11	.6439E-11	.6009E-11	.5607E-11
6.800	.5231E-11	.4880E-11	.4552E-11	.4246E-11	.3960E-11	.3692E-11	.3443E-11	.3210E-11	.2993E-11	.2790E-11
6.900	.2600E-11	.2423E-11	.2258E-11	.2104E-11	.1960E-11	.1826E-11	.1701E-11	.1585E-11	.1476E-11	.1374E-11
7.000	.1280E-11	.1192E-11	.1109E-11	.1033E-11	.9612E-12	.8946E-12	.8325E-12	.7747E-12	.7208E-12	.6706E-12
7.100	.6238E-12	.5802E-12	.5396E-12	.5018E-12	.4667E-12	.4339E-12	.4034E-12	.3750E-12	.3486E-12	.3240E-12
7.200	.3011E-12	.2798E-12	.2599E-12	.2415E-12	.2243E-12	.2084E-12	.1935E-12	.1797E-12	.1669E-12	.1550E-12
7.300	.1439E-12	.1336E-12	.1240E-12	.1151E-12	.1068E-12	.9910E-13	.9196E-13	.8531E-13	.7914E-13	.7341E-13
7.400	.6809E-13	.6315E-13	.5856E-13	.5430E-13	.5034E-13	.4667E-13	.4326E-13	.4010E-13	.3716E-13	.3444E-13
7.500	.3191E-13	.2956E-13	.2739E-13	.2537E-13	.2350E-13	.2176E-13	.2015E-13	.1866E-13	.1728E-13	.1600E-13
7.600	.1481E-13	.1370E-13	.1268E-13	.1174E-13	.1086E-13	.1005E-13	.9297E-14	.8600E-14	.7954E-14	.7357E-14
7.700	.6803E-14	.6291E-14	.5816E-14	.5377E-14	.4971E-14	.4595E-14	.4246E-14	.3924E-14	.3626E-14	.3350E-14
7.800	.3095E-14	.2859E-14	.2641E-14	.2439E-14	.2253E-14	.2080E-14	.1921E-14	.1773E-14	.1637E-14	.1511E-14
7.900	.1395E-14	.1287E-14	.1188E-14	.1096E-14	.1011E-14	.9326E-15	.8602E-15	.7934E-15	.7317E-15	.6747E-15
8.000	.6221E-15	.5735E-15	.5287E-15	.4874E-15	.4492E-15	.4140E-15	.3815E-15	.3515E-15	.3238E-15	.2983E-15
8.100	.2748E-15	.2531E-15	.2331E-15	.2146E-15	.1976E-15	.1820E-15	.1675E-15	.1542E-15	.1419E-15	.1306E-15
8.200	.1202E-15	.1106E-15	.1018E-15	.9361E-16	.8611E-16	.7920E-16	.7284E-16	.6698E-16	.6159E-16	.5662E-16
8.300	.5206E-16	.4785E-16	.4398E-16	.4042E-16	.3715E-16	.3413E-16	.3136E-16	.2881E-16	.2646E-16	.2431E-16
8.400	.2232E-16	.2050E-16	.1882E-16	.1728E-16	.1587E-16	.1457E-16	.1337E-16	.1227E-16	.1126E-16	.1033E-16
8.500	.9480E-17	.8697E-17	.7978E-17	.7317E-17	.6711E-17	.6154E-17	.5643E-17	.5174E-17	.4744E-17	.4348E-17
8.600	.3986E-17	.3653E-17	.3348E-17	.3068E-17	.2811E-17	.2575E-17	.2359E-17	.2161E-17	.1979E-17	.1812E-17
8.700	.1659E-17	.1519E-17	.1391E-17	.1273E-17	.1166E-17	.1067E-17	.9763E-18	.8933E-18	.8174E-18	.7478E-18
8.800	.6841E-18	.6257E-18	.5723E-18	.5234E-18	.4786E-18	.4376E-18	.4001E-18	.3657E-18	.3343E-18	.3055E-18
8.900	.2792E-18	.2552E-18	.2331E-18	.2130E-18	.1946E-18	.1777E-18	.1623E-18	.1483E-18	.1354E-18	.1236E-18
9.000	.1129E-18	.1030E-18	.9404E-19	.8584E-19	.7834E-19	.7148E-19	.6523E-19	.5951E-19	.5429E-19	.4952E-19
9.100	.4517E-19	.4119E-19	.3756E-19	.3425E-19	.3123E-19	.2847E-19	.2595E-19	.2365E-19	.2155E-19	.1964E-19
9.200	.1790E-19	.1631E-19	.1486E-19	.1353E-19	.1232E-19	.1122E-19	.1022E-19	.9307E-20	.8474E-20	.7714E-20
9.300	.7022E-20	.6392E-20	.5817E-20	.5294E-20	.4817E-20	.4382E-20	.3987E-20	.3627E-20	.3299E-20	.3000E-20
9.400	.2728E-20	.2481E-20	.2255E-20	.2050E-20	.1864E-20	.1694E-20	.1540E-20	.1399E-20	.1271E-20	.1155E-20
9.500	.1049E-20	.9533E-21	.8659E-21	.7864E-21	.7142E-21	.6485E-21	.5888E-21	.5345E-21	.4852E-21	.4404E-21
9.600	.3997E-21	.3627E-21	.3292E-21	.2986E-21	.2709E-21	.2458E-21	.2229E-21	.2022E-21	.1834E-21	.1663E-21
9.700	.1507E-21	.1367E-21	.1239E-21	.1123E-21	.1018E-21	.9223E-22	.8358E-22	.7573E-22	.6861E-22	.6215E-22
9.800	.5629E-22	.5098E-22	.4617E-22	.4181E-22	.3786E-22	.3427E-22	.3102E-22	.2808E-22	.2542E-22	.2300E-22
9.900	.2081E-22	.1883E-22	.1704E-22	.1541E-22	.1394E-22	.1261E-22	.1140E-22	.1031E-22	.9323E-23	.8429E-23
10.00	.7620E-23	.6888E-23	.6225E-23	.5626E-23	.5084E-23	.4593E-23	.4150E-23	.3749E-23	.3386E-23	.3058E-23

Notes: (1) E-01 should be read as $\times 10^{-1}$; E-02 should be read as $\times 10^{-2}$, and so on.

(2) This table lists $Q(x)$ for x in the range of 0 to 10 in the increments of 0.01. To find $Q(5.36)$, for example, look up the row starting with $x = 5.3$. The sixth entry in this row (under 0.06) is the desired value 0.4161×10^{-7}.

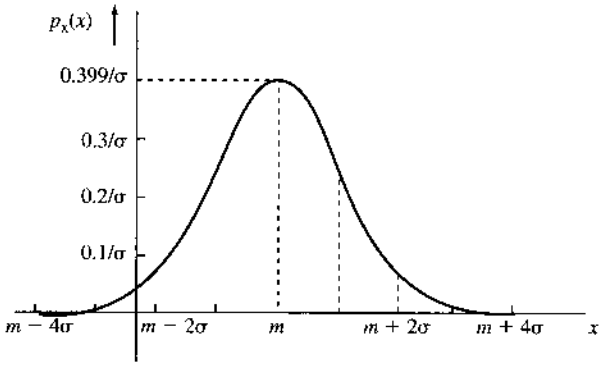


Figure 8.11 Gaussian PDF with mean m and variance σ^2.

Letting $(x - m)/\sigma = z$,

$$F_x(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(x-m)/\sigma} e^{-z^2/2}\, dz$$

$$= 1 - Q\left(\frac{x-m}{\sigma}\right) \tag{8.40a}$$

Therefore,

$$P(x \le x) = 1 - Q\left(\frac{x-m}{\sigma}\right) \tag{8.40b}$$

and

$$P(x > x) = Q\left(\frac{x-m}{\sigma}\right) \tag{8.40c}$$

The Gaussian PDF is perhaps the most important PDF in the field of communications. The majority of the noise processes observed in practice are Gaussian. The amplitude n of a Gaussian noise signal is an RV with a Gaussian PDF. This means the probability of observing n in an interval $(n, n + \Delta n)$ is $p_n(n)\Delta n$, where $p_n(n)$ is of the form in Eq. (8.39) [with $m = 0$].

Example 8.16 Threshold Detection

Over a certain binary channel, messages m= **0** and **1** are transmitted with equal probability by using a positive and a negative pulse, respectively. The received pulse corresponding to **1** is $p(t)$, shown in Fig. 8.12a, and the received pulse corresponding to **0** is $-p(t)$. Let the peak amplitude of $p(t)$ be A_p at $t = T_p$. Because of the channel noise $n(t)$, the received pulses will be (Fig. 8.12c)

$$\pm p(t) + n(t)$$

To detect the pulses at the receiver, each pulse is sampled at its peak amplitude. In the absence of noise, the sampler output is either A_p (for m=**1**) or $-A_p$ (for m=**0**). Because of the channel noise, the sampler output is $\pm A_p + n$, where n, the noise amplitude at the sampling instant (Fig. 8.12b), is an RV. For Gaussian noise, the PDF of n is (Fig. 8.12b)

$$p_n(n) = \frac{1}{\sigma_n\sqrt{2\pi}} e^{-n^2/2\sigma_n^2} \tag{8.41}$$

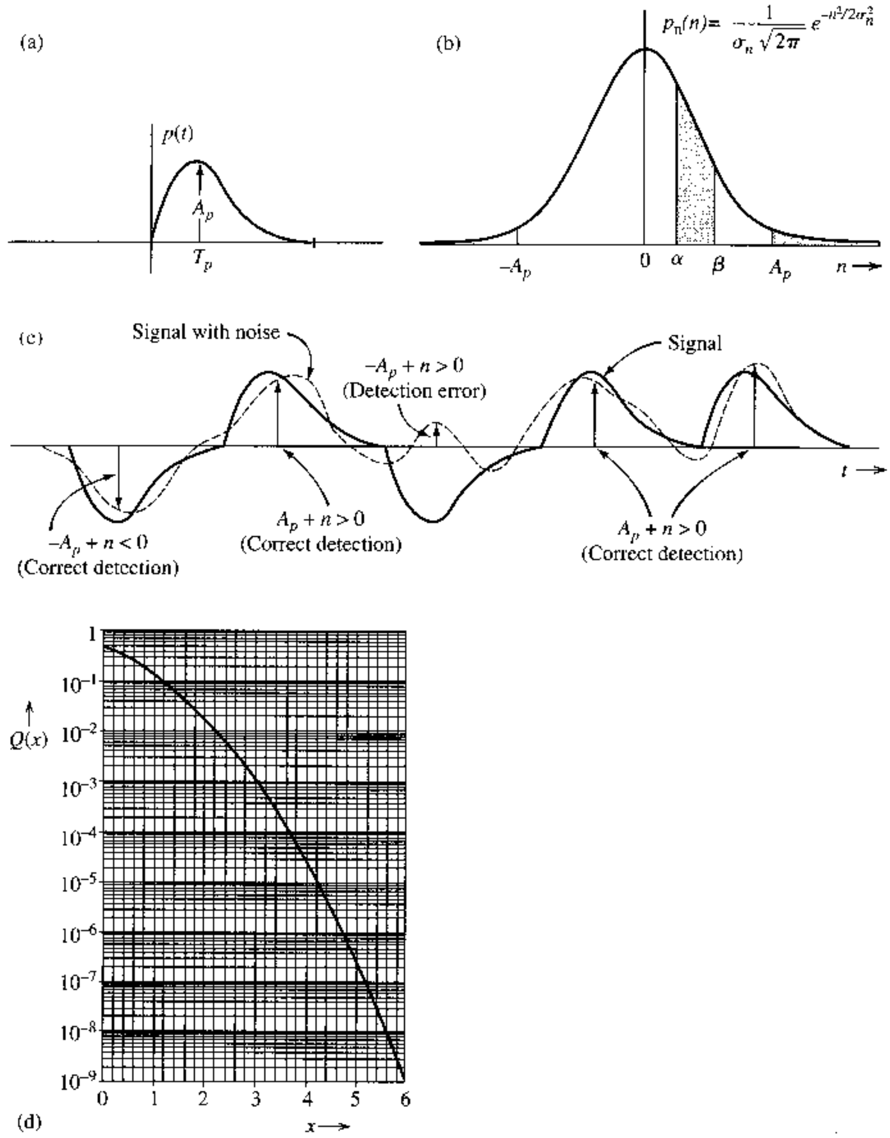


Figure 8.12 Error probability in threshold detection: (a) transmitted pulse; (b) noise PDF; (c) received pulses with noise; (d) detection error probability.

Because of the symmetry of the situation, the optimum detection threshold is zero; that is, the received pulse is detected as a **1** or a **0**, depending on whether the sample value is positive or negative.

Because noise amplitudes range from $-\infty$ to ∞, the sample value $-A_p + \mathrm{n}$ can occasionally be positive, causing the received **0** to be read as **1** (see Fig. 8.12b). Similarly, $A_p + \mathrm{n}$ can occasionally be negative, causing the received **1** to be read as **0**. If **0** is transmitted, it will be detected as **1** if $-A_p + \mathrm{n} > 0$, that is, if $\mathrm{n} > A_p$.

If $P(\epsilon|\mathbf{0})$ is the error probability given that **0** is transmitted, then

$$P(\epsilon|\mathbf{0}) = P(\mathrm{n} > A_p)$$

Because $P(\mathrm{n} > A_p)$ is the shaded area in Fig. 8.12b to the right of A_p, from Eq. (8.40c) [with m = **0**] it follows that

$$P(\epsilon|\mathbf{0}) = Q\left(\frac{A_p}{\sigma_\mathrm{n}}\right) \tag{8.42a}$$

Similarly,

$$\begin{aligned} P(\epsilon|\mathbf{1}) &= P(\mathrm{n} < -A_p) \\ &= Q\left(\frac{A_p}{\sigma_\mathrm{n}}\right) = P(\epsilon|\mathbf{0}) \end{aligned} \tag{8.42b}$$

and

$$\begin{aligned} P_e &= \sum_i P(\epsilon,\, m_i) \\ &= \sum_i P(m_i)P(\epsilon|m_i) \\ &= Q\left(\frac{A_p}{\sigma_n}\right)\sum_i P(m_i) \\ &= Q\left(\frac{A_p}{\sigma_\mathrm{n}}\right) \end{aligned} \tag{8.42c}$$

The error probability P_e can be found from Fig. 8.12d.

Joint Distribution

For two RVs x and y, we define a CDF $F_{\mathrm{xy}}(x,\, y)$ as follows:

$$F_{\mathrm{xy}}(x,\, y) \triangleq P(\mathrm{x} \le x \text{ and } \mathrm{y} \le y) \tag{8.43}$$

and the joint PDF $p_{\mathrm{xy}}(x,\, y)$ as

$$p_{\mathrm{xy}}(x,\, y) = \frac{\partial^2}{\partial x\, \partial y} F_{\mathrm{xy}}(x,\, y) \tag{8.44}$$

Arguing along lines similar to those used for a single variable, we can show that as $\Delta x \to 0$ and $\Delta y \to 0$

$$p_{\mathrm{xy}}(x,\, y)\Delta x\, \Delta y = P(x < \mathrm{x} \le x + \Delta x,\, y < \mathrm{y} \le y + \Delta y) \tag{8.45}$$

Hence, the probability of observing the variables x in the interval $(x,\, x + \Delta x)$ and y in the interval $(y,\, y + \Delta y)$ jointly is given by the volume under the joint PDF $p_{\mathrm{xy}}(x,\, y)$ over the region bounded by $(x,\, x + \Delta x)$ and $(y,\, y + \Delta y)$, as shown in Fig. 8.13a.

From Eq. (8.45), it follows that

$$P(x_1 < x \le x_2,\, y_1 < y \le y_2) = \int_{x_1}^{x_2}\int_{y_1}^{y_2} p_{xy}(x,\, y)\, dx\, dy \tag{8.46}$$

Thus, the probability of jointly observing x in the interval (x_1, x_2) and y in the interval $(y_1,\, y_2)$ is the volume under the PDF over the region bounded by $(x_1,\, x_2)$ and $(y_1,\, y_2)$.

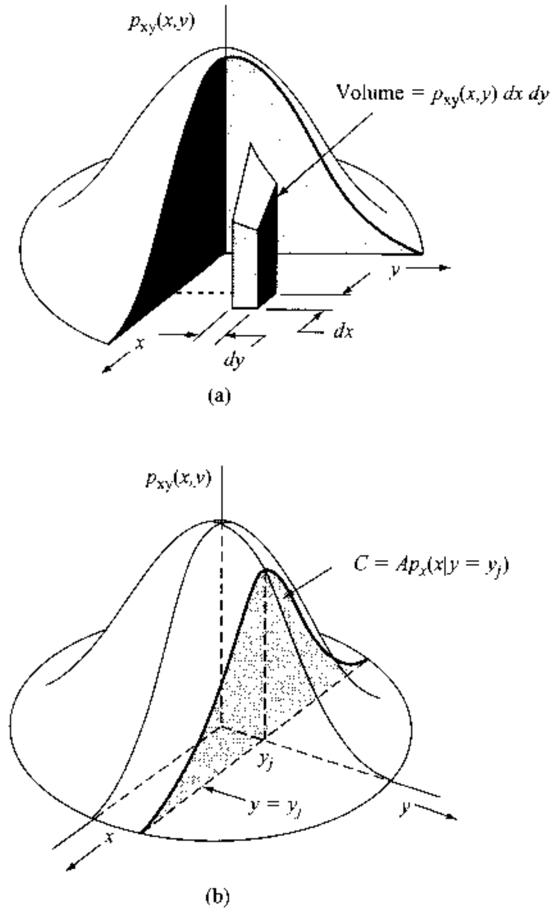


Figure 8.13 (a) Joint PDF. (b) Conditional PDF.

The event of observing x in the interval $(-\infty, \infty)$ and observing y in the interval $(-\infty, \infty)$ is a certainty. Hence,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p_{xy}(x, y)\,dx\,dy = 1 \tag{8.47}$$

Thus, the total volume under the joint PDF must be unity.

When we are dealing with two RVs x and y, the individual probability densities $p_x(x)$ and $p_y(y)$ can be obtained from the joint density $p_{xy}(x, y)$. These individual densities are also called **marginal densities**. To obtain these densities, we note that $p_x(x)\,\Delta x$ is the probability of observing x in the interval $(x, x + \Delta x)$. The value of y may lie anywhere in the interval $(-\infty, \infty)$. Hence,

$$\begin{aligned}
\lim_{\Delta x\to 0} p_x(x)\Delta x &= \lim_{\Delta x\to 0} \text{Probability } (x < \mathrm{x} \le x + \Delta x,\ -\infty < \mathrm{y} \le \infty) \\
&= \lim_{\Delta x\to 0} \int_{x}^{x+\Delta x}\int_{-\infty}^{\infty} p_{xy}(x, y)\,dx\,dy \\
&= \lim_{\Delta x\to 0} \int_{-\infty}^{\infty} p_{xy}(x, y)\,dy \int_{x}^{x+\Delta x} dx \\
&= \lim_{\Delta x\to 0} \Delta x \int_{-\infty}^{\infty} p_{xy}(x, y)\,dy
\end{aligned}$$

The last two steps follow from the fact that $p_{xy}(x, y)$ is constant over $(x, x + \Delta x)$ because $\Delta x \to 0$. Therefore,

$$p_x(x) = \int_{-\infty}^{\infty} p_{xy}(x, y)\, dy \tag{8.48a}$$

Similarly,

$$p_y(y) = \int_{-\infty}^{\infty} p_{xy}(x, y)\, dx \tag{8.48b}$$

In terms of the CDF, we have

$$F_y(y) = F_{xy}(\infty, y) \tag{8.49a}$$

$$F_x(x) = F_{xy}(x, \infty) \tag{8.49b}$$

These results may be generalized for multiple RVs $x_1, x_2, \ldots, x_n$.

Conditional Densities

The concept of conditional probabilities can be extended to the case of continuous RVs. We define the conditional PDF $p_{x|y}(x|y_j)$ as the PDF of x given that y has the value y_j. This is equivalent to saying that $p_{x|y}(x|y_j)\Delta x$ is the probability of observing x in the range $(x, x+\Delta x)$, given that $y = y_j$. The probability density $p_{x|y}(x|y_j)$ is the intersection of the plane $y = y_j$ with the joint PDF $p_{xy}(x, y)$ (Fig. 8.13b). Because every PDF must have unit area, however, we must normalize the area under the intersection curve C to unity to get the desired PDF. Hence, C is $Ap_{x|y}(x|y)$, where A is the area under C. An extension of the results derived for the discrete case yields

$$p_{x|y}(x|y)p_y(y) = p_{xy}(x, y) \tag{8.50a}$$

$$p_{y|x}(y|x)p_x(x) = p_{xy}(x, y) \tag{8.50b}$$

and

$$p_{x|y}(x|y) = \frac{p_{y|x}(y|x)p_x(x)}{p_y(y)} \tag{8.51a}$$

Equation (8.51a) is Bayes' rule for continuous RVs. When we have mixed variables (i.e., discrete and continuous), the mixed form of Bayes' rule is

$$P_{x|y}(x|y)p_y(y) = P_x(x)p_{y|x}(y|x) \tag{8.51b}$$

where x is a discrete RV and y is a continuous RV.*

Note that $p_{x|y}(x|y)$ is still, first and foremost, a probability density function. Thus,

$$\int_{-\infty}^{\infty} p_{x|y}(x|y)\, dx = \frac{\int_{-\infty}^{\infty} p_{xy}(x, y)\, dx}{p_y(y)} = \frac{p_y(y)}{p_y(y)} = 1 \tag{8.52}$$

* It may be worth noting that $P_{x|y}(x|y)$ is conditioned on an event $y = y$ that has probability zero.

Independent Random Variables

The continuous RVs x and y are said to be independent if

$$p_{x|y}(x|y) = p_x(x) \tag{8.53a}$$

In this case from Eqs. (8.53a) and (8.51) it follows that

$$p_{y|x}(y|x) = p_y(y) \tag{8.53b}$$

This implies that for independent RVs x and y,

$$p_{xy}(x, y) = p_x(x)p_y(y) \tag{8.53c}$$

Based on Eq. (8.53c), the joint CDF is also separable:

$$\begin{aligned} F_{xy}(x, y) &= \int_{-\infty}^{x}\int_{-\infty}^{y} p_{xy}(v, w)\,dw\,dv \\ &= \int_{-\infty}^{x} p_x(v)\,dv \cdot \int_{-\infty}^{y} p_y(w)\,dw \\ &= F_x(x) \cdot F_y(y) \end{aligned} \tag{8.54}$$

Example 8.17 Rayleigh Density

The Rayleigh density is characterized by the PDF (Fig. 8.14b)

$$p_r(r) = \begin{cases} \dfrac{r}{\sigma^2} e^{-r^2/2\sigma^2} & r \geq 0 \\ 0 & r < 0 \end{cases} \tag{8.55}$$

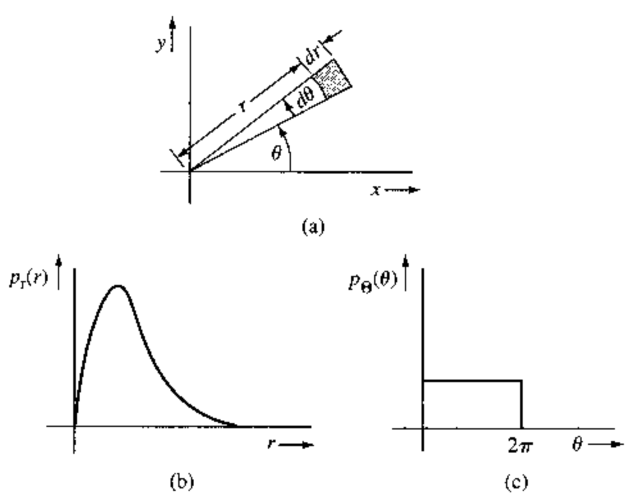


Figure 8.14 Derivation of the Rayleigh density.

A Rayleigh RV can be derived from two independent Gaussian RVs as follows. Let x and y be independent Gaussian variables with identical PDFs:

$$p_x(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2}$$

$$p_y(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-y^2/2\sigma^2}$$

Then

$$p_{xy}(x, y) = p_x(x)p_y(y) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \tag{8.56}$$

The joint density appears somewhat like the bell-shaped surface shown in Fig. 8.13. The points in the (x, y) plane can also be described in polar coordinates as (r, θ), where (Fig. 8.14a)

$$\mathrm{r} = \sqrt{\mathrm{x}^2 + \mathrm{y}^2} \qquad \Theta = \tan^{-1} \frac{\mathrm{y}}{\mathrm{x}}$$

In Fig. 8.14a, the shaded region represents $r < \mathrm{r} \le r + dr$ and $\theta < \Theta \le \theta + d\theta$ (where dr and $d\theta$ both $\to 0$). Hence, if $p_{\mathrm{r}\Theta}(r, \theta)$ is the joint PDF of r and Θ, then by definition [Eq. (8.45)], the probability of observing r and Θ in this region is $p_{\mathrm{r}\Theta}(r, \theta)\, dr\, d\theta$. But we also know that this probability is $p_{xy}(x, y)$ times the area $r\, dr\, d\theta$ of the shaded region.

Hence, [Eq. (8.56)]

$$\frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)2\sigma^2}\, r\, dr\, d\theta = p_{\mathrm{r}\Theta}(r, \theta)\, dr\, d\theta$$

and

$$\begin{aligned} p_{\mathrm{r}\Theta}(r, \theta) &= \frac{r}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \\ &= \frac{r}{2\pi\sigma^2} e^{-r^2/2\sigma^2} \end{aligned} \tag{8.57}$$

and [Eq. (8.48a)]

$$p_{\mathrm{r}}(r) = \int_{-\infty}^{\infty} p_{\mathrm{r}\Theta}(r, \theta)\, d\theta$$

Because Θ exists only in the region $(0, 2\pi)$,

$$\begin{aligned} p_{\mathrm{r}}(r) &= \int_0^{2\pi} \frac{r}{2\pi\sigma^2} e^{-r^2/2\sigma^2}\, d\theta \\ &= \frac{r}{\sigma^2} e^{-r^2/2\sigma^2} u(r) \end{aligned} \tag{8.58a}$$

Note that r is always greater than 0. In a similar way, we find

$$p_\Theta(\theta) = \begin{cases} \frac{1}{2\pi} & 0 \le \Theta < 2\pi \\ 0 & \text{otherwise} \end{cases} \tag{8.58b}$$

RVs r and Θ are independent because $p_{r\Theta}(r, \theta) = p_r(r)p_\Theta(\theta)$. The PDF $p_r(r)$ is the **Rayleigh density function**. We shall later show that the envelope of narrowband Gaussian noise has a Rayleigh density. Both $p_r(r)$ and $p_\Theta(\theta)$ are shown in Fig. 8.14b and c.

8.3 STATISTICAL AVERAGES (MEANS)

Averages are extremely important in the study of RVs. To find a proper definition for the average of a random variable x, consider the problem of determining the average height of the entire population of a country. Let us assume that we have enough resources to gather data about the height of every person. If the data is recorded within the accuracy of an inch, then the height x of every person will be approximated to one of the n numbers $x_1, x_2, \ldots, x_n$. If there are N_i persons of height x_i, then the average height $\bar{x}$ is given by

$$\bar{x} = \frac{N_1 x_1 + N_2 x_2 + \cdots + N_n x_n}{N}$$

where the total number of persons is $N = \sum_i N_i$. Hence,

$$\bar{x} = \frac{N_1}{N} x_1 + \frac{N_2}{N} x_2 + \cdots + \frac{N_n}{N} x_n$$

In the limit as $N \to \infty$, the ratio N_i/N approaches $P_x(x_i)$ according to the relative frequency definition of the probability. Hence,

$$\bar{x} = \sum_{i=1}^{n} x_i P_x(x_i)$$

The mean value is also called the **average value**, or **expected value**, of the RV x and is denoted by $E[x]$. Thus,

$$\bar{x} = E[x] = \sum_i x_i P_x(x_i) \tag{8.59a}$$

We shall use both these notations, our choice depending on the circumstances and convenience.

If the RV x is continuous, an argument similar to that used in arriving at Eq. (8.59a) yields

$$\bar{x} = E[x] = \int_{-\infty}^{\infty} x p_x(x)\, dx \tag{8.59b}$$

This result can be derived by approximating the continuous variable x with a discrete variable by quantizing it in steps of Δx and then letting $\Delta x \to 0$.

Equation (8.59b) is more general and includes Eq. (8.59a), because the discrete RV can be considered as a continuous RV with an impulsive density. In such a case, Eq. (8.59b) reduces to Eq. (8.59a).

As an example, consider the general Gaussian PDF given by (Fig. 8.11)

$$p_x(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-m)^2/2\sigma^2} \tag{8.60a}$$

From Eq. (8.59b) we have

$$\overline{\mathrm{x}} = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-(x-m)^2/2\sigma^2}\, dx$$

Changing the variable to $x = y + m$ yields

$$\begin{aligned}\overline{\mathrm{x}} &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (y+m) e^{-y^2/2\sigma^2}\, dy \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} y e^{-y^2/2\sigma^2}\, dy + m\left[\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2\sigma^2}\, dy\right]\end{aligned}$$

The first integral inside the bracket is zero, because the integrand is an odd function of y. The term inside the square brackets is the integration of the Gaussian PDF, and is equal to 1. Hence,

$$\overline{\mathrm{x}} = m \tag{8.60b}$$

Mean of a Function of a Random Variable

It is often necessary to find the mean value of a function of a RV. For instance, in practice we are often interested in the mean square amplitude of a signal. The mean square amplitude is the mean of the square of the amplitude x, that is, $\overline{\mathrm{x}^2}$.

In general, we may seek the mean value of an RV y that is a function of the RV x; that is, we wish to find $\overline{\mathrm{y}}$ where $\mathrm{y} = g(\mathrm{x})$. Let x be a discrete RV that takes values $x_1, x_2, \ldots, x_n$ with probabilities $P_\mathrm{x}(x_1), P_\mathrm{x}(x_2), \ldots, P_\mathrm{x}(x_n)$, respectively. But because $\mathrm{y} = g(\mathrm{x})$, y takes values $g(x_1), g(x_2), \ldots, g(x_n)$ with probabilities $P_\mathrm{x}(x_1), P_\mathrm{x}(x_2), \ldots, P_\mathrm{x}(x_n)$, respectively. Hence, from Eq. (8.59a) we have

$$\overline{\mathrm{y}} = \overline{g(\mathrm{x})} = \sum_{i=1}^{n} g(x_i) P_\mathrm{x}(x_i) \tag{8.61a}$$

If x is a continuous RV, a similar line of reasoning leads to

$$\overline{g(\mathrm{x})} = \int_{-\infty}^{\infty} g(x) p_\mathrm{x}(x)\, dx \tag{8.61b}$$

Example 8.18 The output voltage of sinusoid generator is $A \cos \omega t$. This output is sampled randomly (Fig. 8.15a). The sampled output is an RV x, which can take on any value in the range $(-A, A)$. Determine the mean value $(\overline{\mathrm{x}})$ and the mean square value $(\overline{\mathrm{x}^2})$ of the sampled output x.

If the output is sampled at a random instant t, the output x is a function of the RV t:

$$\mathrm{x}(t) = A \cos \omega t$$

If we let $\omega t = \Theta$, Θ is also an RV, and if we consider only modulo-2π values of Θ, then the RV Θ lies in the range $(0, 2\pi)$. Because t is randomly chosen, Θ can take any value in the range $(0, 2\pi)$ with uniform probability. Because the area under the PDF must be unity, $p_\Theta(\theta)$ is as shown in Fig. 8.15b.

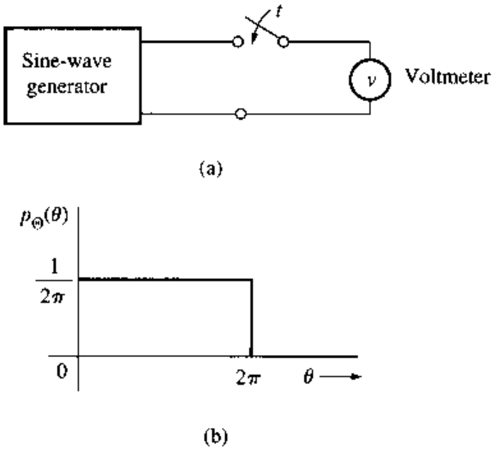


Figure 8.15 Random sampling of a sine-wave generator.

The RV x is thus a function of another RV, Θ,

$$\mathrm{x} = A \cos \Theta$$

Hence, from Eq. (8.61b),

$$\overline{\mathrm{x}} = \int_0^{2\pi} x p_\Theta(\theta)\, d\theta = \frac{1}{2\pi} \int_0^{2\pi} A \cos\theta \, d\theta = 0$$

and

$$\overline{\mathrm{x}^2} = \int_0^{2\pi} x^2 p_\Theta(\theta)\, d\theta = \frac{A^2}{2\pi} \int_0^{2\pi} \cos^2\theta \, d\theta$$

$$= \frac{A^2}{2}$$

Similarly, for the case of two variables x and y, we have

$$\overline{g(\mathrm{x}, \mathrm{y})} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) p_{\mathrm{xy}}(x, y)\, dx\, dy \tag{8.62}$$

Mean of the Sum

If $g_1(\mathrm{x}, \mathrm{y}), g_2(\mathrm{x}, \mathrm{y}), \ldots, g_n(\mathrm{x}, \mathrm{y})$ are functions of the RVs x and y, then

$$\overline{g_1(\mathrm{x}, \mathrm{y}) + g_2(\mathrm{x}, \mathrm{y}) + \cdots + g_n(\mathrm{x}, \mathrm{y})} = \overline{g_1(\mathrm{x}, \mathrm{y})} + \overline{g_2(\mathrm{x}, \mathrm{y})} + \cdots + \overline{g_n(\mathrm{x}, \mathrm{y})} \tag{8.63a}$$

The proof is trivial and follows directly from Eq. (8.62).

Thus, the mean (expected value) of the sum is equal to the sum of the means. An important special case is

$$\overline{\mathrm{x} + \mathrm{y}} = \overline{\mathrm{x}} + \overline{\mathrm{y}} \tag{8.63b}$$

Equation (8.63a) can be extended to functions of any number of RVs.

Mean of the Product of Two Functions

Unfortunately, there is no simple result [as in Eq. (8.63)] for the product of two functions. For the special case where

$$g(\mathrm{x}, \mathrm{y}) = g_1(\mathrm{x})g_2(\mathrm{y}) \tag{8.64a}$$

$$\overline{g_1(\mathrm{x})g_2(\mathrm{y})} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_1(\mathrm{x})g_2(\mathrm{y})p_{\mathrm{xy}}(x, y)\,dx\,dy$$

If x and y are independent, then [Eq. (8.53c)]

$$p_{\mathrm{xy}}(x, y) = p_{\mathrm{x}}(x)p_{\mathrm{y}}(y)$$

and

$$\overline{g_1(\mathrm{x})g_2(\mathrm{y})} = \int_{-\infty}^{\infty} g_1(x)p_{\mathrm{x}}(x)\,dx \int_{-\infty}^{\infty} g_2(y)p_{\mathrm{y}}(y)\,dy$$

$$= \overline{g_1(\mathrm{x})}\;\overline{g_2(\mathrm{y})} \qquad \text{if x and y independent} \tag{8.64b}$$

A special case of this is

$$\overline{\mathrm{xy}} = \overline{\mathrm{x}}\;\overline{\mathrm{y}} \qquad \text{if x and y independent} \tag{8.64c}$$

Moments

The nth **moment** of an RV x is defined as the mean value of x^n. Thus, the nth moment of x is

$$\overline{\mathrm{x}^n} \triangleq \int_{-\infty}^{\infty} x^n p_{\mathrm{x}}(x)\,dx \tag{8.65a}$$

The nth **central moment** of an RV x is defined as

$$\overline{(\mathrm{x}-\overline{\mathrm{x}})^n} \triangleq \int_{-\infty}^{\infty} (x-\overline{\mathrm{x}})^n p_{\mathrm{x}}(x)\,dx \tag{8.65b}$$

The second central moment of an RV x is of special importance. It is called the **variance** of x and is denoted by σ_{x}^2, where σ_{x} is known as the **standard deviation (SD)** of the RV x. By definition,

$$\sigma_{\mathrm{x}}^2 = \overline{(\mathrm{x}-\overline{\mathrm{x}})^2}$$

$$= \overline{\mathrm{x}^2} - 2\overline{\mathrm{x}\overline{\mathrm{x}}} + \overline{\mathrm{x}}^2 = \overline{\mathrm{x}^2} - 2\overline{\mathrm{x}}^2 + \overline{\mathrm{x}}^2$$

$$= \overline{\mathrm{x}^2} - \overline{\mathrm{x}}^2 \tag{8.66}$$

Thus, the variance of x is equal to the mean square value minus the square of the mean. When the mean is zero, the variance is the mean square; that is, $\overline{\mathrm{x}^2} = \sigma_{\mathrm{x}}^2$.

Example 8.19 Find the mean square and the variance of the Gaussian RV with the PDF in Eq. (8.39) [see Fig. 8.11].

We have

$$\overline{\mathrm{x}^2} = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-(x-m)^2/2\sigma^2}\, dx$$

Changing the variable to $y = (x - m)/\sigma$ and integrating, we get

$$\overline{\mathrm{x}^2} = \sigma^2 + m^2 \tag{8.67a}$$

Also, from Eqs. (8.66) and (8.60b),

$$\begin{aligned} \sigma_{\mathrm{x}}^2 &= \overline{\mathrm{x}^2} - \bar{\mathrm{x}}^2 \\ &= (\sigma^2 + m^2) - (m)^2 \\ &= \sigma^2 \end{aligned} \tag{8.67b}$$

Hence, a Gaussian RV described by the density in Eq. (8.60a) has mean m and variance σ^2. In other words, the Gaussian density function is completely specified by the first moment ($\bar{\mathrm{x}}$) and the second moment ($\overline{\mathrm{x}^2}$).

Example 8.20 Mean Square of the Uniform Quantization Error in PCM

In the PCM scheme discussed in Chapter 6, a signal band-limited to B Hz is sampled at a rate of $2B$ samples per second. The entire range $(-m_p,\ m_p)$ of the signal amplitudes is partitioned into L uniform intervals, each of magnitude $2m_p/L$ (Fig. 8.16a). Each sample is approximated to the midpoint of the interval in which it falls. Thus, sample m in Fig. 8.16a is approximated by a value $\hat{m}$, the midpoint of the interval in which m falls. Each sample is thus approximated (quantized) to one of the L numbers.

The difference $\mathrm{q} = \mathrm{m} - \hat{\mathrm{m}}$ is the quantization error and is an RV. We shall determine $\overline{\mathrm{q}^2}$, the mean square value of the quantization error. From Fig. 8.16a it can be seen that q is a continuous RV existing over the range $(-m_p/L,\ m_p/L)$ and is zero outside this range. If we assume that it is equally likely for the sample to lie anywhere in the quantizing interval,* then the PDF of q is uniform

$$p_{\mathrm{q}}(q) = L/2m_p \qquad q \in (-m_p/L,\ m_p/L)$$

* Because the quantizing interval is generally very small, variations in the PDF of signal amplitudes over the interval are small and this assumption is reasonable.

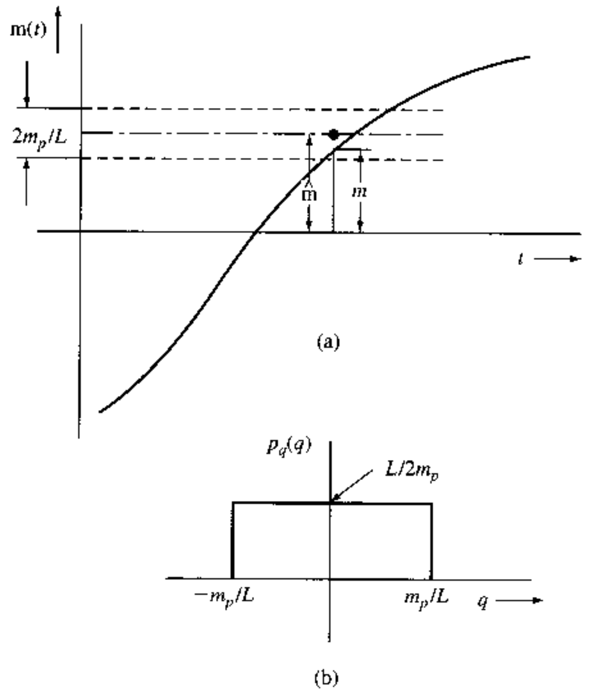


Figure 8.16 (a) Quantization error in PCM and (b) its PDF.

as shown in Fig. 8.16b, and

$$\overline{q^2} = \int_{-m_p/L}^{m_p/L} q^2 p_q(q)\, dq$$

$$= \frac{L}{2m_p} \frac{q^3}{3} \bigg|_{-m_p/L}^{m_p/L}$$

$$= \frac{1}{3}\left(\frac{m_p}{L}\right)^2 \tag{8.68a}$$

From Fig. 8.16b it can be seen that $\overline{q} = 0$. Hence,

$$\sigma_q^2 = \overline{q^2} = \frac{1}{3}\left(\frac{m_p}{L}\right)^2 \tag{8.68b}$$

Example 8.21 Mean Square Error Caused by Channel Noise in PCM

Quantization noise is one of the sources of error in PCM. The other source of error is channel noise. Each quantized sample is coded by a group of n binary pulses. Because of channel noise, some of these pulses are incorrectly detected at the receiver. Hence, the decoded sample value $\tilde{m}$ at the receiver will differ from the quantized sample value $\hat{m}$ that is transmitted. The error $\epsilon = \hat{m} - \tilde{m}$ is a random variable. Let us calculate $\overline{\epsilon^2}$, the mean square error in the sample value caused by the channel noise.

To begin with, let us determine the values that ϵ can take and the corresponding probabilities. Each sample is transmitted by n binary pulses. The value of ϵ depends on the position of the incorrectly detected pulse. Consider, for example, the case of $L = 16$ transmitted by four binary pulses ($n = 4$), as shown in Fig. 1.5. Here the transmitted code **1101** represents a value of 13. A detection error in the first digit changes the received code to **0101**, which is a value of 5. This causes an error $\epsilon = 8$. Similarly, an error in the second digit gives $\epsilon = 4$. Errors in the third and the fourth digits will give $\epsilon = 2$ and $\epsilon = 1$, respectively. In general, the error in the ith digit causes an error $\epsilon_i = (2^{-i})16$. For a general case, the error $\epsilon_i = (2^{-i})F$, where F is the full scale, that is, $2m_p$, in PCM. Thus,

$$\epsilon_i = (2^{-i})(2m_p) \qquad i = 1,\ 2,\ \ldots, n$$

Note that the error ϵ is a discrete RV. Hence,*

$$\overline{\epsilon^2} = \sum_{i=1}^{n} \epsilon_i^2 P_\epsilon(\epsilon_i) \tag{8.69}$$

Because $P_\epsilon(\epsilon_i)$ is the probability that $\epsilon = \epsilon_i$, $P_\epsilon(\epsilon_i)$ is the probability of error in the detection of the ith digit. Because the error probability of detecting any one digit is the same as that of any other, that is, P_e,

$$\begin{aligned}\overline{\epsilon^2} &= P_e \sum_{i=1}^{n} \epsilon_i^2 \\ &= P_e \sum_{i=1}^{n} 4m_p^2(2^{-2i}) \\ &= 4m_p^2 P_e \sum_{i=1}^{n} 2^{-2i}\end{aligned}$$

This summation is a geometric progression with a common ratio $r = 2^{-2}$, with the first term $a_1 = 2^{-2}$ and the last term $a_n = 2^{-2n}$. Hence (see Appendix E.4),

$$\begin{aligned}\overline{\epsilon^2} &= 4m_p^2 P_e \left[\frac{(2^{-2})2^{-2n} - 2^{-2}}{2^{-2} - 1}\right] \\ &= \frac{4m_p^2 P_e (2^{2n} - 1)}{3(2^{2n})}\end{aligned} \tag{8.70a}$$

Note that the magnitude of the error ϵ varies from $2^{-1}(2m_p)$ to $2^{-n}(2m_p)$. The error ϵ can be positive as well as negative. For example, $\epsilon = 8$ because of a first-digit error in **1101**. But the corresponding error ϵ will be -8 if the transmitted code is **0101**. Of course the sign of ϵ does not matter in Eq. (8.69). It must be remembered, however, that ϵ varies from $-2^{-n}(2m_p)$ to $2^{-n}(2m_p)$ and its probabilities are symmetrical about

* Here we are assuming that the error can occur only in one of the n digits. But more than one digit may be in error. Because the digit error probability $P_e \ll 1$ (on the order 10^{-5} or less), however, the probability of more than one wrong digit is extremely small (see Example 8.6), and its contribution $\epsilon_i^2 P_\epsilon(\epsilon_i)$ is negligible.

$\epsilon = 0$. Hence, $\overline{\epsilon} = 0$ and

$$\sigma_\epsilon^2 = \overline{\epsilon^2} = \frac{4m_p^2 P_e(2^{2n} - 1)}{3(2^{2n})} \tag{8.70b}$$

Variance of a Sum of Independent Random Variables

The variance of a sum of independent RVs is equal to the sum of their variances. Thus, if x and y are independent RVs and

$$z = x + y$$

then

$$\sigma_z^2 = \sigma_x^2 + \sigma_y^2 \tag{8.71}$$

This can be shown as follows:

$$\begin{aligned}\sigma_z^2 = \overline{(z - \bar{z})^2} &= \overline{[x + y - (\bar{x} + \bar{y})]^2} \\ &= \overline{[(x - \bar{x}) + (y - \bar{y})]^2} \\ &= \overline{(x - \bar{x})^2} + \overline{(y - \bar{y})^2} + \overline{2(x - \bar{x})(y - \bar{y})} \\ &= \sigma_x^2 + \sigma_y^2 + \overline{2(x - \bar{x})(y - \bar{y})}\end{aligned}$$

Because x and y are independent RVs, $(x - \bar{x})$ and $(y - \bar{y})$ are also independent RVs. Hence, from Eq. (8.64b) we have

$$\overline{(x - \bar{x})(y - \bar{y})} = \overline{(x - \bar{x})} \cdot \overline{(y - \bar{y})}$$

But

$$\overline{(x - \bar{x})} = \bar{x} - \bar{x} = \bar{x} - \bar{x} = 0$$

Similarly,

$$\overline{(y - \bar{y})} = 0$$

and

$$\sigma_z^2 = \sigma_x^2 + \sigma_y^2$$

This result can be extended to any number of variables. If RVs x and y both have zero means (i.e., $\bar{x} = \bar{y} = 0$), then $\bar{z} = \bar{x} + \bar{y} = 0$. Also, because the variance equals the mean square value when the mean is zero, it follows that

$$\overline{z^2} = \overline{(x + y)^2} = \overline{x^2} + \overline{y^2} \tag{8.72}$$

provided $\bar{x} = \bar{y} = 0$, and provided x and y are independent RVs.

Example 8.22 Total Mean Square Error in PCM

In PCM, as seen in Examples 8.20 and 8.21, a signal sample m is transmitted as a quantized sample $\hat{m}$, causing a quantization error $q = m - \hat{m}$. Because of channel noise, the transmitted sample $\hat{m}$ is read as $\tilde{m}$, causing a detection error $\epsilon = \hat{m} - \tilde{m}$. Hence, the actual signal sample m is received as $\tilde{m}$ with a total error

$$m - \tilde{m} = (m - \hat{m}) + (\hat{m} - \tilde{m}) = q + \epsilon$$

where both q and ϵ are zero mean RVs. Because the quantization error q and the channel-noise error ϵ are independent, the mean square of the sum is [see Eq. (8.72)]

$$\begin{aligned}\overline{(m - \tilde{m})^2} &= \overline{(q + \epsilon)^2} = \overline{q^2} + \overline{\epsilon^2} \\ &= \frac{1}{3}\left(\frac{m_p}{L}\right)^2 + \frac{4m_p^2 P_\epsilon (2^{2n} - 1)}{3(2^{2n})}\end{aligned}$$

Also, because $L = 2^n$,

$$\overline{(m - \tilde{m})^2} = \overline{q^2} + \overline{\epsilon^2} = \frac{m_p{}^2}{3(2^{2n})}[1 + 4P_\epsilon(2^{2n} - 1)] \tag{8.73}$$

Chebyshev's Inequality

The standard deviation σ_x of an RV x is a measure of the width of its PDF. The larger the σ_x, the wider the PDF. Figure 8.17 illustrates this effect for a Gaussian PDF. Chebyshev's inequality is a statement of this fact. It states that for a zero mean RV x

$$P(|x| \leq k\sigma_x) \geq 1 - \frac{1}{k^2} \tag{8.74}$$

This means the probability of observing x within a few standard deviations is very high. For example, the probability of finding $|x|$ within $3\sigma_x$ is equal to or greater than 0.88. Thus, for a PDF with $\sigma_x = 1$, $P(|x| \leq 3) \geq 0.88$, whereas for a PDF with $\sigma_x = 3$, $P(|x| \leq 9) \geq 0.88$. It is clear that the PDF with $\sigma_x = 3$ is spread out much more than the PDF with $\sigma_x = 1$. Hence,

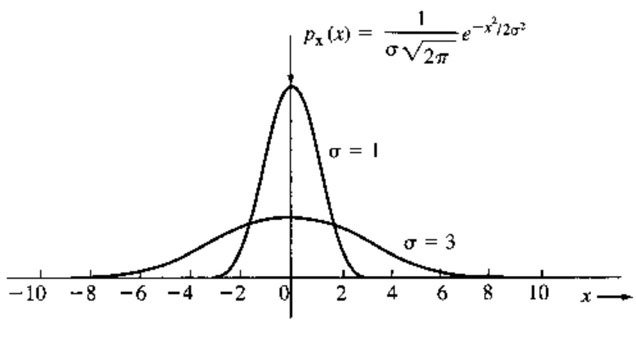


Figure 8.17 Gaussian PDF with standard deviations $\sigma = 1$ and $\sigma = 3$.

σ_x or σ_x^2 is often used as a measure of the width of a PDF. In Chapter 10, we shall use this measure to estimate the bandwidth of a signal spectrum. The proof of Eq. (8.74) is as follows:

$$\sigma_x^2 = \int_{-\infty}^{\infty} x^2 p_x(x)\, dx$$

Because the integrand is positive,

$$\sigma_x^2 \geq \int_{|x| \geq k\sigma_x} x^2 p_x(x)\, dx$$

If we replace x by its smallest value $k\sigma_x$, the inequality still holds,

$$\sigma_x^2 \geq k^2\sigma_x^2 \int_{|x| \geq k\sigma_x} p_x(x)\, dx = k^2\sigma_x^2 P(|x| \geq k\sigma_x)$$

or

$$P(|x| \geq k\sigma_x) \leq \frac{1}{k^2}$$

Hence,

$$P(|x| < k\sigma_x) \geq 1 - \frac{1}{k^2}$$

This inequality can be generalized for a nonzero mean RV as:

$$P(|x - \bar{x}| < k\sigma_x) \geq 1 - \frac{1}{k^2} \tag{8.75}$$

Example 8.23 Estimate the width, or spread, of a Gaussian PDF [Eq. (8.60a)]

For a Gaussian RV [see Eqs. (8.35) and (8.40b)]

$$P(|x - \bar{x}| < \sigma) = 1 - 2Q(1) = 0.6826$$
$$P(|x - \bar{x}| < 2\sigma) = 1 - 2Q(2) = 0.9546$$
$$P(|x - \bar{x}| < 3\sigma) = 1 - 2Q(3) = 0.9974$$

This means that the area under the PDF over the interval $(\bar{x} - 3\sigma,\ \bar{x} + 3\sigma)$ is 99.74% of the total area. A negligible fraction (0.26%) of the area lies outside this interval. Hence, the width, or spread, of the Gaussian PDF may be considered roughly $\pm 3\sigma$ about its mean, giving a total width of roughly 6σ.

8.4 CORRELATION

Often we are interested in determining the nature of dependence between two entities, such as smoking and lung cancer. Consider a random experiment with two outcomes described by

RVs x and y. We conduct several trials of this experiment and record values of x and y for each trial. From this data, it may be possible to determine the nature of a dependence between x and y. The covariance of RVs x and y is one measure that is simple to compute and can yield useful information about the dependence between x and y.

The covariance σ_{xy} of two RVs is defined as

$$\sigma_{xy} \triangleq \overline{(x - \bar{x})(y - \bar{y})} \tag{8.76}$$

Note that the concept of covariance is a natural extension of the concept of variance, which is defined as

$$\sigma_x^2 = \overline{(x - \bar{x})(x - \bar{x})}$$

Let us consider a case of two variables x and y that are dependent such that they tend to vary in harmony; that is, if x increases y increases, and if x decreases y also decreases. For instance, x may be the average daily temperature of a city and y the volume of soft drink sales that day in the city. It is reasonable to expect the two quantities to vary in harmony for a majority of the cases. Suppose we consider the following experiment: pick a random day and record the average temperature of that day as the value of x and the soft drink sales volume that day as the value of y. We perform this measurement over several days (several trials of the experiment) and record the data x and y for each trial. We now plot points (x, y) for all the trials. This plot, known as the **scatter diagram**, may appear as shown in Fig. 8.18a. The plot shows that when x is large, y is likely to be large. Note the use of the word *likely*. It is not *always* true that y will be large if x is large, but it is true most of the time. In other words, in a few cases, a low average temperature will be paired with higher soft drink sales owing to some atypical situation, such as a major soccer match. This is quite obvious from the scatter diagram in Fig. 8.18a.

To continue this example, the variable $x - \bar{x}$ represents the difference between actual and average values of x, and $y - \bar{y}$ represents the difference between actual and average values of y. It is more instructive to plot $(y - \bar{y})$ vs. $(x - \bar{x})$. This is the same as the scatter diagram in Fig. 8.18a with the origin shifted to $(\bar{x}, \bar{y})$, as in Fig. 8.18b, which shows that a day with an above-average temperature is likely to produce above-average soft drink sales, and a day with a below-average temperature is likely to produce below-average soft drink sales.

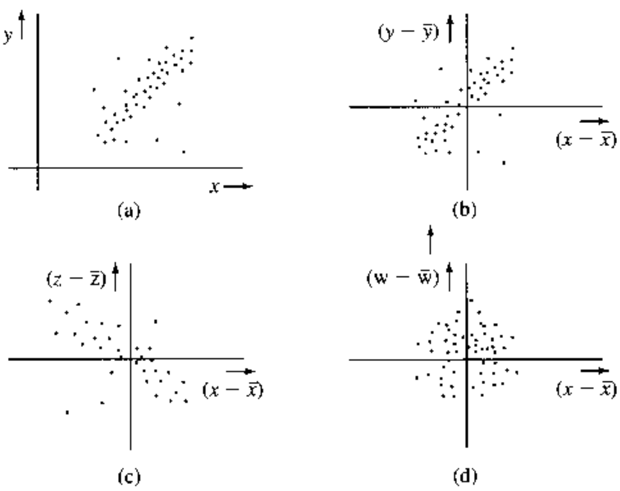


Figure 8.18 Scatter diagrams: (a), (b) positive correlation; (c) negative correlation; (d) zero correlation.

That is, if $x - \bar{x}$ is positive, $y - \bar{y}$ is likely to be positive, and if $x - \bar{x}$ is negative, $y - \bar{y}$ is more likely to be negative. Thus, the quantity $(x - \bar{x})(y - \bar{y})$ will be positive for most trials. We compute this product for every pair, add these products, and then divide by the number of trials. The result is the mean value of $(x - \bar{x})(y - \bar{y})$, that is, the covariance $\sigma_{xy} = \overline{(x - \bar{x})(y - \bar{y})}$. The covariance will be positive in the example under consideration. In such cases, we say that a positive correlation exists between variables x and y. We can conclude that a positive correlation implies variation of two variables in harmony (in the same direction, up or down).

Next, we consider the case of the two variables: x, the average daily temperature, and z, the sales volume of sweaters that day. It is reasonable to believe that as x (daily average temperature) increases, z (the sweater sales volume) tends to decrease. A hypothetical scatter diagram for this experiment is shown in Fig. 8.18c. Thus, if $x - \bar{x}$ is positive (above-average temperature), $z - \bar{z}$ is likely to be negative (below-average sweater sales). Similarly, when $x - \bar{x}$ is negative, $z - \bar{z}$ is likely to be positive. The product $(x - \bar{x})(z - \bar{z})$ will be negative for most of the trials, and the mean $\overline{(x - \bar{x})(z - \bar{z})} = \sigma_{xz}$ will be negative. In such a case, we say that negative correlation exists between x and y. It should be stressed here that negative correlation does not mean that x and y are unrelated. It means that they are dependent, but when one increases, the other decreases, and vice versa.

Last, consider the variables x (the average daily temperature) and w (the number of births). It is reasonable to expect that the daily temperature has little to do with the number of children born. A hypothetical scatter diagram for this case will appear as shown in Fig. 8.18d. If $x - \bar{x}$ is positive, $w - \bar{w}$ is equally likely to be positive or negative. The product $(x - \bar{x})(w - \bar{w})$ is therefore equally likely to be positive or negative, and the mean $\overline{(x - \bar{x})(w - \bar{w})} = \sigma_{xw}$ will be zero. In such a case, we say that RVs x and w are **uncorrelated.**

To reiterate, if σ_{xy} is positive (or negative), then x and y are said to have a positive (or negative) correlation, and if $\sigma_{xy} = 0$, then the variables x and y are said to be uncorrelated.

From this discussion, it appears that under suitable conditions, covariance can serve as a measure of the dependence of two variables. It often provides *some* information about the interdependence of the two RVs and proves useful in a number of applications.

The covariance σ_{xy} may be expressed in another way, as follows. By definition,

$$\begin{aligned}\sigma_{xy} &= \overline{(x - \bar{x})(y - \bar{y})} \\ &= \overline{xy} - \overline{\bar{x}y} - \overline{x\bar{y}} + \overline{\bar{x}\bar{y}} \\ &= \overline{xy} - \bar{x}\bar{y} - \bar{x}\bar{y} + \bar{x}\bar{y} \\ &= \overline{xy} - \bar{x}\bar{y}\end{aligned} \tag{8.77}$$

From Eq. (8.77) it follows that the variables x and y are uncorrelated ($\sigma_{xy} = 0$) if

$$\overline{xy} = \bar{x}\bar{y} \tag{8.78}$$

The correlation between x and y cannot be directly compared with the correlation between z and w. This is because different RVs may differ in strength. To be fair, the covariance value should be normalized appropriately. For this reason, the definition of **correlation coefficient** is particularly useful. **Correlation coefficient** ρ_{xy} is σ_{xy} normalized by $\sigma_x\sigma_y$,

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x\sigma_y} \tag{8.79}$$

Thus, if x and y are uncorrelated, then $\rho_{xy} = 0$. Also, it can be shown that (Prob. 8.5-5) that

$$-1 \leq \rho_{xy} \leq 1 \tag{8.80}$$

Independence vs. Uncorrelatedness

Note that for independent RVs [Eq. (8.64c)]

$$\overline{xy} = \bar{x}\bar{y} \qquad \text{and} \qquad \sigma_{xy} = 0$$

Hence, independent RVs are uncorrelated. This supports the heuristic argument presented earlier. It should be noted that whereas independent variables are uncorrelated, the converse is not necessarily true—uncorrelated variables are generally not independent (Prob. 8.5-3). Independence is, in general, a stronger and more restrictive condition than uncorrelatedness. For independent variables, we have shown [Eq. (8.64b)] that, when the expectations exist,

$$\overline{g_1(x)g_2(y)} = \overline{g_1(x)}\;\overline{g_2(y)}$$

for any functions $g_1(\cdot)$ and $g_2(\cdot)$, whereas for uncorrelatedness, the only requirement is that

$$\overline{xy} = \bar{x}\,\bar{y}$$

There is only one *special* case for which independence and uncorrelatedness are equivalent—when random variables x and y are jointly Gaussian. Note that when x and y are jointly Gaussian, individually x and y are also Gaussian.

Mean Square of the Sum of Uncorrelated Variables

If x and y are uncorrelated, then for $z = x + y$ we show that

$$\sigma_z^2 = \sigma_x^2 + \sigma_y^2 \tag{8.81}$$

That is, the variance of the sum is the sum of variances for uncorrelated RVs. We have proved this result earlier for independent variables x and y. Following the development after Eq. (8.71), we have

$$\begin{aligned}
\sigma_z^2 &= \overline{[(x - \bar{x}) + (y - \bar{y})]^2} \\
&= \overline{(x - \bar{x})^2} + \overline{(y - \bar{y})^2} + 2\overline{(x - \bar{x})(y - \bar{y})} \\
&= \sigma_x^2 + \sigma_y^2 + 2\sigma_{xy}
\end{aligned}$$

Because x and y are uncorrelated, $\sigma_{xy} = 0$, and Eq. (8.81) follows. If x and y have zero means, then z also has a zero mean, and the mean square values of these variables are equal to their variances. Hence,

$$\overline{(x + y)^2} = \overline{x^2} + \overline{y^2} \tag{8.82}$$

if x and y are uncorrelated and have zero means. Thus, Eqs. (8.81) and (8.82) are valid not only when x and y are independent, but also under the less restrictive condition that x and y be uncorrelated.

8.5 LINEAR MEAN SQUARE ESTIMATION

When two random variables x and y are related (or dependent), then a knowledge of one gives certain information about the other. Hence, it is possible to estimate the value of (parameter or signal) y from a knowledge of the value of x. The estimate of y will be another random variable $\hat{y}$. The estimated random variable $\hat{y}$ will in general be different from the actual y. One may choose various criteria of goodness for estimation. Minimum mean square error is one possible criterion. The optimum estimate in this case minimizes the mean square error $\overline{\epsilon^2}$ given by

$$\overline{\epsilon^2} = \overline{(y - \hat{y})^2}$$

In general, the optimum estimate $\hat{y}$ is a nonlinear function of x.* We simplify the problem by constraining the estimate $\hat{y}$ to be a linear function of x of the form

$$\hat{y} = a x$$

assuming that $\overline{x} = 0$.† In this case,

$$\begin{aligned} \overline{\epsilon^2} &= \overline{(y - \hat{y})^2} = \overline{(y - a x)^2} \\ &= \overline{y^2} + a^2\overline{x^2} - 2a\overline{xy} \end{aligned}$$

To minimize $\overline{\epsilon^2}$, we have

$$\frac{\partial \overline{\epsilon^2}}{\partial a} = 2a\overline{x^2} - 2\overline{xy} = 0$$

Hence,

$$a = \frac{\overline{xy}}{\overline{x^2}} = \frac{R_{xy}}{R_{xx}} \tag{8.83}$$

where $R_{xy} = \overline{xy}$, $R_{xx} = \overline{x^2}$, and $R_{yy} = \overline{y^2}$. Note that for this constant choice of a,

$$\epsilon = y - a x = y - \frac{R_{xy}}{R_{xx}} x$$

Hence,

$$\overline{x\epsilon} = \overline{x\left(y - \frac{R_{xy}}{R_{xx}} x\right)} = \overline{xy} - \frac{R_{xy}}{R_{xx}}\overline{x^2}$$

---

* It can be shown that[5] the optimum estimate $\hat{y}$ is the conditional mean of y when x = x, that is,

$$\hat{y} = E[y \mid x = x]$$

In general, this is a nonlinear function of x.

† Throughout the discussion, the variables x, y, ... will be assumed to have zero mean. This can be done without loss of generality. If the variables have nonzero means, we can form new variables $x' = x - \overline{x}$ and $y' = y - \overline{y}$, and so on. The new variables obviously have zero mean values.

Since by definition $\overline{xy} = R_{xy}$ and $\overline{xx} = \overline{x^2} = R_{xx}$, we have

$$\overline{x\epsilon} = R_{xy} - R_{xy} = 0 \tag{8.84}$$

The condition of Eq. (8.84) is known as the principle of orthogonality. The physical interpretation is that the data (x) used in estimation and the (minimum) error (ϵ) are orthogonal (implying uncorrelatedness in this case) when the mean square error is minimum.

Given the principle of orthogonality, the minimum mean square error is given by

$$\begin{aligned}\overline{\epsilon^2} &= \overline{(y - ax)^2} \\ &= \overline{(y - ax)y} - a \cdot \overline{\epsilon x} \\ &= \overline{(y - ax)y} \\ &= \overline{y^2} - a \cdot \overline{yx} \\ &= R_{yy} - aR_{xy}\end{aligned} \tag{8.85}$$

Using n Random Variables to Estimate a Random Variable

If a random variable x_0 is related to n RVs $x_1, x_2, \ldots, x_n$, then we can estimate x_0 using a linear combination* of $x_1, x_2, \ldots, x_n$:

$$\hat{x}_0 = a_1x_1 + a_2x_2 + \cdots + a_nx_n = \sum_{i=1}^{n} a_ix_i \tag{8.86}$$

The mean square error is given by

$$\overline{\epsilon^2} = \overline{[x_0 - (a_1x_1 + a_2x_2 + \cdots + a_nx_n)]^2}$$

To minimize $\overline{\epsilon^2}$, we must set

$$\frac{\partial\overline{\epsilon^2}}{\partial a_1} = \frac{\partial\overline{\epsilon^2}}{\partial a_2} = \cdots = \frac{\partial\overline{\epsilon^2}}{\partial a_n} = 0$$

that is,

$$\frac{\partial\overline{\epsilon^2}}{\partial a_i} = \frac{\partial}{\partial a_i}\overline{[x_0 - (a_1x_1 + a_2x_2 + \cdots + a_nx_n)]^2} = 0$$

Interchanging the order of differentiation and averaging, we have

$$\frac{\partial\overline{\epsilon^2}}{\partial a_i} = -2\overline{[x_0 - (a_1x_1 + a_2x_2 + \cdots + a_nx_n)]x_i} = 0 \tag{8.87a}$$

Equation (8.87a) can be written as

$$\overline{\epsilon \cdot x_i} = 0 \qquad i = 1, 2, \ldots, n \tag{8.87b}$$

* Throughout this section as before, we assume that all the random variables have zero mean values. This can be done without loss of generality.

It can be rewritten into Yule-Walker equations

$$R_{0i} = a_1 R_{i1} + a_2 R_{i2} + \cdots + a_n R_{in} \tag{8.88}$$

where

$$R_{ij} = \overline{x_i x_j}$$

Differentiating $\overline{\epsilon^2}$ with respect to $a_1, a_2, \ldots, a_n$ and equating to zero, we obtain n simultaneous equations of the form shown in Eq. (8.88). The desired constants $a_1, a_2, \ldots, a_n$ can be found from these equations by matrix inversion

$$\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1n} \\ R_{21} & R_{22} & \cdots & R_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ R_{n1} & R_{n2} & \cdots & R_{nn} \end{bmatrix}^{-1} \begin{bmatrix} R_{01} \\ R_{02} \\ \vdots \\ R_{0n} \end{bmatrix} \tag{8.89}$$

Equation (8.87) shows that ϵ (the error) is orthogonal to data $(x_1, x_2, \ldots, x_n)$ for optimum estimation. This gives the more general form for the principle of orthogonality in mean square estimation. Consequently, the mean square error (under optimum conditions) is

$$\overline{\epsilon^2} = \overline{\epsilon\epsilon} = \overline{\epsilon[x_0 - (a_1 x_1 + a_2 x_2 + \cdots + a_n x_n)]}$$

Because $\overline{\epsilon x_i} = 0$ $(i = 1, 2, \ldots, n)$,

$$\begin{aligned} \overline{\epsilon^2} &= \overline{\epsilon x_0} \\ &= \overline{x_0[x_0 - (a_1 x_1 + a_2 x_2 + \cdots + a_n x_n)]} \\ &= R_{00} - (a_1 R_{01} + a_2 R_{02} + \cdots + a_n R_{0n}) \end{aligned} \tag{8.90}$$

Example 8.24 In differential pulse code modulation (DPCM), instead of transmitting sample values directly, we estimate (predict) the value of each sample from the knowledge of previous n samples. The estimation error ϵ_k, the difference between the actual value and the estimated value of the kth sample, is quantized and transmitted (Fig. 8.19). Because the estimation error ϵ_k is smaller than the sample value m_k, for the same number of quantization levels (the same number of PCM code bits), the SNR is increased. It was shown in Sec. 6.5 that the SNR improvement is equal to $\overline{m^2}/\overline{\epsilon^2}$, where $\overline{m^2}$ and $\overline{\epsilon^2}$ are the mean square values of the speech signal and the estimation error ϵ, respectively. In this example, we shall find the optimum linear second-order predictor and the corresponding SNR improvement.

The equation of a second-order estimator (predictor), shown in Fig. 8.19, is

$$\hat{m}_k = a_1 m_{k-1} + a_2 m_{k-2}$$

where $\hat{m}_k$ is the best linear estimate of m_k. The estimation error ϵ_k is given by

$$\epsilon_k = \hat{m}_k - m_k = a_1 m_{k-1} + a_2 m_{k-2} - m_k$$

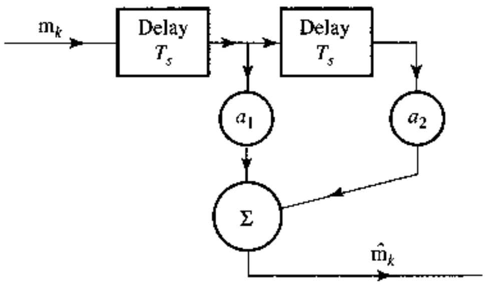


Figure 8.19 Second-order predictor in Example 8.24.

For speech signals, Jayant and Noll[5] give the values of correlations of various samples as:

$$\overline{m_k m_k} = \overline{m^2}, \quad \overline{m_k m_{k-1}} = 0.825\overline{m^2}, \quad \overline{m_k m_{k-2}} = 0.562\overline{m^2},$$

$$\overline{m_k m_{k-3}} = 0.308\overline{m^2}, \quad \overline{m_k m_{k-4}} = 0.004\overline{m^2}, \quad \overline{m_k m_{k-5}} = -0.243\overline{m^2}$$

Note that $R_{ij} = \overline{m_k m_{k-(j-i)}}$. Hence,

$$R_{11} = R_{22} = \overline{m^2}$$
$$R_{12} = R_{21} = R_{01} = 0.825\overline{m^2}$$
$$R_{02} = 0.562\overline{m^2}$$

The optimum values of a_1 and a_2 are found from Eq. (8.89) as $a_1 = 1.1314$ and $a_2 = -0.3714$, and the mean square error in the estimation is given by Eq. (8.90) as

$$\overline{\epsilon^2} = [1 - (0.825a_1 + 0.562a_2)]\overline{m^2} = 0.2753\overline{m^2} \tag{8.91}$$

The SNR improvement is $10 \log_{10} \overline{m^2}/0.2752\overline{m^2} = 5.6$ dB.

8.6 SUM OF RANDOM VARIABLES

In many applications, it is useful to characterize the RV z that is the sum of two RVs x and y:

$$z = x + y$$

Because $z = x + y$, $y = z - x$ regardless of the value of x. Hence, the event $z \le z$ is the joint event [$y \le z - x$ and x to have any value in the range $(-\infty, \infty)$]. Hence,

$$F_z(z) = P(z \le z) = P(x \le \infty, y \le z - x)$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{z-x} p_{xy}(x, y)\, dy\, dx$$
$$= \int_{-\infty}^{\infty} dx \int_{-\infty}^{z-x} p_{xy}(x, y)\, dy$$

and

$$p_z(z) = \frac{dF_z(z)}{dz} = \int_{-\infty}^{\infty} p_{xy}(x,\, z - x)\, dx$$

If x and y are independent RVs, then

$$p_{xy}(x,\, z - x) = p_x(x)p_y(z - x)$$

and

$$p_z(z) = \int_{-\infty}^{\infty} p_x(x)p_y(z - x)\, dx \tag{8.92}$$

The PDF $p_z(z)$ is then the convolution of PDFs $p_x(z)$ and $p_y(z)$. We can extend this result to a sum of n independent RVs $x_1,\, x_2,\, \ldots,\, x_n$. If

$$z = x_1 + x_2 + \cdots + x_n$$

then the PDF $p_z(z)$ will be the convolution of PDFs $p_{x_1}(x),\, p_{x_2}(x),\, \ldots,\, p_{x_n}(x)$, that is,

$$p_z(x) = p_{x_1}(x) * p_{x_2}(x) * \ldots * p_{x_n}(x) \tag{8.93}$$

Sum of Gaussian Random Variables

Gaussian random variables have several very important properties. For example, a Gaussian random variable x and its probability density function $p_x(x)$ are fully described by the mean μ_x and the variance σ_x^2. Furthermore, the sum of any number of jointly distributed Gaussian random variables is also a Gaussian random variable, regardless of their relationships (such as dependency). Again, note that when the members of a set of random variables $\{x_i\}$ are jointly Gaussian, each individual random variable x_i also has Gaussian distribution.

As an example, we will show that the sum of two independent, zero mean, Gaussian random variables is Gaussian. Let x_1 and x_2 be two zero mean and independent Gaussian random variables with probability density functions

$$p_{x_1}(x) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-x^2/(2\sigma_1^2)} \qquad \text{and} \qquad p_{x_2}(x) = \frac{1}{\sqrt{2\pi}\sigma_2} e^{-x^2/(2\sigma_2^2)}$$

Let

$$y = x_1 + x_2$$

The probability density function of y is therefore

$$p_y(y) = \int_{-\infty}^{\infty} p_{x_1}(x)p_{x_2}(y - x)\, dx$$

Upon carrying out this convolution (integration), we have

$$p_{\mathrm{y}}(y) = \frac{1}{2\pi\sigma_1\sigma_2}\int_{-\infty}^{\infty}\exp\left(-\frac{x^2}{2\sigma_1^2} - \frac{(y-x)^2}{2\sigma_2}\right)dx$$

$$= \frac{1}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}}e^{-\frac{y^2}{2(\sigma_1^2+\sigma_2^2)}}\frac{1}{\sqrt{2\pi}\frac{\sigma_1\sigma_2}{\sqrt{\sigma_1^2+\sigma_1^2}}}\int_{-\infty}^{\infty}\exp\left(-\frac{1}{2\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}}\left[x-\frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}y\right]^2\right)dx \tag{8.94}$$

By a simple change of variable

$$w = \frac{\left[x - \frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}y\right]}{\frac{\sigma_1\sigma_2}{\sqrt{\sigma_1^2+\sigma_2^2}}}$$

we can rewrite the integral of Eq. (8.94) as

$$p_{\mathrm{y}}(y) = \frac{1}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}}e^{-\frac{y^2}{2(\sigma_1^2+\sigma_2^2)}}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{1}{2}w^2}dw = \frac{1}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}}e^{-\frac{y^2}{2(\sigma_1^2+\sigma_2^2)}} \tag{8.95}$$

By examining Eq. (8.95), it can be seen that y is a Gaussian RV with zero mean and variance:

$$\sigma_{\mathrm{y}}^2 = \sigma_1^2 + \sigma_2^2$$

In fact, because x_1 and x_2 are independent, they must be uncorrelated. This relationship can be obtained from Eq. (8.81).

More generally,[5] if x_1 and x_2 are jointly Gaussian but not necessarily independent, then $\mathrm{y} = \mathrm{x}_1 + \mathrm{x}_2$ is Gaussian RV with mean

$$\overline{\mathrm{y}} = \overline{\mathrm{x}_1} + \overline{\mathrm{x}_2}$$

and variance

$$\sigma_{\mathrm{y}}^2 = \sigma_{x_1}^2 + \sigma_{x_2}^2 + 2\sigma_{\mathrm{x}_1\mathrm{x}_2}$$

Based on induction, the sum of any number of jointly Gaussian distributed RV's is still Gaussian. More importantly, for any fixed constants $\{a_i,\ i = 1, \ldots, m\}$ and jointly Gaussian RVs $\{\mathrm{x}_i,\ i = 1, \ldots, m\}$,

$$\sum_{i=1}^{m} a_i\mathrm{x}_i$$

remains Gaussian. This result has important practical implications. For example, if x_k is a sequence of jointly Gaussian signal samples passing through a discrete time filter with impulse response $\{h_i\}$, then the filter output

$$\mathrm{y} = \sum_{i=0}^{\infty} h_i\mathrm{x}_{k-i} \tag{8.96}$$

will continue to be Gaussian. The fact that linear filter output to a Gaussian signal input will be a Gaussian signal is highly significant and is one of the most useful results in communication analysis.

8.7 CENTRAL LIMIT THEOREM

Under certain conditions, the sum of a large number of independent RVs tends to be a Gaussian random variable, independent of the probability densities of the variables added.* The rigorous statement of this tendency is what is known as the **central limit theorem.**† Proof of this theorem can be found in the Refs. 6 and 7. We shall give here only a simple plausibility argument.

The tendency toward a Gaussian distribution when a large number of functions are convolved in shown in Fig. 8.20. For simplicity, we assume all PDFs to be identical, that is, a gate function $0.5\,\Pi(x/2)$. Figure 8.20 shows the successive convolutions of gate functions. The tendency toward a bell-shaped density is evident.

This important result that the **distribution** of the sum of n independent Bernoulli random variables, when properly normalized, converges toward Gaussian distribution was established first by A. de Moivre in the early 1700s. The more general proof for an arbitrary distribution was credited to J. W. Lindenber and P. Lévy in the 1920s. Note that the "normalized sum" is the sample average (or sample mean) of n random variables.

Central Limit Theorem (for the sample mean):
Let $x_1, \ldots, x_n$ be independent random samples from a given distribution with mean μ and variance σ^2 with $0 < \sigma^2 < \infty$. Then for any value x, we have

$$\lim_{n\to\infty} P\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{x_i-\mu}{\sigma} \le x\right] = \int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-v^2/2}\,dv \tag{8.97}$$

or equivalently,

$$\lim_{n\to\infty} P\left[\frac{\tilde{x}_n-\mu}{\sigma/\sqrt{n}} > x\right] = Q(x) \tag{8.98}$$

Note that

$$\tilde{x}_n = \frac{x_1+\cdots+x_n}{n}$$

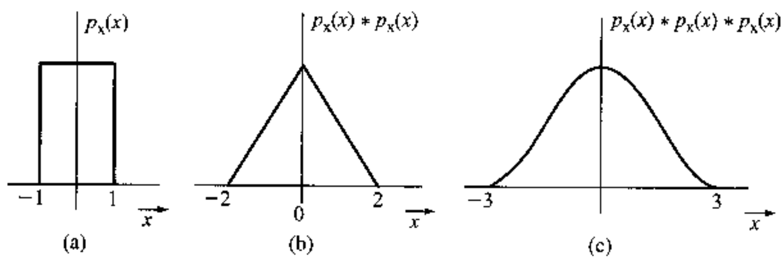


Figure 8.20 Demonstration of the central limit theorem.

* If the variables are Gaussian, this is true even if the variables are not independent.
† Actually, a group of theorems collectively called the central limit theorem.

is known as the sample mean. The interpretation is that the sample mean of any distribution with nonzero finite variance converges to Gaussian distribution with fixed mean μ and decreasing variance σ^2/n. In other words, regardless of the true distribution of x_i, $\sum_{i=1}^{n} x_i$ can be approximated by a Gaussian distribution with mean $n\mu$ and variance $n\sigma^2$.

Example 8.25 Consider a communication system that transmits a data packet of 1024 bits. Each bit can be in error with probability of 10^{-2}. Find the (approximate) probability that more than 30 of the 1024 bits are in error.

Define a random variable x_i such that $x_i = 1$ if the ith bit is in error and $x_i = 0$ if not. Hence

$$v = \sum_{i=1}^{1024} x_i$$

is the number of errors in the data packet. We would like to find $P(v > 30)$.

Since $P(x_i = 1) = 10^{-2}$ and $P(x_i = 0) = 1 - 10^{-2}$, strictly speaking we would need to find

$$P(v > 30) = \sum_{m=31}^{1024} \binom{1024}{m} \left(10^{-2}\right)^m \left(1 - 10^{-2}\right)^{1024-m}$$

This calculation is time-consuming. We now apply the central limit theorem to solve this problem approximately.

First, we find

$$\overline{x_i} = 10^{-2} \times (1) + (1 - 10^{-2}) \times (0) = 10^{-2}$$

$$\overline{x_i^2} = 10^{-2} \times (1)^2 + (1 - 10^{-2}) \times (0) = 10^{-2}$$

As a result,

$$\sigma_i^2 = \overline{x_i^2} - (\overline{x_i})^2 = 0.0099$$

Based on the central limit theorem, $v = \sum_i x_i$ is approximately Gaussian with mean of $1024 \cdot 10^{-2} = 10.24$ and variance $1024 \times 0.0099 = 10.1376$. Since

$$y = \frac{v - 10.24}{\sqrt{10.1376}}$$

is a standard Gaussian with zero mean and unit variance,

$$\begin{aligned} P(v > 30) &= P\left(y > \frac{30 - 10.24}{\sqrt{10.1376}}\right) \\ &= P(y > 6.20611) \\ &= Q(6.20611) \\ &\simeq 1.925 \times 10^{-10} \end{aligned}$$

Now is a good time to further relax the conditions in the central limit theorem for the sample mean. This highly important generalization is proved by the famous Russian mathematician A. Lyapunov in 1901.

Central Limit Theorem (for the sum of independent random variables): *Let random variables* $\mathrm{x}_1, \ldots, \mathrm{x}_n$ *be independent but not necessarily identically distributed. Each of the random variable* x_i *has mean* μ_i *and nonzero variance* $\sigma_i^2 < \infty$. *Furthermore, suppose that each third-order central moment*

$$\overline{|\mathrm{x}_i - \mu_i|^3} < \infty, \qquad i = 1, \ldots, n$$

and suppose

$$\lim_{n\to\infty} \sum_{i=1}^{n} \overline{|\mathrm{x}_i - \mu_i|^3} \left(\sum_{i=1}^{n} \sigma_i^2\right)^{3/2} = 0$$

Then random variable

$$\mathrm{y}(n) = \frac{\sum_{i=1}^{n} \mathrm{x}_i - \sum_{i=1}^{n} \mu_i}{\sqrt{\sum_{i=1}^{n} \sigma_i^2}}$$

converges to a standard Gaussian density as $n \to \infty$, *that is,*

$$\lim_{n\to\infty} P\left[\mathrm{y}(n) > x\right] = Q(x) \tag{8.99}$$

The central limit theorem provides a plausible explanation for the well-known fact that many random variables in practical experiments are approximately Gaussian. For example, communication channel noise is the sum effect of many different random disturbance sources (e.g., sparks, lightning, static electricity). Based on the central limit theorem, noise as the sum of all these random disturbances should be approximately Gaussian.

REFERENCES

1. J. Singh, *Great Ideas of Modern Mathematics*, Dover, Boston, 1959.
2. M. Abramowitz and I. A. Stegun, Eds., *Handbook of Mathematical Functions*, National Bureau of Standards, Washington, DC, 1964, sec. 26.
3. The Chemical Rubber Co., *CRC Standard Mathematical Tables*, 26th ed., 1980.
4. J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering*, Wiley, New York, 1965, p. 83.
5. N. S. Jayant and P. Noll, *Digital Coding of Waveforms: Principles and Applications to Speech and Video*, Prentice-Hall, 1984.
6. A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed., McGraw-Hill, New York, 1995.
7. M. H. DeGroot, *Probabilities and Statistics*, 2nd ed., Addison Wesley, Reading, MA, 1987.

PROBLEMS

8.1-1 A card is drawn randomly from a regular deck of cards. Assign probability to the event that the card drawn is: **(a)** a red card; **(b)** a black queen; **(c)** a picture card (count an ace as a picture card); **(d)** a number card with number 7; **(e)** a number card with number ≤ 5.

8.1-2 Three regular dice are thrown. Assign probabilities to the following events: the sum of the points appearing on the three dice is **(a)** 4; **(b)** 9; **(c)** 15.

8.1-3 The probability that the number i appears on a throw of a certain loaded dice is k_i ($i = 1, 2, \ldots, 6$). Assign probabilities to all six outcomes.

8.1-4 A bin contains three oscillator microchips, marked O_1, O_2, and O_3, and two PLL microchips, marked P_1 and P_2. Two chips are picked randomly in succession without replacement.

(a) How many outcomes are possible (i.e., how many points are in the sample space)? List all the outcomes and assign probabilities to each of them.

(b) Express the following events as unions of the outcomes in part **(a)**: **(i)** one chip drawn is marked oscillator and the other PLL **(ii)** both chips are PLL; **(iii)** both chips are oscillators; and **(iv)** both chips are of the same kind. Assign probabilities to each of these events.

8.1-5 Use Eq. (8.12) to find the probabilities in Prob. 8.1-4, part **(b)**.

8.1-6 In Prob. 8.1-4, determine the probability that:

(a) The second pick is an oscillator chip given that the first pick is a PLL chip.

(b) The second pick is an oscillator chip given that the first pick is also an oscillator chip.

8.1-7 A binary source generates digits **1** and **0** randomly with equal probability. Assign probabilities to the following events with respect to 10 digits generated by the source: **(a)** there are exactly two **1**s and eight **0**s; **(b)** there are at least four **0**s.

8.1-8 In the California lottery (Lotto), a player chooses any 6 numbers out of 49 numbers (1 through 49). Six balls are drawn randomly (without replacement) from the 49 balls numbered 1 through 49.

(a) Find the probability of matching all 6 balls to the 6 numbers chosen by the player.

(b) Find the probability of matching exactly 5 balls.

(c) Find the probability of matching exactly 4 balls.

(d) Find the probability of matching exactly 3 balls.

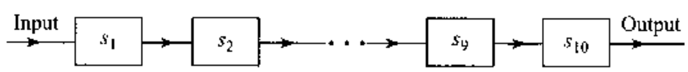
8.1-9 A network consists of 10 links $s_1, s_2, \ldots, s_{10}$ in cascade (Fig. P8.1-9). If any one of the links fails, the entire system fails. All links are independent, with equal probability of failure $p = 0.01$.

(a) What is the probability of failure of the network?

Hint: Consider the probability that none of the links fails.

(b) The reliability of a network is the probability of not failing. If the system reliability is required to be 0.99, what must be the failure probability of each link?

Figure P.8.1-9

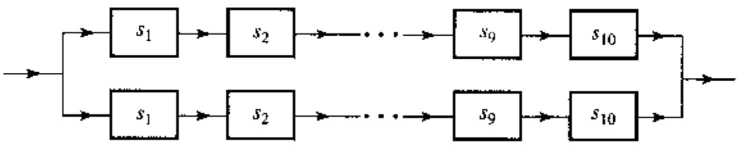


8.1-10 Network reliability improves when redundant links are used. The reliability of the network in Prob. 8.1-9 (Fig. P8.1-9) can be improved by building two subnetworks in parallel (Fig. P8.1-9). Thus, if one subnetwork fails, the other one will still connect.

(a) Using the data in Prob. 8.1-9, determine the reliability of the network in Fig. P8.1-10.

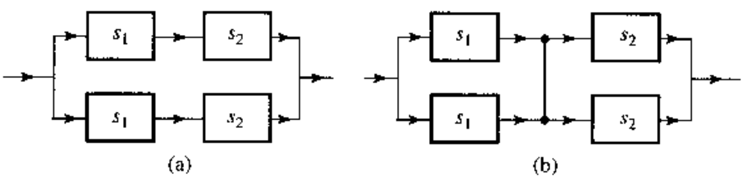
(b) If the reliability of this new network is required to be 0.999, what must be the failure probability of each link?

Figure P.8.1-10



8.1-11 Compare the reliability of the two networks in Fig. P8.1-11, given that the failure probability of links s_1 and s_2 is p each.

Figure P.8.1-11



8.1-12 In a poker game each player is dealt five cards from a regular deck of 52 cards. What is the probability that a player will get a flush (all five cards of the same suit)?

8.1-13 Two dice are thrown. One die is regular and the other is biased with the following probabilities:

$$P(1) = P(6) = \frac{1}{6}, \qquad P(2) = P(4) = 0, \qquad P(3) = P(5) = \frac{1}{3}$$

Determine the probabilities of obtaining a sum: **(a)** 4; **(b)** 5.

8.1-14 In Sec. 8.1, Example 8.5, determine:

(a) $P(B)$, the probability of drawing an ace in the second draw.

(b) $P(A|B)$, the probability that the first draw was a red ace given that the second draw is an ace.

Hint: Event B can occur in two ways: the first draw is a red ace and the second draw is an ace, or the first draw is not a red ace and the second draw is an ace. This is $A \cap B \cup A^c B$ (see Fig. 8.2).

8.1-15 A binary source generates digits **1** and **0** randomly with probabilities $P(1) = 0.8$ and $P(0) = 0.2$.

(a) What is the probability that exactly two **1**s will occur in a n-digit sequence?

(b) What is the probability that at least three **1**s will occur in a n-digit sequence?

8.1-16 In a binary communication channel, the receiver detects binary pulses with an error probability P_e. What is the probability that out of 100 received digits, no more than four digits are in error?

8.1-17 A PCM channel consists of 10 links, with a regenerative repeater at the end of each link. If the detection error probabilities of the 15 detectors are $p_1, p_2, \ldots, p_{15}$, determine the detection error probability of the entire channel if $p_i \ll 1$.

8.1-18 Example 8.8 considers the possibility of improving reliability by repeating a digit three times. Repeat this analysis for five repetitions.

8.1-19 A box contains nine bad microchips. A good microchip is thrown into the box by mistake. Someone is trying to retrieve the good chip. He draws a chip randomly and tests it. If the chip is bad, he throws it out and draws another chip randomly, repeating the procedure until he finds the good chip.

(a) What is the probability that he will find the good chip in the first trial?

(b) What is the probability that he will find the good chip in five trials?

8.1-20 One out of a group of 10 people is to be selected for a suicide mission by drawing straws. There are 10 straws: nine are of the same length and the tenth is shorter than the others. Each of the 10 people draws a straw, one by one. The person who draws the short straw is selected for the mission. Determine which position in the sequence favors the most and which favors the least drawing the short straw.

8.2-1 For a certain binary nonsymmetric channel it is given that

$$P_{\mathrm{y|x}}(\mathbf{0}|\mathbf{1}) = 0.1 \qquad \text{and} \qquad P_{\mathrm{y|x}}(\mathbf{1}|\mathbf{0}) = 0.2$$

where x is the transmitted digit and y is the received digit. If $P_{\mathrm{x}}(\mathbf{0}) = 0.4$, determine $P_{\mathrm{y}}(\mathbf{0})$ and $P_{\mathrm{y}}(\mathbf{1})$.

8.2-2 A binary symmetric channel (see Example 8.13) has an error probability P_e. The probability of transmitting **1** is Q. If the receiver detects an incoming digit as **1**, what is the probability that the originally transmitted digit was: **(a) 1**; **(b) 0**?

Hint: If x is the transmitted digit and y is the received digit, you are given $P_{\mathrm{y|x}}(\mathbf{0}|\mathbf{1}) = P_{\mathrm{y|x}}(\mathbf{1}|\mathbf{0}) = P_e$. Now using Bayes' rule, find $P_{\mathrm{x|y}}(\mathbf{1}|\mathbf{1})$ and $P_{\mathrm{x|y}}(\mathbf{0}|\mathbf{1})$.

8.2-3 The PDF of amplitude x of a certain signal $x(t)$ is given by $p_{\mathrm{x}}(x) = 0.5|x|e^{-|x|}$.

(a) Find the probability that $\mathrm{x} \geq 1$.

(b) Find the probability that $-1 < \mathrm{x} \leq 2$.

(c) Find the probability that $\mathrm{x} \leq -2$.

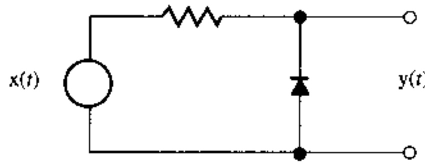
8.2-4 The PDF of an amplitude x of a Gaussian signal x(t) is given by

$$p_{\mathrm{x}}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2}$$

This signal is applied to the input of a half-wave rectifier circuit (Fig. P8.2-4).

Assuming an ideal diode, determine $F_{\mathrm{y}}(y)$ and $p_{\mathrm{y}}(y)$ of the output signal amplitude $\mathrm{y} = \mathrm{x} \cdot u(\mathrm{x})$. Notice that the probability of x= 0 is not zero.

Figure P.8.2-4



8.2-5 The PDF of a Gaussian variable x is given by

$$p_x(x) = \frac{1}{3\sqrt{2\pi}} e^{-(x-4)^2/18}$$

Determine: **(a)** $P(x \geq 4)$; **(b)** $P(x \leq 0)$; **(c)** $P(x \geq -2)$.

8.2-6 For an RV x with PDF

$$p_x(x) = \frac{1}{2\sqrt{2\pi}} e^{-x^2/32} u(x)$$

(a) Sketch $p_x(x)$, and state (with reasons) if this is a Gaussian RV.

(b) Determine: **(i)** $P(x \geq 1)$, **(ii)** $P(1 < x \leq 2)$.

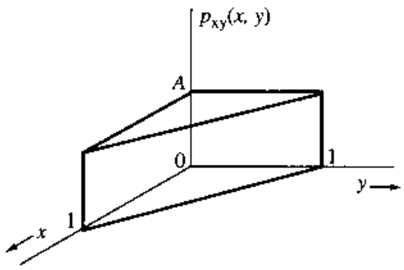
(c) How to generate RV x from another Gaussian RV? Show block diagram and explain.

8.2-7 The joint PDF of RVs x and y is shown in Fig. P8.2-7.

(a) Determine: **(i)** A; **(ii)** $p_x(x)$; **(iii)** $p_y(y)$; **(iv)** $P_{x|y}(x|y)$; **(v)** $P_{y|x}(y|x)$.

(b) Are x and y independent? Explain.

Figure P.8.2-7



8.2-8 The joint PDF $p_{xy}(x, y)$ of two continuous RVs is given by

$$p_{xy}(x, y) = xy\, e^{-(x^2)} e^{-y^2/2} u(x)u(y)$$

(a) Find $p_x(x)$, $p_y(y)$, $p_{x|y}(x|y)$, and $p_{y|x}(y|x)$.

(b) Are x and y independent?

8.2-9 RVs x and y are said to be jointly Gaussian if their joint PDF is given by

$$p_{xy}(x, y) = \frac{1}{2\pi\sqrt{M}} e^{-(ax^2+by^2-2cxy)/2M}$$

where $M = ab - c^2$. Show that $p_x(x)$, $p_y(y)$, $p_{x|y}(x|y)$, and $p_{y|x}(y|x)$ are all Gaussian and that $\overline{x^2} = b$, $\overline{y^2} = a$, and $\overline{xy} = c$.

Hint: Use

$$\int_{-\infty}^{\infty} e^{-px^2+qx}\,dx = \sqrt{\frac{\pi}{p}}e^{q^2/4p}$$

8.2-10 The joint PDF of RVs x and y is given by

$$p_{xy}(x, y) = ke^{-(x^2+xy+y^2)}$$

Determine: **(a)** the constant k; **(b)** $p_x(x)$; **(c)** $p_y(y)$; **(d)** $p_{x|y}(x, y)$; **(e)** $p_{y|x}(y|x)$. Are x and y independent?

8.2-11 In the example on threshold detection (Example 8.16), it was assumed that the digits **1** and **0** were transmitted with equal probability. If $P_x(\mathbf{1})$ and $P_x(\mathbf{0})$, the probabilities of transmitting **1** and **0**, respectively, are not equal, show that the optimum threshold is not 0 but is a, where

$$a = \frac{\sigma_n^2}{2A_p}\ln\frac{P_x(\mathbf{0})}{P_x(\mathbf{1})}$$

Hint: Assume that the optimum threshold is a, and write P_e in terms of the Q functions. For the optimum case, $dP_e/da = 0$. Use the fact that

$$Q(x) = 1 - \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-y^2/2}\,dy$$

and

$$\frac{dQ(x)}{dx} = -\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$$

8.3-1 If an amplitude x of a Gaussian signal x(t) has a mean value of 2 and an RMS value of 3, determine its PDF.

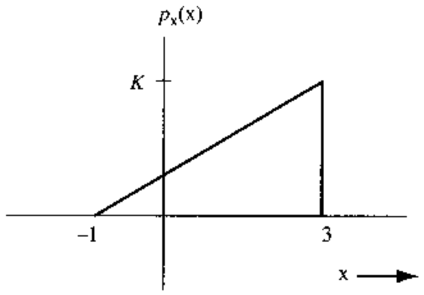
8.3-2 Determine the mean, the mean square, and the variance of the RV x in Prob. 8.2-3.

8.3-3 Determine the mean and the mean square value of RV x in Prob. 8.2-4.

8.3-4 Determine the mean and the mean square value of RV x in Prob. 8.2-6.

8.3-5 Find the mean, the mean square, and the variance of the RV x in Fig. P8.3-5.

Figure P.8.3-5



8.3-6 The sum of points on two tossed dice is a discrete RV x, as analyzed in Example 8.12. Determine the mean, the mean square, and the variance of the RV x.

8.3-7 For a Gaussian PDF $p_x(x) = (1/\sigma_x\sqrt{2\pi})e^{-x^2/2\sigma_x^2}$, show that

$$\overline{x^n} = \begin{cases} (1)(3)(5)\cdots(n-1)\sigma_x^n & n \text{ even} \\ 0 & n \text{ odd} \end{cases}$$

Hint: See appropriate definite integrals in any standard mathematical table.

8.3-8 Ten regular dice are thrown. The sum of the numbers appearing on these 10 dice is an RV x. Find $\bar{x}$, $\overline{x^2}$, and σ_x^2.

Hint: Remember that the outcome of each die is independent.

8.5-1 Show that $|\rho_{xy}| \leq 1$, where ρ_{xy} is the correlation coefficient [Eq. (8.79)] of RVs x and y.

Hint: For any real number a,

$$\overline{[a(x-\bar{x})-(y-\bar{y})]^2} \geq 0$$

The discriminant of this quadratic in a is nonpositive.

8.5-2 Show that if two RVs x and y are related by

$$y = k_1 x + k_2$$

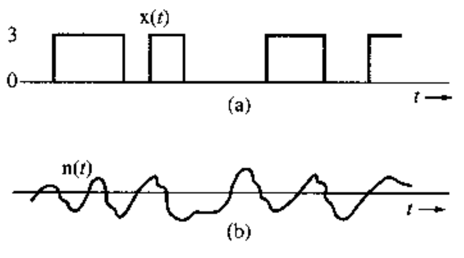
where k_1 and k_2 are arbitrary constants, the correlation coefficient $\rho_{xy} = 1$ if k_1 is positive, and $\rho_{xy} = -1$ if k_1 is negative.

8.5-3 Given $x = \cos\Theta$ and $y = \sin\Theta$, where Θ is an RV uniformly distributed in the range $(0, 2\pi)$, show that x and y are uncorrelated but are not independent.

8.6-1 The random binary signal x(t), shown in Fig. P8.6-1a, can take on only two values, 3 and 0, with equal probability. An exponential channel noise n(t) shown in Fig. P8.6-1b is added to this signal, giving the received signal y(t). The PDF of the noise amplitude n is exponential with a zero mean and a variance of 2. Determine and sketch the PDF of the amplitude y.

Hint: Use of Eq. (8.92) yields $p_y(y) = p_x(x) * p_n(n)$.

Figure P.8.6-1



8.6-2 Repeat Prob. 8.6-1 if the amplitudes 3 and 0 of x(t) are not equiprobable but $P_x(3) = 0.6$ and $P_x(0) = 0.4$.

8.6-3 If $x(t)$ and $y(t)$ are both independent binary signals, each taking on values −1 and 1 only, with

$$P_x(1) = Q = 1 - P_x(-1)$$
$$P_y(1) = P = 1 - P_y(-1)$$

determine $P_z(z_i)$ where $z = x + y$.

8.6-4 If $z = x + y$, where x and y are independent Gaussian RVs with

$$p_x(x) = \frac{1}{\sigma_x\sqrt{2\pi}} e^{-(x-\bar{x})^2/2\sigma_x^2} \quad \text{and} \quad p_y(y) = \frac{1}{\sigma_y\sqrt{2\pi}} e^{-(y-\bar{y})^2/2\sigma_y^2}$$

then show that z is also Gaussian with

$$\bar{z} = \bar{x} + \bar{y} \qquad \textit{and} \qquad \sigma_z^2 = \sigma_x^2 + \sigma_y^2$$

Hint: Convolve $p_x(x)$ and $p_y(y)$. See pair 22 in Table 3.1.

8.6-5 In Example 8.24, design the optimum third-order predictor processor for speech signals and determine the SNR improvement. Values of various correlation coefficients for speech signals are given in Example 8.24.

9 RANDOM PROCESSES AND SPECTRAL ANALYSIS

The notion of a random process is a natural extension of the random variable (RV). Consider, for example, the temperature x of a certain city at noon. The temperature x is an RV and takes on different values every day. To get the complete statistics of x, we need to record values of x at noon over many days (a large number of trials). From this data, we can determine $p_x(x)$, the PDF of the RV x (the temperature at noon).

But the temperature is also a function of time. At 1 p.m., for example, the temperature may have an entirely different distribution from that of the temperature at noon. Still, the two temperatures may be related, via a joint probability density function. Thus, this random temperature x is a function of time and can be expressed as $x(t)$. If the random variable is defined for a time interval $t \in [t_a, t_b]$, then $x(t)$ is a function of time and is random for every instant $t \in [t_a, t_b]$. An RV that is a function of time* is called a **random process**, or **stochastic process**. Thus, a random process is a collection of an infinite number of RVs. Communication signals as well as noises, typically random and varying with time, are well characterized by random processes. For this reason, random process is the subject of this chapter before we study the performance analysis of different communication systems.

9.1 FROM RANDOM VARIABLE TO RANDOM PROCESS

To specify an RV x, we run multiple trials of the experiment and from the outcomes estimate $p_x(x)$. Similarly, to specify the random process $x(t)$, we do the same thing for each time instant t. To continue with our example of the random process $x(t)$, the temperature of the city, we need to record daily temperatures for each value of t (for each time of the day). This can be done by recording temperatures at every instant of the day, which gives a waveform $x(t, \zeta_i)$, where ζ_i indicates the day for which the record was taken. We need to repeat this procedure every day for a large number of days. The collection of all possible waveforms is known as the **ensemble** (corresponding to the sample space) of the random process $x(t)$. A waveform in this collection is a **sample function** (rather than a sample point) of the random process (Fig. 9.1).

* Actually, to qualify as a random process, x could be a function of any practical variable, such as distance. In fact, a random process may also be a function of more than one variable.

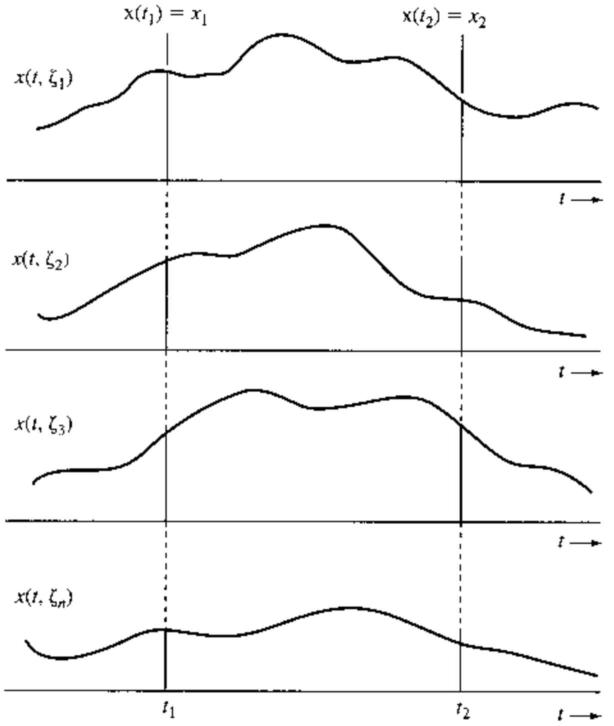


Figure 9.1 Random process to represent the temperature of a city.

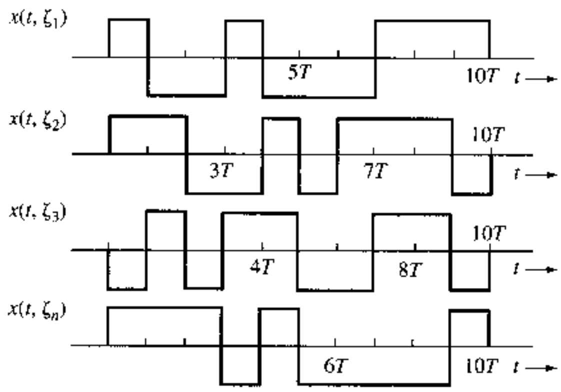


Figure 9.2 Ensemble with a finite number of sample functions.

Sample function amplitudes at some instant $t = t_1$ are the values taken by the RV $x(t_1)$ in various trials.

We can view a random process in another way. In the case of an RV, the outcome of each trial of the experiment is a number. We can view a random process also as the outcome of an experiment, where the outcome of each trial is a waveform (a sample function) that is a function of t. The number of waveforms in an ensemble may be finite or infinite. In the case of the random process $x(t)$ (the temperature of a city), the ensemble has infinitely many waveforms. On the other hand, if we consider the output of a binary signal generator (over the period 0 to $10T$), there are at most 2^{10} waveforms in this ensemble (Fig. 9.2).

One fine point that needs clarification is that the waveforms (or sample functions) in the ensemble are not random. They have occurred and are therefore deterministic. Randomness in this situation is associated not with the waveform but with the uncertainty as to which waveform would occur in a given trial. This is completely analogous to the situation of an RV. For example, in the experiment of tossing a coin four times in succession (Example 8.4), 16 possible outcomes exist, all of which are known. The randomness in this situation is associated not with the outcomes but with the uncertainty as to which of the 16 outcomes will occur in a given trial. Indeed, the random process is basically an infinite long vector of random variables. Once an experiment is completed, the sampled vector is deterministic. However, since each element in the vector is random, the experimental outcome is also random, leading to uncertainty over what vector (or function) will be generated in each experiment.

Characterization of a Random Process

The next important question is how to characterize (describe) a random process. In some cases, we may be able to describe it analytically. Consider, for instance, a random process described by $\mathrm{x}(t) = A\cos(\omega_c t + \Theta)$, where Θ is an RV uniformly distributed over the range $(0, 2\pi)$. This analytical expression completely describes a random process (and its ensemble). Each sample function is a sinusoid of amplitude A and frequency ω_c. But the phase is random (see later, Fig. 9.5). It is equally likely to take any value in the range $(0, 2\pi)$. Such an analytical description requires well-defined models such that the random process is characterized by specific parameters that are random variables.

Unfortunately, it is not always possible to be able to describe a random process analytically. Without a specific model, we may have just an ensemble obtained experimentally. The ensemble has the complete information about the random process. From this ensemble, we must find some quantitative measure that will specify or characterize the random process. In this case, we consider the random process as an RV x that is a function of time. Thus, a random process is just a collection of an infinite number of RVs, which are generally dependent. We know that the complete information of several dependent RVs is provided by the joint PDF of those variables. Let x_i represent the RV $\mathrm{x}(t_i)$ generated by the amplitudes of the random process at instant $t = t_i$. Thus, x_1 is the RV generated by the amplitudes at $t = t_1$, and x_2 is the RV generated by the amplitudes at $t = t_2$, and so on, as shown in Fig. 9.1. The n RVs $\mathrm{x}_1, \mathrm{x}_2, \mathrm{x}_3, \ldots, \mathrm{x}_n$ generated by the amplitudes at $t = t_1, t_2, t_3, \ldots, t_n$, respectively, are dependent in general. For the n samples, they are fully characterized by the n th-order joint probability density function or the n th-order joint cumulative distribution function (CDF)

$$F_{\mathrm{x}}(x_1, x_2, \ldots, x_n; t_1, t_2, \ldots, t_n) = P[\mathrm{x}(t) \le x_1; \mathrm{x}(t) \le x_2; \ldots; \mathrm{x}(t_n) \le x_n]$$

The definition of the joint CDF of the n random samples leads to the joint PDF

$$p_{\mathrm{x}}(x_1, x_2, \ldots, x_n; t_1, t_2, \ldots, t_n) = \frac{\partial^n}{\partial x_1 \partial x_2 \ldots \partial x_n} F_{\mathrm{x}}(x_1, x_2, \ldots, x_n; t_1, t_2, \ldots, t_n) \tag{9.1}$$

This discussion provides some good insight. It can be shown that the random process is completely described by the n th-order joint PDF (9.1) for all n (up to ∞) and for any choice of $t_1, t_2, t_3, \ldots, t_n$. Determining this PDF (of infinite order) is a formidable task. Fortunately, we shall soon see that when analyzing random signals and noises in conjunction with linear systems, we are often content with the specifications of the first- and second-order statistics.

A higher order PDF is the joint PDF of the random process at multiple time instants. Hence, we can always derive a lower order PDF from a higher order PDF by simple integration. For instance,

$$p_{\mathrm{x}}(x_1; t_1) = \int_{-\infty}^{\infty} p_{\mathrm{x}}(x_1, x_2; t_1, t_2)\, dx_2$$

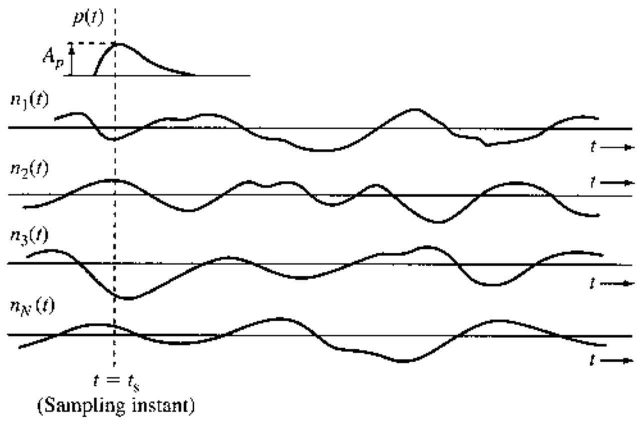


Figure 9.3 A random process to represent a channel noise.

Hence, when the nth-order PDF is available, there is no need to specify PDFs of order lower than n.

The mean $\overline{\mathrm{x}(t)}$ of a random process x(t) can be determined from the first-order PDF as

$$\overline{\mathrm{x}(t)} = \int_{-\infty}^{\infty} x\, p_{\mathrm{x}}(x;\ t)\, dx \tag{9.2}$$

which is typically a deterministic function of time t.

Why Do We Need Ensemble Statistics?

The preceding discussion shows that to specify a random process, we need ensemble statistics. For instance, to determine the PDF $p_{\mathrm{x}_1}(x_1)$, we need to find the values of all the sample functions at $t = t_1$. This is ensemble statistics. In the same way, the inclusion of all possible statistics in the specification of a random process necessitates some kind of ensemble statistics. In deterministic signals, we are used to studying the data of a waveform (or waveforms) as a function of time. Hence, the idea of investigating ensemble statistics makes us feel a bit uncomfortable at first. Theoretically, we may accept it, but does it have any practical significance? How is this concept useful in practice? We shall now answer this question.

To understand the necessity of ensemble statistics, consider the problem of threshold detection in Example 8.16. A **1** is transmitted by $p(t)$ and a **0** is transmitted by $-p(t)$ (polar signaling). The peak pulse amplitude is A_p. When **1** is transmitted, the received sample value is $A_p + \mathrm{n}$, where n is the noise. We would make a decision error if the noise value at the sampling instant t_s were less than $-A_p$, forcing the sum of signal and noise to fall below the threshold. To find this error probability, we repeat the experiment N times ($N \to \infty$) and see how many times the noise at $t = t_s$ is less than $-A_p$ (Fig. 9.3). This information is precisely one of ensemble statistics of the noise process $n(t)$ at instant t_s.

The importance of ensemble statistics is clear from this example. When we are dealing with a random process or processes, we do not know which sample function will occur in a given trial. Hence, for any statistical specification and characterization of the random process, we need to average over the entire ensemble. This is the basic physical reason for the appearance of ensemble statistics in random processes.

Autocorrelation Function of a Random Process

For the purpose of signal analysis, one of the most important (statistical) characteristics of a random process is its **autocorrelation function**, which leads to the spectral information of the

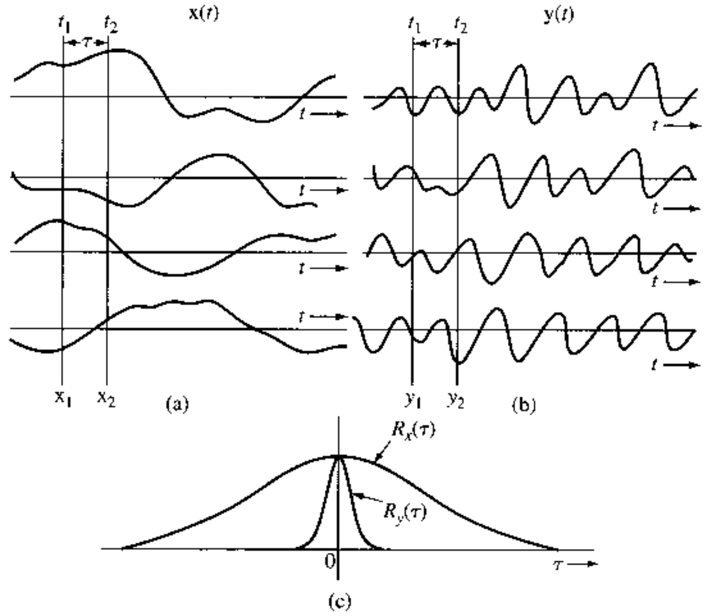


Figure 9.4 Autocorrelation functions for a slowly varying and a rapidly varying random process.

random process. The spectral content of a process depends on the rapidity of the amplitude change with time. This can be measured by correlating amplitudes at t_1 and $t_1 + \tau$. On average, the random process x(t) in Fig. 9.4a is a slowly varying process in comparison to the process y(t) in Fig. 9.4b. For x(t), the amplitudes at t_1 and $t_1 + \tau$ are similar (Fig. 9.4a), that is, have stronger correlation. On the other hand, for y(t), the amplitudes at t_1 and $t_1 + \tau$ have little resemblance (Fig. 9.4b), that is, have weaker correlation. Recall that correlation is a measure of the similarity of two RVs. Hence, we can use correlation to measure the similarity of amplitudes at t_1 and $t_2 = t_1 + \tau$. If the RVs x(t_1) and x(t_2) are denoted by x_1 and x_2, respectively, then for a real random process,* the autocorrelation function $R_x(t_1, t_2)$ is defined as

$$R_x(t_1, t_2) = \overline{x(t_1)x(t_2)} = \overline{x_1 x_2} \tag{9.3a}$$

This is the correlation of RVs x(t_1) and x(t_2), indicating the similarity between RVs x(t_1) and x(t_2). It is computed by multiplying amplitudes at t_1 and t_2 of a sample function and then averaging this product over the ensemble. It can be seen that for a small τ, the product $x_1 x_2$ will be positive for most sample functions of x(t), but the product $y_1 y_2$ is equally likely to be positive or negative. Hence, $\overline{x_1 x_2}$ will be larger than $\overline{y_1 y_2}$. Moreover, x_1 and x_2 will show correlation for considerably larger values of τ, whereas y_1 and y_2 will lose correlation quickly, even for small τ, as shown in Fig. 9.4c. Thus, $R_x(t_1, t_2)$, the autocorrelation function of x(t), provides valuable information about the frequency content of the process. In fact, we shall show that the PSD of x(t) is the Fourier transform of its autocorrelation function, given by (for real processes)

$$\begin{aligned} R_x(t_1, t_2) &= \overline{x_1 x_2} \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2 p_x(x_1, x_2; t_1, t_2)\, dx_1\, dx_2 \end{aligned} \tag{9.3b}$$

* For a complex random process x(t), the autocorrelation function is defined as

$$R_x(t_1, t_2) = \overline{x^*(t_1)x(t_2)}$$

Hence, $R_x(t_1, t_2)$ can be derived from the joint PDF of x_1 and x_2, which is the second-order PDF.

9.2 CLASSIFICATION OF RANDOM PROCESSES

Random processes may be classified into the following broad categories.

Stationary and Nonstationary Random Processes

A random process whose statistical characteristics do not change with time is classified as a **stationary random process.** For a stationary process, we can say that a shift of time origin will be impossible to detect; the process will appear to be the same. Suppose we determine $p_x(x;\ t_1)$, then shift the origin by t_0, and again determine $p_x(x;\ t_1)$. The instant t_1 in the new frame of reference is $t_2 = t_1 + t_0$ in the old frame of reference. Hence, the PDFs of x at t_1 and $t_2 = t_1 + t_0$ must be the same; that is, $p_x(x;\ t_1)$ and $p_x(x;\ t_2)$ must be identical for a stationary random process. This is possible only if $p_x(x;\ t)$ is independent of t. Thus, the first-order density of a stationary random process can be expressed as

$$p_x(x;\ t) = p_x(x)$$

Similarly, for a stationary random process the autocorrelation function $R_x(t_1, t_2)$ must depend on t_1 and t_2 only through the difference $t_2 - t_1$. If not, we could determine a unique time origin. Hence, for a real stationary process,

$$R_x(t_1, t_2) = R_x(t_2 - t_1)$$

Therefore,

$$R_x(\tau) = \overline{x(t)x(t + \tau)} \tag{9.4}$$

For a stationary process, the joint PDF for x_1 and x_2 must also depend only on $t_2 - t_1$. Similarly, higher order PDFs are all independent of the choice of origin, that is,

$$\begin{aligned} p_x(x_1, x_2, \ldots, x_n;\ t_1, t_2, \ldots, t_n) &= p_x(x_1, x_2, \ldots, x_n;\ t_1 - t, t_2 - t, \ldots, t_n - t) \qquad \forall t \\ &= p_x(x_1, x_2, \ldots, x_n;\ 0, t_2 - t_1, \ldots t_n - t_1) \end{aligned} \tag{9.5}$$

The random process $x(t)$ representing the temperature of a city is an example of a nonstationary random process because the temperature statistics (mean value, for example) depend on the time of the day. On the other hand, we can say that the noise process in Fig. 9.3 is stationary because its statistics (the mean and the mean square values, for example) do not change with time. In general, it is not easy to determine whether or not a process is stationary because the nth-order ($n = 1, 2, \ldots, \infty$) statistics must be investigated. In practice, we can ascertain stationarity if there is no change in the signal-generating mechanism. Such is the case for the noise process in Fig. 9.3.

Wide-Sense (or Weakly) Stationary Processes

A process that is not stationary in the strict sense, as discussed in the last subsection, may yet have a mean value and an autocorrelation function that are independent of the shift of time origin. This means

$$\overline{x(t)} = \text{constant}$$

and

$$R_x(t_1, t_2) = R_x(\tau) \qquad \tau = t_2 - t_1 \tag{9.6}$$

Such a process is known as a **wide-sense stationary**, or **weakly stationary**, process. Note that stationarity is a stronger condition than wide-sense stationarity. Stationary processes with well-defined autocorrelation functions are wide-sense stationary; exception for Gaussian random processes, however, the converse is not necessarily true.

Just as no sinusoidal signal exists in actual practice, no truly stationary process can occur in real life. All processes in practice are nonstationary because they must begin at some finite time and terminate at some finite time. A truly stationary process must start at $t = -\infty$ and go on forever. Many processes can be considered stationary for the time interval of interest, however, and the stationarity assumption allows a manageable mathematical model. The use of a stationary model is analogous to the use of a sinusoidal model in deterministic analysis.

Example 9.1 Show that the random process

$$x(t) = A\cos(\omega_c t + \Theta)$$

where Θ is an RV uniformly distributed in the range $(0, 2\pi)$, is a wide-sense stationary process.

The ensemble (Fig. 9.5) consists of sinusoids of constant amplitude A and constant frequency ω_c, but the phase Θ is random. For any sample function, the phase is equally likely to have any value in the range $(0, 2\pi)$. Because Θ is an RV uniformly distributed over the range $(0, 2\pi)$, one can determine[1] $p_x(x, t)$ and, hence, $\overline{x(t)}$, as in Eq. (9.2). For this particular case, however, $\overline{x(t)}$ can be determined directly as a function of random variable Θ:

$$\overline{x(t)} = \overline{A\cos(\omega_c t + \Theta)} = A\,\overline{\cos(\omega_c t + \Theta)}$$

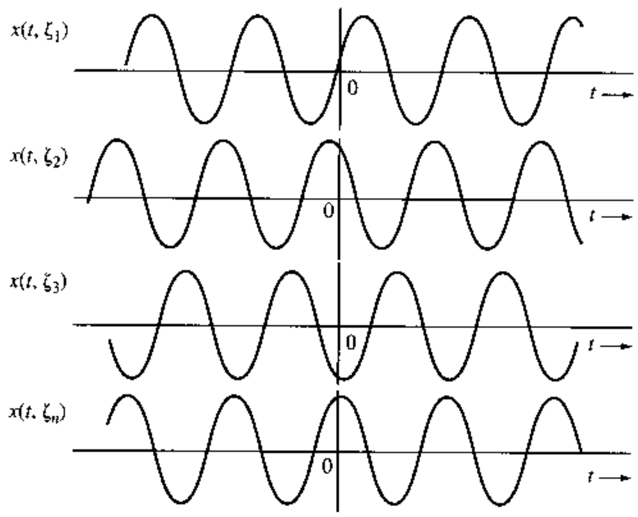


Figure 9.5 Ensemble for the random process $A\cos(\omega_c t + \Theta)$.

Because $\cos(\omega_c t + \Theta)$ is a function of an RV Θ, we have [see Eq. (8.61b)]

$$\overline{\cos(\omega_c t + \Theta)} = \int_0^{2\pi} \cos(\omega_c t + \theta) p_\Theta(\theta)\, d\theta$$

Because $p_\Theta(\theta) = 1/2\pi$ over $(0, 2\pi)$ and 0 outside this range,

$$\overline{\cos(\omega_c t + \Theta)} = \frac{1}{2\pi}\int_0^{2\pi} \cos(\omega_c t + \theta)\, d\theta = 0$$

Hence,

$$\overline{\mathrm{x}(t)} = 0 \tag{9.7a}$$

Thus, the ensemble mean of sample function amplitudes at any instant t is zero.

The autocorrelation function $R_{\mathrm{x}}(t_1, t_2)$ for this process also can be determined directly from Eq. (9.3a),

$$\begin{aligned} R_{\mathrm{x}}(t_1, t_2) &= \overline{A^2 \cos(\omega_c t_1 + \Theta)\cos(\omega_c t_2 + \Theta)} \\ &= A^2\, \overline{\cos(\omega_c t_1 + \Theta)\cos(\omega_c t_2 + \Theta)} \\ &= \frac{A^2}{2}\left\{\overline{\cos[\omega_c(t_2 - t_1)]} + \overline{\cos[\omega_c(t_2 + t_1) + 2\Theta]}\right\} \end{aligned}$$

The first term on the right-hand side contains no RV. Hence, $\overline{\cos[\omega_c(t_2 - t_1)]}$ is $\cos[\omega_c(t_2 - t_1)]$ itself. The second term is a function of the uniform RV Θ, and its mean is

$$\overline{\cos[\omega_c(t_2 + t_1) + 2\Theta]} = \frac{1}{2\pi}\int_0^{2\pi} \cos[\omega_c(t_2 + t_1) + 2\theta]\, d\theta = 0$$

Hence,

$$R_{\mathrm{x}}(t_1, t_2) = \frac{A^2}{2}\cos[\omega_c(t_2 - t_1)] \tag{9.7b}$$

or

$$R_{\mathrm{x}}(\tau) = \frac{A^2}{2}\cos \omega_c \tau \qquad \tau = t_2 - t_1 \tag{9.7c}$$

From Eqs. (9.7a) and (9.7b) it is clear that $\mathrm{x}(t)$ is a wide-sense stationary process.

Ergodic Wide-Sense Stationary Processes

We have studied the mean and the autocorrelation function of a random process. These are ensemble averages. For example, $\overline{\mathrm{x}(t)}$ is the ensemble average of sample function amplitudes at t, and $R_{\mathrm{x}}(t_1, t_2) = \overline{\mathrm{x}_1\mathrm{x}_2}$ is the ensemble average of the product of sample function amplitudes $\mathrm{x}(t_1)$ and $\mathrm{x}(t_2)$.

We can also define time averages for each sample function. For example, a time mean $\widetilde{x(t)}$ of a sample function $x(t)$ is*

$$\widetilde{x(t)} = \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} x(t)\, dt \tag{9.8a}$$

* Here a sample function $x(t, \zeta_i)$ is represented by $x(t)$ for convenience.

Similarly, the time autocorrelation function $\mathcal{R}_x(\tau)$ defined in Eq. (3.82b) is

$$\mathcal{R}_x(\tau) = \widetilde{x(t)x(t+\tau)} = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)x(t+\tau)\,dt \tag{9.8b}$$

For **ergodic (wide-sense) stationary processes**, ensemble averages are equal to the time averages of any sample function. Thus, for an ergodic process x(t),

$$\overline{\mathrm{x}(t)} = \widetilde{x(t)} \tag{9.9a}$$

$$R_x(\tau) = \mathcal{R}_x(\tau) \tag{9.9b}$$

These are the two averages for ergodic wide-sense stationary processes. For the broader definition of an ergodic process, all possible ensemble averages are equal to the corresponding time averages of one of its sample functions. Figure 9.6 illustrates the relationship among different classes of (ergodic) processes. In the coverage of this book, our focus lies in the class of ergodic wide-sense stationary processes.

It is difficult to test whether a process is ergodic or not, because we must test all possible orders of time and ensemble averages. Nevertheless, in practice many of the stationary processes are ergodic with respect to at least low-order statistics, such as the mean and the autocorrelation. For the process in Example 9.1 (Fig. 9.5), we can show that $\widetilde{x(t)} = 0$ and $\mathcal{R}_x(\tau) = (A^2/2)\cos\omega_c\tau$ (see Prob. 3.8-1). Therefore, this process is ergodic at least with respect to the first- and second-order averages.

The ergodicity concept can be explained by a simple example of traffic lights in a city. Suppose the city is well planned, with all its streets in E–W and N–S directions only and with traffic lights at each intersection. Assume that each light stays green for 0.75 second in the E–W direction and 0.25 second in the N–S direction and that switching of any light is independent of the other lights. For the sake of simplicity, we ignore the orange light.

If we consider a certain person driving a car arriving at any traffic light randomly in the E–W direction, the probability that the person will have a green light is 0.75; that is, on the average, 75% of the time the person will observe a green light. On the other hand, if we consider a large number of drivers arriving at a traffic light in the E–W direction at some instant t, then 75% of the drivers will have a green light, and the remaining 25% will have a red light. Thus, the experience of a single driver arriving randomly many times at a traffic light will contain the same statistical information (sample function statistics) as that of a large number of drivers arriving simultaneously at various traffic lights (ensemble statistics) at one instant.

The ergodicity notion is extremely important because we do not have a large number of sample functions available in practice from which to compute ensemble averages. If the process is known to be ergodic, then we need only one sample function to compute ensemble averages. As mentioned earlier, many of the stationary processes encountered in practice are ergodic with

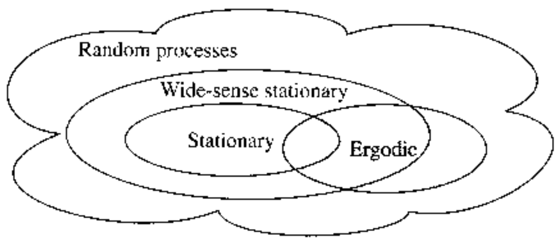


Figure 9.6 Classification of random processes.

respect to at least second-order averages. As we shall see in dealing with stationary processes in conjunction with linear systems, we need only the first- and second-order averages. This means that in most cases we can get by with a single sample function, as is often the case in practice.

9.3 POWER SPECTRAL DENSITY

An electrical engineer instinctively thinks of signals and linear systems in terms of their frequency domain descriptions. Linear systems are characterized by their frequency response (the transfer function), and signals are expressed in terms of the relative amplitudes and phases of their frequency components (the Fourier transform). From a knowledge of the input spectrum and transfer function, the response of a linear system to a given signal can be obtained in terms of the frequency content of that signal. This is an important analytical procedure for deterministic signals. We may wonder if similar methods may be found for random processes. Ideally, all the sample functions of a random process are assumed to exist over the entire time interval $(-\infty, \infty)$ and, thus, are power signals.* We therefore inquire about the existence of a power spectral density (PSD). Superficially, the concept of a random process having a PSD may appear ridiculous for the following reasons. In the first place, we may not be able to describe a sample function analytically. Second, for a given process, every sample function may be different from another one. Hence, even if a PSD does exist for each sample function, it may be different for different sample functions. Fortunately, both problems can be neatly resolved, and it is possible to define a meaningful PSD for a stationary (at least in the wide sense) random process. For nonstationary processes, the PSD may not exist.

Whenever randomness is involved, our inquiries can at best provide answers in terms of averages. When tossing a coin, for instance, the most we can say about the outcome is that on the average we will obtain heads in about half the trials and tails in the remaining half of the trials. For random signals or RVs, we do not have enough information to predict the outcome with certainty, and we must accept answers in terms of averages. It is not possible to transcend this limit of knowledge because of our fundamental ignorance of the process. It seems reasonable to define the PSD of a random process as a weighted mean of the PSDs of all sample functions. This is the only sensible solution, since we do not know exactly which of the sample functions may occur in a given trial. We must be prepared for any sample function. Consider, for example, the problem of filtering a certain random process. We would not want to design a filter with respect to any one particular sample function because any of the sample functions in the ensemble may be present at the input. A sensible approach is to design the filter with respect to the mean parameters of the input process. In designing a system to perform certain operations, one must design it with respect to the whole ensemble. We are therefore justified in defining the PSD $S_{\mathrm{x}}(f)$ of a random process $\mathrm{x}(t)$ as the ensemble average of the PSDs of all sample functions. Thus [see Eq. (3.80)],

$$S_{\mathrm{x}}(f) = \lim_{T\to\infty} \overline{\left[\frac{|\mathrm{X}_T(f)|^2}{T}\right]} \quad \text{W/Hz} \tag{9.10a}$$

where $\mathrm{X}_T(f)$ is the Fourier transform of the time-truncated random process

$$\mathrm{x}_T(t) = \mathrm{x}(t)\,\Pi(t/T)$$

* As we shall soon see, for the PSD to exist, the process must be stationary (at least in the wide sense). Stationary processes, because their statistics do not change with time, are power signals.

and the bar atop represents ensemble average. Note that ensemble averaging is done before the limiting operation. We shall now show that the PSD as defined in Eq. (9.10a) is the Fourier transform of the autocorrelation function $R_{\mathrm{x}}(\tau)$ of the process $\mathrm{x}(t)$; that is,

$$R_{\mathrm{x}}(\tau) \Longleftrightarrow S_{\mathrm{x}}(f) \tag{9.10b}$$

This can be proved as follows:

$$\mathrm{X}_T(f) = \int_{-\infty}^{\infty} \mathrm{x}_T(t)e^{-j2\pi ft}\,dt = \int_{-T/2}^{T/2} \mathrm{x}(t)e^{-j2\pi ft}\,dt \tag{9.11}$$

Thus, for real $\mathrm{x}(t)$,

$$\begin{aligned} |\mathrm{X}_T(f)|^2 &= \mathrm{X}_T(-f)\mathrm{X}_T(f) \\ &= \int_{-T/2}^{T/2} \mathrm{x}(t_1)e^{j2\pi ft_1}\,dt_1 \int_{-T/2}^{T/2} \mathrm{x}(t_2)e^{-j2\pi ft_2}\,dt_2 \\ &= \int_{-T/2}^{T/2}\int_{-T/2}^{T/2} \mathrm{x}(t_1)\mathrm{x}(t_2)e^{-j2\pi f(t_2-t_1)}\,dt_1 dt_2 \end{aligned}$$

and

$$\begin{aligned} S_{\mathrm{x}}(f) &= \lim_{T\to\infty} \overline{\left[\frac{|\mathrm{X}_T(f)|^2}{T}\right]} \\ &= \lim_{T\to\infty} \overline{\left[\frac{1}{T}\int_{-T/2}^{T/2}\int_{-T/2}^{T/2} \mathrm{x}(t_1)\mathrm{x}(t_2)e^{-j2\pi f(t_2-t_1)}\,dt_1\,dt_2\right]} \end{aligned} \tag{9.12}$$

Interchanging the operation of integration and ensemble averaging,* we get

$$\begin{aligned} S_{\mathrm{x}}(f) &= \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2}\int_{-T/2}^{T/2} \overline{\mathrm{x}(t_1)\mathrm{x}(t_2)}e^{-j2\pi f(t_2-t_1)}\,dt_1\,dt_2 \\ &= \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2}\int_{-T/2}^{T/2} R_{\mathrm{x}}(t_2-t_1)e^{-j2\pi f(t_2-t_1)}\,dt_1\,dt_2 \end{aligned}$$

Here we are assuming that the process $\mathrm{x}(t)$ is at least wide-sense stationary, so that $\overline{\mathrm{x}(t_1)\mathrm{x}(t_2)} = R_{\mathrm{x}}(t_2 - t_1)$. For convenience, let

$$R_{\mathrm{x}}(t_2 - t_1)e^{-j2\pi f(t_2-t_1)} = \varphi(t_2 - t_1) \tag{9.13}$$

Then,

$$S_{\mathrm{x}}(f) = \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2}\int_{-T/2}^{T/2} \varphi(t_2 - t_1)\,dt_1\,dt_2 \tag{9.14}$$

* The operation of ensemble averaging is also an operation of integration. Hence, interchanging integration with ensemble averaging is equivalent to interchanging the order of integration.

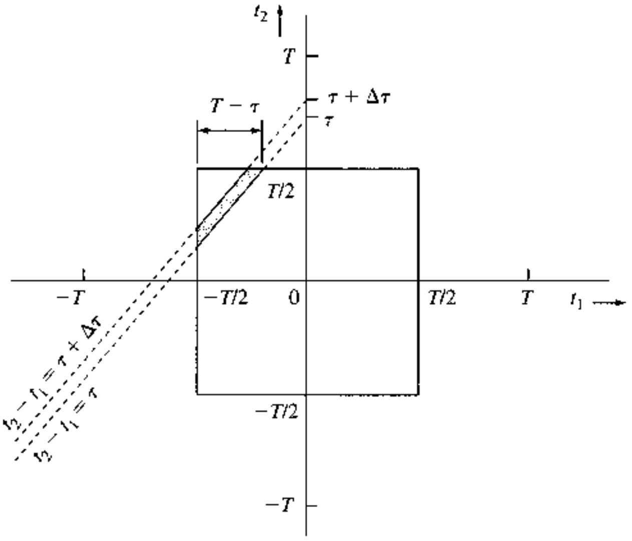


Figure 9.7 Derivation of the Wiener-Khintchine theorem.

The integral on the right-hand side is a double integral over the range $(-T/2,\ T/2)$ for each of the variables t_1 and t_2. The square region of integration in the t_1–t_2 plane is shown in Fig. 9.7. The integral in Eq. (9.14) is a volume under the surface $\varphi(t_2 - t_1)$ over the square region in Fig. 9.7. The double integral in Eq. (9.14) can be converted to a single integral by observing that $\varphi(t_2 - t_1)$ is constant along any line $t_2 - t_1 = \tau$ (a constant) in the t_1–t_2 plane (Fig. 9.7).

Let us consider two such lines, $t_2 - t_1 = \tau$ and $t_2 - t_1 = \tau + \Delta\tau$. If $\Delta\tau \to 0$, $\varphi(t_2 - t_1) \simeq \varphi(\tau)$ over the shaded region whose area is $(T - \tau)\,\Delta\tau$. Hence, the volume under the surface $\varphi(t_2 - t_1)$ over the shaded region is $\varphi(\tau)(T - \tau)\,\Delta\tau$. If τ were negative, the volume would be $\varphi(\tau)(T + \tau)\,\Delta\tau$. Hence, in general, the volume over the shaded region is $\varphi(\tau)(T - |\tau|)\,\Delta\tau$. The desired volume over the square region in Fig. 9.7 is the sum of the volumes over the shaded strips and is obtained by integrating $\varphi(\tau)(T - |\tau|)$ over the range of τ, which is $(-T,\ T)$ (see Fig. 9.7). Hence,

$$\begin{aligned} S_x(f) &= \lim_{T\to\infty} \frac{1}{T} \int_{-T}^{T} \varphi(\tau)(T - |\tau|)\, d\tau \\ &= \lim_{T\to\infty} \int_{-T/2}^{T/2} \varphi(\tau)\left(1 - \frac{|\tau|}{T}\right) d\tau \\ &= \int_{-\infty}^{\infty} \varphi(\tau)\, d\tau \end{aligned}$$

provided $\int_{-\infty}^{\infty} |\tau|\varphi(\tau)\, d\tau$ is bounded. Substituting Eq. (9.13) into this equation, we have

$$S_x(f) = \int_{-\infty}^{\infty} R_x(\tau) e^{-j2\pi f \tau}\, d\tau \tag{9.15}$$

provided $\int_{-\infty}^{\infty} |\tau| R_x(\tau) e^{-j2\pi f \tau}\, d\tau$ is bounded. Thus, the PSD of a wide-sense stationary random process is the Fourier transform of its autocorrelation function,*

$$R_x(\tau) \Longleftrightarrow S_x(f) \tag{9.16}$$

This is the well-known **Wiener-Khintchine theorem**, first presented in Chapter 3.

From the discussion thus far, the autocorrelation function emerges as one of the most significant entities in the spectral analysis of a random process. Earlier we showed heuristically how the autocorrelation function is connected with the frequency content of a random process.

The autocorrelation function $R_x(\tau)$ for real processes is an even function of τ. This can be proved in two ways. First, because $|X_T(f)|^2 = |X_T(f)X_T^*(f)| = |X_T(f)X_T(-f)|$ is an even function of f, $S_x(f)$ is also an even function of f, and $R_x(\tau)$, its inverse transform, is also an even function of τ (see Prob. 3.1-1). Alternately, we may argue that

$$R_x(\tau) = \overline{x(t)x(t+\tau)} \qquad \text{and} \qquad R_x(-\tau) = \overline{x(t)x(t-\tau)}$$

Letting $t - \tau = \sigma$, we have

$$R_x(-\tau) = \overline{x(\sigma)x(\sigma+\tau)} = R_x(\tau) \tag{9.17}$$

The PSD $S_x(f)$ is also a real and even function of f.

The mean square value $\overline{x^2(t)}$ of the random process $x(t)$ is $R_x(0)$,

$$R_x(0) = \overline{x(t)x(t)} = \overline{x^2(t)} = \overline{x^2} \tag{9.18}$$

The mean square value $\overline{x^2}$ is not the time mean square of a sample function but the ensemble average of the squares of all sample function amplitudes at any instant t.

The Power of a Random Process

The power P_x (average power) of a wide-sense random process $x(t)$ is its mean square value $\overline{x^2}$. From Eq. (9.16),

$$R_x(\tau) = \int_{-\infty}^{\infty} S_x(f) e^{j2\pi f \tau}\, df$$

Hence, from Eq. (9.18),

$$P_x = \overline{x^2} = R_x(0) = \int_{-\infty}^{\infty} S_x(f)\, df \tag{9.19a}$$

Because $S_x(f)$ is an even function of f, we have

$$P_x = \overline{x^2} = 2\int_{0}^{\infty} S_x(f)\, df \tag{9.19b}$$

where f is the frequency in hertz. This is the same relationship as that derived for deterministic signals in Chapter 3 [Eq. (3.81)]. The power P_x is the area under the PSD. Also, $P_x = \overline{x^2}$ is the ensemble mean of the square amplitudes of the sample functions at any instant.

---

* It can be shown that Eq. (9.15) holds also for complex random processes, for which we define $R_x(\tau) = \overline{x^*(t)x(t+\tau)}$.

It is helpful to repeat here, once again, that the PSD may not exist for processes that are not wide-sense stationary. Hence, in our future discussion, random processes will be assumed to be at least wide-sense stationary unless specifically stated otherwise.

Example 9.2 Determine the autocorrelation function $R_X(\tau)$ and the power P_X of a low-pass random process with a white noise PSD $S_X(f) = \mathcal{N}/2$ (Fig. 9.8a).

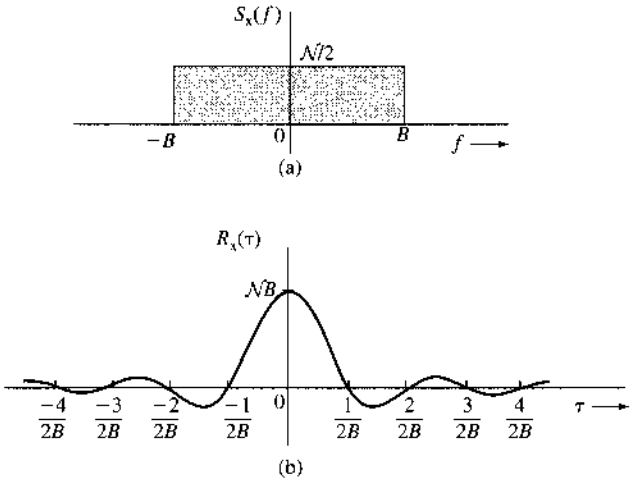


Figure 9.8 Bandpass white noise PSD and its autocorrelation function.

We have

$$S_X(f) = \frac{\mathcal{N}}{2}\Pi\left(\frac{f}{2B}\right) \tag{9.20a}$$

Hence, from Table 3.1 (pair 18),

$$R_X(\tau) = \mathcal{N}B \operatorname{sinc}(2\pi B\tau) \tag{9.20b}$$

This is shown in Fig. 9.8b. Also,

$$P_X = \overline{x^2} = R_X(0) = \mathcal{N}B \tag{9.20c}$$

Alternately,

$$\begin{aligned} P_X &= 2\int_0^\infty S_X(f)\, df \\ &= 2\int_0^B \frac{\mathcal{N}}{2}\, df \\ &= \mathcal{N}B \end{aligned} \tag{9.20d}$$

Example 9.3 Determine the PSD and the mean square value of a random process

$$x(t) = A\cos(\omega_c t + \Theta) \tag{9.21a}$$

where Θ is an RV uniformly distributed over $(0, 2\pi)$.

For this case $R_x(\tau)$ is already determined [Eq. (9.7c)],

$$R_x(\tau) = \frac{A^2}{2}\cos\omega_c\tau \tag{9.21b}$$

Hence,

$$S_x(f) = \frac{A^2}{4}[\delta(f+f_c) + \delta(f-f_c)] \tag{9.21c}$$

$$P_x = \overline{x^2} = R_x(0) = \frac{A^2}{2} \tag{9.21d}$$

Thus, the power, or the mean square value, of the process $x(t) = A\cos(\omega_c t + \Theta)$ is $A^2/2$. The power P_x can also be obtained by integrating $S_x(f)$ with respect to f.

Example 9.4 **Amplitude Modulation**

Determine the autocorrelation function and the PSD of the DSB-SC-modulated process $m(t)\cos(\omega_c t + \Theta)$, where $m(t)$ is a wide-sense stationary random process, and Θ is an RV uniformly distributed over $(0, 2\pi)$ and independent of $m(t)$.

Let

$$\varphi(t) = m(t)\cos(\omega_c t + \Theta)$$

Then

$$R_\varphi(\tau) = \overline{m(t)\cos(\omega_c t + \Theta)\cdot m(t+\tau)\cos[\omega_c(t+\tau) + \Theta]}$$

Because $m(t)$ and Θ are independent, we can write [see Eqs. (8.64b) and (9.7c)]

$$\begin{aligned} R_\varphi(\tau) &= \overline{m(t)m(t+\tau)}\;\overline{\cos(\omega_c t + \Theta)\cos[\omega_c(t+\tau) + \Theta]} \\ &= \frac{1}{2}R_m(\tau)\cos\omega_c\tau \end{aligned} \tag{9.22a}$$

Consequently,*

$$S_\varphi(f) = \frac{1}{4}[S_m(f+f_c) + S_m(f-f_c)] \tag{9.22b}$$

From Eq. (9.22a) it follows that

$$\overline{\varphi^2(t)} = R_\varphi(0) = \frac{1}{2}R_m(0) = \frac{1}{2}\overline{m^2(t)} \tag{9.22c}$$

* We obtain the same result even if $\varphi(t) = m(t)\sin(\omega_c t + \Theta)$.

Hence, the power of the DSB-SC-modulated signal is half the power of the modulating signal. We derived the same result earlier [Eq. (3.93)] for deterministic signals.

We note that, without the random phase Θ, a DSB-SC amplitude-modulated signal $m(t)\cos(\omega_c t)$ is in fact not wide-sense stationary. To find its PSD, we can resort to the time autocorrelation concept of Chapter 3.

Example 9.5 Random Binary Process

In this example we shall consider a random binary process for which a typical sample function is shown in Fig. 9.9a. The signal can assume only two states (values), 1 or -1, with equal probability. The transition from one state to another can take place only at node points, which occur every T_b seconds. The probability of a transition from one state to the other is 0.5. The first node is equally likely to be situated at any instant within the interval 0 to T_b from the origin. Analytically, we can represent $x(t)$ as

$$x(t) = \sum_n a_n p(t - nT_b - \alpha)$$

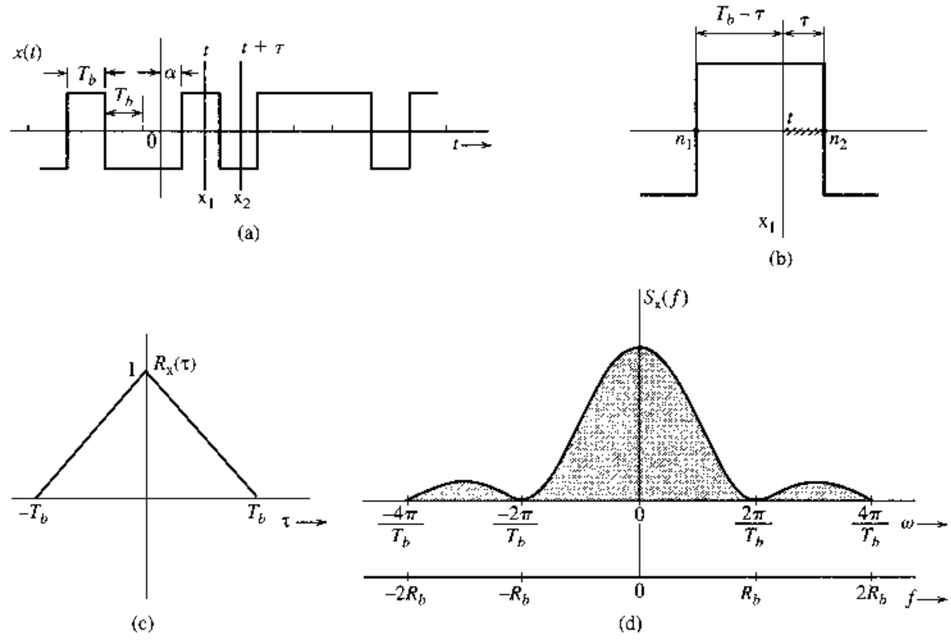


Figure 9.9 Derivation of autocorrelation function and PSD of a random binary process.

where α is an RV uniformly distributed over the range $(0, T_b)$ and $p(t)$ is the basic pulse (in this case $\Pi[(t - T_b/2)/T_b]$). Note that α is the distance of the first node from the origin, and it varies randomly from sample function to sample function. In addition, a_n is random, taking values 1 or -1 with equal probability. The amplitudes at t represent RV x_1, and those at $t+\tau$ represent RV x_2. Note that x_1 and x_2 are discrete and each can assume only two values, −1 and 1. Hence,

$$R_x(\tau) = \overline{x_1 x_2} = \sum_{x_1}\sum_{x_2} x_1 x_2 P_{x_1x_2}(x_1, x_2)$$

$$= P_{x_1x_2}(1, 1) + P_{x_1x_2}(-1, -1) - P_{x_1x_2}(-1, 1) - P_{x_1x_2}(1, -1) \qquad (9.23a)$$

By symmetry, the first two terms and the last two terms on the right-hand side are equal. Therefore,

$$R_x(\tau) = 2[P_{x_1x_2}(1, 1) - P_{x_1x_2}(1, -1)] \qquad (9.23b)$$

From Bayes' rule, we have

$$R_x(\tau) = 2P_{x_1}(1)[P_{x_2|x_1}(1|1) - P_{x_2|x_1}(-1|1)]$$

$$= P_{x_2|x_1}(1|1) - P_{x_2|x_1}(-1|1) \qquad (9.23c)$$

Moreover,

$$P_{x_2|x_1}(1|1) = 1 - P_{x_2|x_1}(-1|1)$$

Hence,

$$R_x(\tau) = 1 - 2P_{x_2|x_1}(-1|1)$$

It is helpful to compute $R_x(\tau)$ for small values of τ first. Let us consider the case of $\tau < T_b$, where, at most, one node is in the interval t to $t+\tau$. In this case, the event $x_2 = -1$ given $x_1 = 1$ is a joint event $A \cap B$, where the event A is "a node in the interval $(t,\ t+\tau)$" and B is "the state change at this node." Because A and B are independent events,

$$P_{x_2|x_1}(-1|1) = P(\text{a node lies in } t \text{ to } t+\tau)P(\text{state change})$$

$$= \tfrac{1}{2}P(\text{a node lies in } t \text{ to } t+\tau)$$

Figure 9.9b shows adjacent nodes n_1 and n_2, between which t lies. We mark off the interval τ from the node n_2. If t lies anywhere in this interval (sawtooth line), the node n_2 lies within t and $t+\tau$. But because the instant t is chosen arbitrarily between nodes n_1 and n_2, it is equally likely to be at any instant over the T_b seconds between n_1 and n_2, and the probability that t lies in the shaded interval is simply τ/T_b. Therefore,

$$P_{x_2|x_1}(-1|1) = \frac{1}{2}\left(\frac{\tau}{T_b}\right) \qquad (9.24)$$

and

$$R_x(\tau) = 1 - \frac{\tau}{T_b} \qquad \tau < T_b \qquad (9.25)$$

Because $R_x(\tau)$ is an even function of τ, we have

$$R_x(\tau) = 1 - \frac{|\tau|}{T_b} \qquad |\tau| < T_b \qquad (9.26)$$

Next, consider the range $\tau > T_b$. In this case at least one node lies in the interval t to $t+\tau$. Hence, x_1 and x_2 become independent, and

$$R_x(\tau) = \overline{x_1 x_2} = \overline{x_1}\,\overline{x_2} = 0 \qquad \tau > T_b$$

where, by inspection, we observe that $\bar{x}_1 = \bar{x}_2 = 0$ (Fig. 9.9a). This result can also be obtained by observing that for $|\tau| > T_b$, x_1 and x_2 are independent, and it is equally likely that $x_2 = 1$ or -1 given that $x_1 = 1$ (or -1). Hence, all four probabilities in Eq. (9.23a) are equal to $1/4$, and

$$R_x(\tau) = 0 \qquad \tau > T_b$$

Therefore,

$$R_x(\tau) = \begin{cases} 1 - |\tau|/T_b & |\tau| < T_b \\ 0 & |\tau| > T_b \end{cases} \tag{9.27a}$$

and

$$S_x(f) = T_b \operatorname{sinc}^2(\pi f T_b) \tag{9.27b}$$

The autocorrelation function and the PSD of this process are shown in Fig. 9.9c and d. Observe that $\overline{x^2} = R_x(0) = 1$, as expected.

The random binary process described in Example 9.5 is sometimes known as the telegraph signal. This process also coincides with the polar signaling of Sec. 7.2.2 when the pulse shape is a rectangular NRZ pulse (Fig. 7.2). For wide-sense stationarity, the signal's initial starting point α is randomly distributed.

Let us now consider a more general case of the pulse train $y(t)$, discussed in Sec. 7.2 (Fig. 7.4). From the knowledge of the PSD of this train, we can derive the PSD of on-off, polar, bipolar, duobinary, split-phase, and many more important digital signals.

Example 9.6 Random PAM Pulse Train

Digital data is transmitted by using a basic pulse $p(t)$, as shown in Fig. 9.10a. The successive pulses are separated by T_b seconds, and the kth pulse is $a_k p(t)$, where a_k is an RV. The distance α of the first pulse (corresponding to $k = 0$) from the origin is equally likely to be any value in the range $(0, T_b)$. Find the autocorrelation function and the PSD of such a random pulse train $y(t)$ whose sample function is shown in Fig. 9.10b. The random process $y(t)$ can be described as

$$y(t) = \sum_{k=-\infty}^{\infty} a_k p(t - kT_b - \alpha)$$

where α is an RV uniformly distributed in the interval $(0, T_b)$. Thus, α is different for each sample function. Note that $p(\alpha) = 1/T_b$ over the interval $(0, T_b)$ and is zero everywhere else.* It can be shown that $\overline{y(t)} = (\overline{a_k}/T_b)\int_{-\infty}^{\infty} p(t)\,dt$ is a constant.†

---

* If $\alpha = 0$, the process can be expressed as $y(t) = \sum_{k=-\infty}^{\infty} a_k p(t - kT_b)$. In this case $\overline{y(t)} = \overline{a_k}\sum_{k=-\infty}^{\infty} p(t - kT_b)$ is not constant, but is periodic with period T_b. Similarly, we can show that the autocorrelation function is periodic with the same period T_b. This is an example of a **cyclostationary**, or periodically stationary, process (a process whose statistics are invariant to a shift of the time origin by integral multiples of a constant T_b). Cyclostationary processes, as seen here, are clearly not wide-sense stationary. But they can be made wide-sense stationary with slight modification by adding the RV α in the expression of $y(t)$, as in this example.

† Using exactly the same approach, as seen shortly in the derivation of Eq. (9.28), we can show that $\overline{y(t)} = (\overline{a_k}/T_b)\int_{-\infty}^{\infty} p(t)\,dt$.

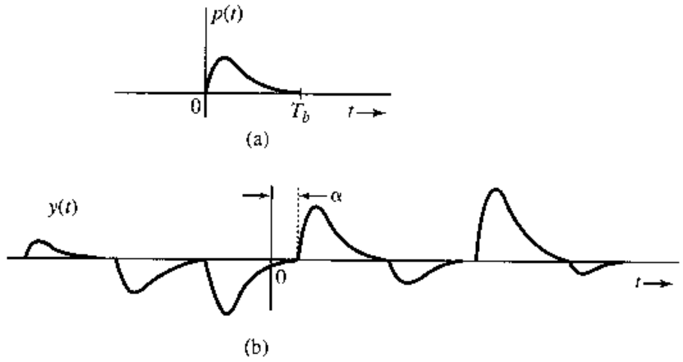


Figure 9.10 Random PAM process.

We have the expression

$$\begin{aligned}
R_y(\tau) &= \overline{y(t)y(t+\tau)} \\
&= \overline{\sum_{k=-\infty}^{\infty} a_k p(t - kT_b - \alpha) \sum_{m=-\infty}^{\infty} a_m p(t + \tau - mT_b - \alpha)} \\
&= \sum_{k=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \overline{a_k a_m p(t - kT_b - \alpha) p(t + \tau - mT_b - \alpha)}
\end{aligned}$$

Because a_k and a_m are independent of α,

$$R_y(\tau) = \sum_{k=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \overline{a_k a_m} \cdot \overline{p(t - kT_b - \alpha) \cdot p(t + \tau - mT_b - \alpha)}$$

Both k and m are integers. Letting $m = k + n$, this expression can be written

$$R_y(\tau) = \sum_{k=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \overline{a_k a_{k+n}} \cdot \overline{p(t - kT_b - \alpha) \cdot p(t + \tau - [k+n]T_b - \alpha)}$$

The first term under the double sum is the correlation of RVs a_k and a_{k+n} and will be denoted by $\mathcal{R}_n$. The second term, being a mean with respect to the RV α, can be expressed as an integral. Thus,

$$R_y(\tau) = \sum_{n=-\infty}^{\infty} \mathcal{R}_n \sum_{k=-\infty}^{\infty} \int_0^{T_b} p(t - kT_b - \alpha) p(t + \tau - [k+n]T_b - \alpha) p(\alpha)\, d\alpha$$

Recall that α is uniformly distributed over the interval 0 to T_b. Hence, $p(\alpha) = 1/T_b$ over the interval $(0, T_b)$, and is zero otherwise. Therefore,

$$\begin{aligned} R_{\mathrm{y}}(\tau) &= \sum_{n=-\infty}^{\infty} \mathcal{R}_n \sum_{k=-\infty}^{\infty} \frac{1}{T_b} \int_0^{T_b} p(t - kT_b - \alpha)p(t + \tau - [k+n]T_b - \alpha)\, d\alpha \\ &= \frac{1}{T_b} \sum_{n=-\infty}^{\infty} \mathcal{R}_n \sum_{k=-\infty}^{\infty} \int_{t-(k+1)T_b}^{t-kT_b} p(\beta)p(\beta + \tau - nT_b)\, d\beta \\ &= \frac{1}{T_b} \sum_{n=-\infty}^{\infty} \mathcal{R}_n \int_{-\infty}^{\infty} p(\beta)p(\beta + \tau - nT_b)\, d\beta \end{aligned}$$

The integral on the right-hand side is the time autocorrelation function of the pulse $p(t)$ with the argument $\tau - nT_b$. Thus,

$$R_{\mathrm{y}}(\tau) = \frac{1}{T_b} \sum_{n=-\infty}^{\infty} \mathcal{R}_n \, \psi_p(\tau - nT_b) \tag{9.28}$$

where

$$\mathcal{R}_n = \overline{\mathrm{a}_k \mathrm{a}_{k+n}} \tag{9.29}$$

and

$$\psi_p(\tau) = \int_{-\infty}^{\infty} p(t)p(t+\tau)\, dt \tag{9.30}$$

As seen in Eq. (3.74), if $p(t) \Longleftrightarrow P(f)$, then $\psi_p(\tau) \Longleftrightarrow |P(f)|^2$. Therefore, the PSD of $\mathrm{y}(t)$, which is the Fourier transform of $R_{\mathrm{y}}(\tau)$, is given by

$$\begin{aligned} S_{\mathrm{y}}(f) &= \frac{1}{T_b} \sum_{n=-\infty}^{\infty} \mathcal{R}_n |P(f)|^2 e^{-jn2\pi f T_b} \\ &= \frac{|P(f)|^2}{T_b} \sum_{n=-\infty}^{\infty} \mathcal{R}_n e^{-jn2\pi f T_b} \end{aligned} \tag{9.31}$$

This result is similar to that found in Eq. (7.11b). The only difference is the use of the ensemble average in defining $\mathcal{R}_n$ in this chapter, whereas R_n in Chapter 7 is the time average.

Example 9.7 Find the PSD $S_{\mathrm{y}}(f)$ for a polar binary random signal where **1** is transmitted by a pulse $p(t)$ (Fig. 9.11) whose Fourier transform is $P(f)$, and **0** is transmitted by $-p(t)$. The digits **1** and **0** are equally likely, and one digit is transmitted every T_b seconds. Each digit is independent of the other digits.

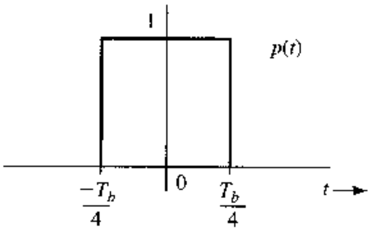


Figure 9.11 Basic pulse for a random binary polar signal.

In this case, a_k can take on values 1 and −1 with probability 1/2 each. Hence,

$$\overline{a_k} = \sum_{k=1,-1} a_k P(a_k) = (1)P_{a_k}(1) + (-1)P_{a_k}(-1)$$

$$= \frac{1}{2} - \frac{1}{2} = 0$$

$$\mathcal{R}_0 = \overline{a_k^2} = \sum_{k=1,-1} a_k^2 P(a_k) = (1)^2 P_{a_k}(1) + (-1)^2 P_{a_k}(-1)$$

$$= \frac{1}{2}(1)^2 + \frac{1}{2}(-1)^2 = 1$$

and because each digit is independent of the remaining digits,

$$\mathcal{R}_n = \overline{a_k a_{k+n}} = \overline{a_k}\ \overline{a_{k+n}} = 0 \qquad n \geq 1$$

Hence, from Eq. (9.31),

$$S_y(f) = \frac{|P(f)|^2}{T_b}$$

We already found this result in Eq. (7.13), where we used time averaging instead of ensemble averaging. When a process is ergodic of second order (or higher), the ensemble and time averages yield the same result. Note that Example 9.5 is a special case of this result, where $p(t)$ is a full-width rectangular pulse $\Pi(t/T_b)$ with $P(f) = T_b \operatorname{sinc}(\pi f T_b)$, and

$$S_y(f) = \frac{|P(f)|^2}{T_b} = T_b \operatorname{sinc}^2(\pi f T_b)$$

Example 9.8 Find the PSD $S_y(f)$ for on-off and bipolar random signals which use a basic pulse for $p(t)$, as shown in Fig. 9.11. The digits 1 and 0 are equally likely, and digits are transmitted every T_b seconds. Each digit is independent of the remaining digits. All these line codes are described in Sec. 7.2.

In each case we shall first determine $\mathcal{R}_0, \mathcal{R}_1, \mathcal{R}_2, \ldots, \mathcal{R}_n$.

(a) *On-off signaling:* In this case, a_n can take on values 1 and 0 with probability 1/2 each. Hence,

$$\overline{a_k} = (1)P_{a_k}(1) + (0)P_{a_k}(0) = \frac{1}{2}(1) + \frac{1}{2}(0) = \frac{1}{2}$$

$$\mathcal{R}_0 = \overline{a_k^2} = (1)^2 P_{a_k}(1) + (0)^2 P_{a_k}(0) = \frac{1}{2}(1)^2 + \frac{1}{2}(0)^2 = \frac{1}{2}$$

and because each digit is independent of the remaining digits,

$$\mathcal{R}_n = \overline{a_k a_{k+n}} = \overline{a_k}\,\overline{a_{k+n}} = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{1}{4} \qquad n \geq 1$$

Therefore, from Eq. (9.31),

$$S_y(f) = \frac{|P(f)|^2}{T_b}\left[\frac{1}{2} + \frac{1}{4}\sum_{\substack{n=-\infty \\ n\neq 0}}^{\infty} e^{-jn2\pi f T_b}\right] \tag{9.32a}$$

$$= \frac{|P(f)|^2}{T_b}\left[\frac{1}{4} + \frac{1}{4}\sum_{n=-\infty}^{\infty} e^{-jn2\pi f T_b}\right] \tag{9.32b}$$

Equation (9.32b) is obtained from Eq. (9.32a) by splitting the term 1/2 corresponding to $\mathcal{R}_0$ into two: 1/4 outside the summation and 1/4 inside the summation (corresponding to $n = 0$). This result is identical to Eq. (7.18b) found earlier by using time averages.

We now use a Poisson summation formula,*

$$\sum_{n=-\infty}^{\infty} e^{-jn2\pi f T_b} = \frac{1}{T_b}\sum_{n=-\infty}^{\infty} \delta\left(f - \frac{n}{T_b}\right)$$

Substitution of this result into Eq. (9.32b) yields

$$S_y(f) = \frac{|P(f)|^2}{4T_b}\left[1 + \frac{1}{T_b}\sum_{n=-\infty}^{\infty} \delta\left(f - \frac{n}{T_b}\right)\right] \tag{9.32c}$$

Note that the spectrum $S_y(f)$ consists of both a discrete and a continuous part. A discrete component of clock frequency ($R_b = 1/T_b$) is present in the spectrum. The continuous component of the spectrum is $|P(f)|^2/4T_b$ is identical (except for a scaling factor 1/4) to the spectrum of the polar signal in Example 9.7. This is a logical result because as Fig. 7.3 shows, an on-off signal can be expressed as a sum of a

---

* The impulse train in Fig. 3.23a is $\delta_{T_b}(t)$ can be expressed as $\delta_{T_b}(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT_b)$. Also $\delta(t - nT_b) \Leftrightarrow e^{-jn2\pi f T_b}$. Hence, the Fourier transform of this impulse train is $\sum_{n=-\infty}^{\infty} e^{-jn2\pi f T_b}$. But we found the alternate form of the Fourier transform of this train in Eq. (3.43) (Example 3.11). Hence,

$$\sum_{n=-\infty}^{\infty} e^{jn2\pi f T_b} = \frac{1}{T_b}\sum_{n=-\infty}^{\infty} \delta\left(f - \frac{n}{T_b}\right)$$

polar and a periodic component. The polar component is exactly half the polar signal discussed earlier. Hence, the PSD of this component is one-fourth of the PSD of the polar signal. The periodic component is of clock frequency R_b, and consists of discrete components of frequency R_b and its harmonics.

(b) *Bipolar signaling:* In this case, a_k can take on values 0, 1, and -1 with probabilities 1/2, 1/4, and 1/4, respectively. Hence,

$$\overline{a_k} = (0)P_{a_k}(0) + (1)P_{a_k}(1) + (-1)P_{a_k}(-1)$$
$$= \frac{1}{2}(0) + \frac{1}{4}(1) + \frac{1}{4}(-1) = 0$$
$$\mathcal{R}_0 = \overline{a_k^2} = (0)^2 P_{a_k}(0) + (1)^2 P_{a_k}(1) + (-1)^2 P_{a_k}(-1)$$
$$= \frac{1}{2}(0)^2 + \frac{1}{4}(1)^2 + \frac{1}{4}(-1)^2 = \frac{1}{2}$$

Also,

$$\mathcal{R}_1 = \overline{a_k a_{k+1}} = \sum_k \sum_{k+1} a_k a_{k+1} P_{a_k a_{k+1}}(a_k a_{k+1})$$

Because a_k and a_{k+1} can take three values each, the sum on the right-hand side has nine terms, of which only four terms (corresponding to values ± 1 for a_k and a_{k+1}) are nonzero. Thus,

$$\mathcal{R}_1 = (1)(1)P_{a_k a_{k+1}}(1, 1) + (-1)(1)P_{a_k a_{k+1}}(-1, 1)$$
$$+ (1)(-1)P_{a_k a_{k+1}}(1, -1) + (-1)(-1)P_{a_k a_{k+1}}(-1, -1)$$

Because of the bipolar rule,

$$P_{a_k a_{k+1}}(1, 1) = P_{a_k a_{k+1}}(-1, -1) = 0$$

and

$$P_{a_k a_{k+1}}(-1, 1) = P_{a_k}(-1)P_{a_{k+1}|a_k}(1|-1) = \left(\frac{1}{4}\right)\left(\frac{1}{2}\right) = \frac{1}{8}$$

Similarly, we find $P_{a_k a_{k+1}}(1, -1) = 1/8$. Substitution of these values in $\mathcal{R}_1$ yields

$$\mathcal{R}_1 = -\frac{1}{4}$$

For $n \geq 2$, the pulse strengths a_k and a_{k+1} become independent. Hence,

$$\mathcal{R}_n = \overline{a_k a_{k+n}} = \overline{a_k}\ \overline{a_{k+n}} = (0)(0) = 0 \qquad n \geq 2$$

Substitution of these values in Eq. (9.31) and noting that $\mathcal{R}_n$ is an even function of n, yields

$$S_y(f) = \frac{|P(f)|^2}{T_b} \sin^2(\pi f T_b)$$

This result is identical to Eq. (7.21b) found earlier by using time averages.

9.4 MULTIPLE RANDOM PROCESSES

For two real random processes $x(t)$ and $y(t)$, we define the **cross-correlation function*** $R_{xy}(t_1, t_2)$ as

$$R_{xy}(t_1, t_2) = \overline{x(t_1)y(t_2)} \tag{9.33a}$$

The two processes are said to be **jointly stationary** (in the wide sense) if each of the processes is individually wide-sense stationary and if

$$\begin{aligned} R_{xy}(t_1, t_2) &= R_{xy}(t_2 - t_1) \\ &= R_{xy}(\tau) \end{aligned} \tag{9.33b}$$

Uncorrelated, Orthogonal (Incoherent), and Independent Processes

Two processes $x(t)$ and $y(t)$ are said to be **uncorrelated** if their cross-correlation function is equal to the product of their means; that is,

$$R_{xy}(\tau) = \overline{x(t)y(t+\tau)} = \bar{x}\bar{y} \tag{9.34}$$

This implies that RVs $x(t)$ and $y(t+\tau)$ are uncorrelated for all t and τ.

Processes $x(t)$ and $y(t)$ are said to be **incoherent**, or **orthogonal**, if

$$R_{xy}(\tau) = 0 \tag{9.35}$$

Incoherent, or orthogonal, processes are uncorrelated processes with $\bar{x}$ and/or $\bar{y} = 0$.

Processes $x(t)$ and $y(t)$ are **independent** random processes if the random variables $x(t_1)$ and $y(t_2)$ are independent for all possible choices of t_1 and t_2.

Cross-Power Spectral Density

We define the **cross-power spectral density** $S_{xy}(f)$ for two random processes $x(t)$ and $y(t)$ as

$$S_{xy}(f) = \lim_{T\to\infty} \frac{\overline{X_T^*(f)Y_T(f)}}{T} \tag{9.36}$$

where $X_T(f)$ and $Y_T(f)$ are the Fourier transforms of the truncated processes $x(t)\,\Pi(t/T)$ and $y(t)\,\Pi(t/T)$, respectively. Proceeding along the lines of the derivation of Eq. (9.16), it can be shown that†

$$R_{xy}(\tau) \Longleftrightarrow S_{xy}(f) \tag{9.37a}$$

It can be seen from Eqs. (9.33) that for real random processes $x(t)$ and $y(t)$,

$$R_{xy}(\tau) = R_{yx}(-\tau) \tag{9.37b}$$

Therefore,

$$S_{xy}(f) = S_{yx}(-f) \tag{9.37c}$$

---

* For complex random processes, the cross-correlation function is defined as

$$R_{xy}(t_1, t_2) = \overline{x^*(t_1)y(t_2)}$$

† Equation (9.37a) is valid for complex processes as well.

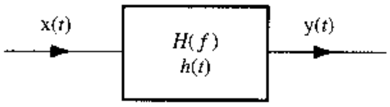


Figure 9.12 Transmission of a random process through a linear time-invariant system.

9.5 TRANSMISSION OF RANDOM PROCESSES THROUGH LINEAR SYSTEMS

If a random process $x(t)$ is applied at the input of a *stable* linear time-invariant system (Fig. 9.12) with transfer function $H(f)$, we can determine the autocorrelation function and the PSD of the output process $y(t)$. We now show that

$$R_y(\tau) = h(\tau) * h(-\tau) * R_x(\tau) \tag{9.38}$$

and

$$S_y(f) = |H(f)|^2 S_x(f) \tag{9.39}$$

To prove this, we observe that

$$y(t) = \int_{-\infty}^{\infty} h(\alpha)x(t-\alpha)\,d\alpha$$

and

$$y(t+\tau) = \int_{-\infty}^{\infty} h(\alpha)x(t+\tau-\alpha)\,d\alpha$$

Hence,*

$$\begin{aligned} R_y(\tau) = \overline{y(t)y(t+\tau)} &= \overline{\int_{-\infty}^{\infty} h(\alpha)x(t-\alpha)\,d\alpha \int_{-\infty}^{\infty} h(\beta)x(t+\tau-\beta)\,d\beta} \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(\alpha)h(\beta)\overline{x(t-\alpha)x(t+\tau-\beta)}\,d\alpha\,d\beta \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(\alpha)h(\beta)R_x(\tau+\alpha-\beta)\,d\alpha\,d\beta \end{aligned}$$

This double integral is precisely the double convolution $h(\tau)*h(-\tau)*R_x(\tau)$. Hence, Eqs. (9.38) and (9.39) follow.

Example 9.9 Thermal Noise

Random thermal motion of electrons in a resistor R causes a random voltage across its terminals. This voltage $n(t)$ is known as the **thermal noise**. Its PSD $S_n(f)$ is practically flat over a very large band (up to 1000 GHz at room temperature) and is given by[1]

$$S_n(f) = 2kTR \tag{9.40}$$

* In this development, we interchange the operations of averaging and integrating. Because averaging is really an operation of integration, we are really changing the order of integration, and we assume that such a change is permissible.

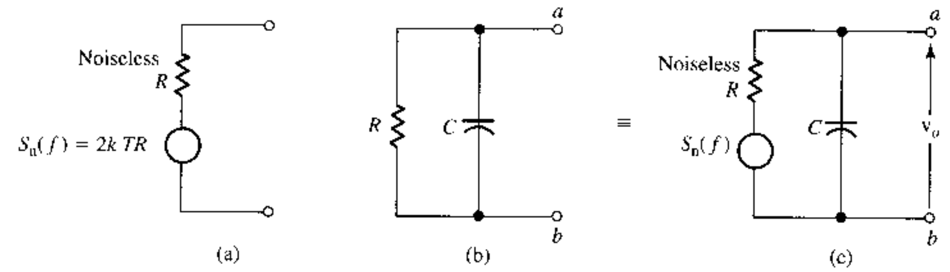


Figure 9.13 Thermal noise representation in a resistor.

where k is the Boltzmann constant (1.38×10^{-23}) and T is the ambient temperature in kelvins. A resistor R at a temperature T kelvin can be represented by a noiseless resistor R in series with a random white-noise voltage source (thermal noise) having a PSD of $2kTR$ (Fig. 9.13a). Observe that the thermal noise power over a band Δf is $(2kTR)\,2\Delta f = 4kTR\Delta f$.

Let us calculate the thermal noise voltage (rms value) across the simple RC circuit in Fig. 9.13b. The resistor R is replaced by an equivalent noiseless resistor in series with the thermal noise voltage source. The transfer function $H(f)$ relating the voltage v_o at terminals a–b to the thermal noise voltage is given by

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

If $S_0(f)$ is the PSD of the voltage v_o, then from Eq. (9.39) we have

$$S_0(f) = \left|\frac{1}{1 + j2\pi fRC}\right|^2 2kTR$$

$$= \frac{2kTR}{1 + 4\pi^2 f^2 R^2 C^2}$$

The mean square value $\overline{v_o^2}$ is given by

$$\overline{v_o^2} = \int_{-\infty}^{\infty} \frac{2kTR}{1 + 4\pi^2 f^2 R^2 C^2}\, df$$

$$= \frac{kT}{C} \tag{9.41}$$

Hence, the rms thermal noise voltage across the capacitor is $\sqrt{kT/C}$.

Sum of Random Processes

If two stationary processes (at least in the wide sense) $x(t)$ and $y(t)$ are added to form a process $z(t)$, the statistics of $z(t)$ can be determined in terms of those of $x(t)$ and $y(t)$. If

$$z(t) = x(t) + y(t) \tag{9.42a}$$

then

$$R_z(\tau) = \overline{z(t)z(t+\tau)} = \overline{[x(t) + y(t)][x(t+\tau) + y(t+\tau)]}$$

$$= R_x(\tau) + R_y(\tau) + R_{xy}(\tau) + R_{yx}(\tau) \tag{9.42b}$$

If $x(t)$ and $y(t)$ are uncorrelated, then from Eq. (9.34),

$$R_{xy}(\tau) = R_{yx}(\tau) = \bar{x}\bar{y}$$

and

$$R_z(\tau) = R_x(\tau) + R_y(\tau) + 2\bar{x}\bar{y} \tag{9.43}$$

Most processes of interest in communication problems have zero means. If processes $x(t)$ and $y(t)$ are uncorrelated with either $\bar{x}$ or $\bar{y} = 0$ [i.e., if $x(t)$ and $y(t)$ are incoherent], then

$$R_z(\tau) = R_x(\tau) + R_y(\tau) \tag{9.44a}$$

and

$$S_z(f) = S_x(f) + S_y(f) \tag{9.44b}$$

It also follows from Eqs. (9.44a) and (9.19) that

$$\overline{z^2} = \overline{x^2} + \overline{y^2} \tag{9.44c}$$

Hence, the mean square of a sum of incoherent (or orthogonal) processes is equal to the sum of the mean squares of these processes.

Example 9.10 Two independent random voltage processes $x_1(t)$ and $x_2(t)$ are applied to an RC network, as shown in Fig. 9.14. It is given that

$$S_{x_1}(f) = K \qquad S_{x_2}(f) = \frac{2\alpha}{\alpha^2 + (2\pi f)^2}$$

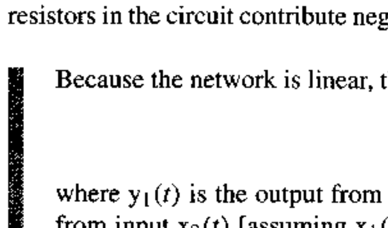


Figure 9.14 Noise calculations in a resistive circuit.

Determine the PSD and the power P_y of the output random process $y(t)$. Assume that the resistors in the circuit contribute negligible thermal noise (i.e., assume that they are noiseless).

Because the network is linear, the output voltage $y(t)$ can be expressed as

$$y(t) = y_1(t) + y_2(t)$$

where $y_1(t)$ is the output from input $x_1(t)$ [assuming $x_2(t) = 0$] and $y_2(t)$ is the output from input $x_2(t)$ [assuming $x_1(t) = 0$]. The transfer functions relating $y(t)$ to $x_1(t)$ and $x_2(t)$ are $H_1(f)$ and $H_2(f)$, respectively, given by

$$H_1(f) = \frac{1}{3(3 \cdot j2\pi f + 1)}, \qquad H_2(f) = \frac{1}{2(3 \cdot j2\pi f + 1)}$$

Hence,

$$S_{y_1}(f) = |H_1(f)|^2 S_{x_1}(f) = \frac{K}{9[9(2\pi f)^2 + 1]}$$

and

$$S_{y_2}(f) = |H_2(f)|^2 S_{x_2}(f) = \frac{\alpha}{2[9(2\pi f)^2 + 1][\alpha^2 + (2\pi f)^2]}$$

Because the input processes $x_1(t)$ and $x_2(t)$ are independent, the outputs $y_1(t)$ and $y_2(t)$ generated by them will also be independent. Also, the PSDs of $y_1(t)$ and $y_2(t)$ have no impulses at $f = 0$, implying that they have no dc components [i.e., $\overline{y_1(t)} = \overline{y_2(t)} = 0$]. Hence, $y_1(t)$ and $y_2(t)$ are incoherent, and

$$\begin{aligned} S_y(f) &= S_{y_1}(f) + S_{y_2}(f) \\ &= \frac{2K[\alpha^2 + (2\pi f)^2] + 9\alpha}{18[9(2\pi f)^2 + 1][\alpha^2 + (2\pi f)^2]} \end{aligned}$$

The power P_y (or the mean square value $\overline{y^2}$) can be determined in two ways. We can find $R_y(\tau)$ by taking the inverse transforms of $S_{y_1}(f)$ and $S_{y_2}(f)$ as

$$R_y(\tau) = \underbrace{\frac{K}{54} e^{-|\tau|/3}}_{R_{y_1}(\tau)} + \underbrace{\frac{3\alpha - e^{-\alpha|\tau|}}{4(9\alpha^2 - 1)}}_{R_{y_2}(\tau)}$$

and

$$P_y = \overline{y^2} = R_y(0) = \frac{K}{54} + \frac{3\alpha - 1}{4(9\alpha^2 - 1)}$$

Alternatively, we can determine $\overline{y^2}$ by integrating $S_y(f)$ with respect to f (or f) [see Eq. (9.19)].

9.6 APPLICATION: OPTIMUM FILTERING (WIENER-HOPF FILTER)

When a desired signal is mixed with noise, the SNR can be improved by passing it through a filter that suppresses frequency components where the signal is weak but the noise is strong. The SNR improvement in this case can be explained qualitatively by considering a case of white noise mixed with a signal $m(t)$ whose PSD decreases at high frequencies. If the filter attenuates higher frequencies more, the signal will be reduced—in fact, distorted. The distortion component $m_\epsilon(t)$ may be considered as bad as added noise. Thus, attenuation of higher frequencies will cause additional noise (from signal distortion), but, in compensation, it will reduce the channel noise, which is strong at high frequencies. Because at higher frequencies the signal has a small power content, the distortion component will be small in comparison to the reduction in channel noise, and the total distortion may be smaller than before.

Let $H_{opt}(f)$ be the optimum filter (Fig. 9.15a). This filter, not being ideal, will cause signal distortion. The distortion signal $m_\epsilon(t)$ can be found from Fig. 9.15b. The distortion signal power N_D appearing at the output is given by

$$N_D = \int_{-\infty}^{\infty} S_m(f)|H_{opt}(f) - 1|^2 \, df$$

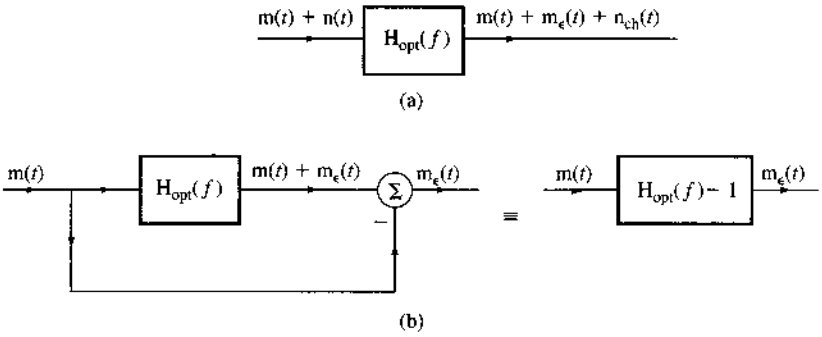


Figure 9.15 Wiener-Hopf filter calculations.

where $S_m(f)$ is the signal PSD at the input of the receiving filter. The channel noise power N_{ch} appearing at the filter output is given by

$$N_{ch} = \int_{-\infty}^{\infty} S_n(f)|H_{opt}(f)|^2 \, df$$

where $S_n(f)$ is the noise PSD appearing at the input of the receiving filter. The distortion component acts as a noise. Because the signal and the channel noise are incoherent, the total noise N_o at the receiving filter output is the sum of the channel noise N_{ch} and the distortion noise N_D,

$$\begin{aligned} N_o &= N_{ch} + N_D \\ &= \int_{-\infty}^{\infty} \left[|H_{opt}(f)|^2 S_n(f) + |H_{opt}(f) - 1|^2 S_m(f) \right] df \end{aligned} \tag{9.45a}$$

Using the fact that $|A + B|^2 = (A + B)(A^* + B^*)$, and noting that both $S_m(f)$ and $S_n(f)$ are real, we can rearrange Eq. (9.45a) as

$$N_o = \int_{-\infty}^{\infty} \left[\left| H_{opt}(f) - \frac{S_m(f)}{S_r(f)} \right|^2 S_r(f) + \frac{S_m(f)S_n(f)}{S_r(f)} \right] df \tag{9.45b}$$

where $S_r(f) = S_m(f) + S_n(f)$. The integrand on the right-hand side of Eq. (9.45b) is nonnegative. Moreover, it is a sum of two nonnegative terms. Hence, to minimize N_o, we must minimize each term. Because the second term $S_m(f)S_n(f)/S_r(f)$ is independent of $H_{opt}(f)$, only the first term can be minimized. From Eq. (9.45b) it is obvious that this term is minimum at zero when

$$\begin{aligned} H_{opt}(f) &= \frac{S_m(f)}{S_r(f)} \\ &= \frac{S_m(f)}{S_m(f) + S_n(f)} \end{aligned} \tag{9.46a}$$

For this optimum choice, the output noise power N_o is given by

$$\begin{aligned} N_o &= \int_{-\infty}^{\infty} \frac{S_m(f)S_n(f)}{S_r(f)} \, df \\ &= \int_{-\infty}^{\infty} \frac{S_m(f)S_n(f)}{S_m(f) + S_n(f)} \, df \end{aligned} \tag{9.46b}$$

The optimum filter is known as the **Wiener-Hopf filter** in the literature. Equation (9.46a) shows that $H_{opt}(f) \approx 1$ (no attenuation) when $S_m(f) \gg S_n(f)$. But when $S_m(f) \ll S_n(f)$, the filter has high attenuation. In other words, the optimum filter attenuates heavily the band where noise is relatively stronger. This causes some signal distortion, but at the same time it attenuates the noise more heavily so that the overall SNR is improved.

Comments on the Optimum Filter

If the SNR at the filter input is reasonably large—for example, $S_m(f) > 100S_n(f)$ (SNR of 20 dB)—the optimum filter [Eq. (9.46a)] in this case is practically an ideal filter, and N_o [Eq. (9.46b)] is given by

$$N_o \simeq \int_{-\infty}^{\infty} S_n(f)\, df$$

Hence for a large input SNR, optimization yields insignificant improvement. The Wiener-Hopf filter is therefore practical only when the input SNR is small (large-noise case).

Another issue is the realizability of the optimum filter in Eq. (9.46a). Because $S_m(f)$ and $S_n(f)$ are both even functions of f, the optimum filter $H_{opt}(f)$ is an even function of f. Hence, the unit impulse response $h_{opt}(t)$ is an even function of t (see Prob. 3.1-1). This makes $h_{opt}(t)$ noncausal and the filter unrealizable. As noted earlier, such a filter can be realized approximately if we are willing to tolerate some delay in the output. If delay cannot be tolerated, the derivation of $H_{opt}(f)$ must be repeated with a realizability constraint. Note that the realizable optimum filter can never be superior to the unrealizable optimum filter [Eq. (9.46a)]. Thus, the filter in Eq. (9.46a) gives the upper bound on performance (output SNR). Discussion of realizable optimum filters can be readily found in the literature[1,2].

Example 9.11 A random process m(t) (the signal) is mixed with a white channel noise n(t). Given

$$S_m(2f) = \frac{2\alpha}{\alpha^2 + (2\pi f)^2} \qquad \text{and} \qquad S_n(f) = \frac{\mathcal{N}}{2}$$

find the Wiener-Hopf filter to maximize the SNR. Find the resulting output noise power N_o.

From Eq. (9.46a),

$$\begin{aligned} H_{opt}(f) &= \frac{4\alpha}{4\alpha + \mathcal{N}[\alpha^2 + (2\pi f)^2]} \\ &= \frac{4\alpha}{\mathcal{N}[\beta^2 + (2\pi f)^2]} \qquad \beta^2 = \frac{4\alpha}{\mathcal{N}} + \alpha^2 \end{aligned} \tag{9.47a}$$

Hence,

$$h_{opt}(t) = \frac{2\alpha}{\mathcal{N}\beta} e^{-\beta|t|} \tag{9.47b}$$

Figure 9.16a shows $h_{opt}(t)$. It is evident that this is an unrealizable filter. However, a delayed version (Fig. 9.16b) of this filter, that is, $h_{opt}(t - t_0)$, is closely realizable if we make $t_0 \geq 3/\beta$ and eliminate the tail for $t < 0$ (Fig. 9.16c).

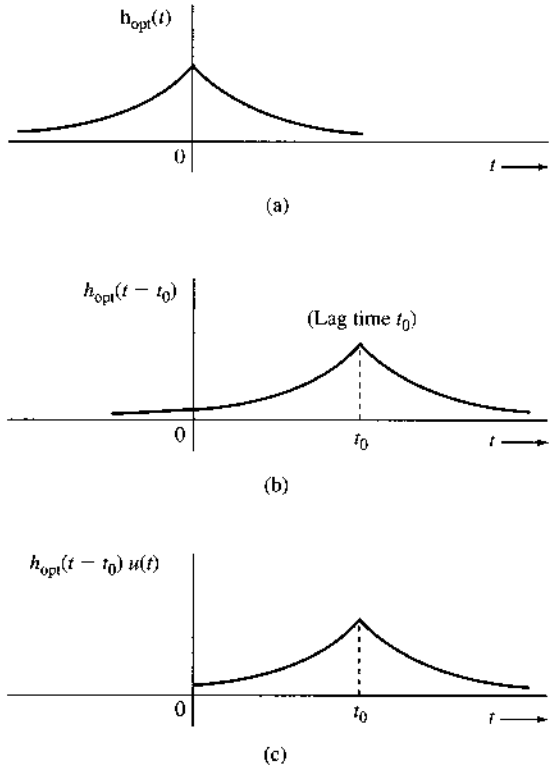


Figure 9.16 Close realization of an unrealizable filter using delay.

The output noise power N_o is [Eq. (9.46b)]

$$N_o = \int_0^{\infty} \frac{2\alpha}{\beta^2 + (2\pi f)^2}\, df = \frac{\alpha}{\beta} = \frac{\alpha}{\sqrt{\alpha^2 + (4\alpha/\mathcal{N})}} \tag{9.48}$$

9.7 APPLICATION: PERFORMANCE ANALYSIS OF BASEBAND ANALOG SYSTEMS

We now apply the concept of power spectral density (PSD) to analyze the performance of baseband analog communication systems. In analog signals, the SNR is basic in specifying the signal quality. For voice signals, an SNR of 5 to 10 dB at the receiver implies a barely intelligible signal. Telephone-quality signals have an SNR of 25 to 35 dB, whereas for television, an SNR of 45 to 55 dB is required.

Figure 9.17 shows a simple communication system in which analog signal $m(t)$ is transmitted at power S_T through a channel (representing a transmission medium). The transmitted signal is corrupted by additive channel noise during transmission. The channel also attenuates (and may also distort) the signal. At the receiver input, we have a signal mixed with noise. The signal and noise powers at the receiver input are S_i and N_i, respectively.

The receiver processes (filters) the signal to yield the output $s_o(t) + n_o(t)$. The noise component $n_o(t)$ came from the processing of $n(t)$ by the receiver, while the signal component $s_o(t)$ came from the message $m(t)$. The signal and noise powers at the receiver output are S_o

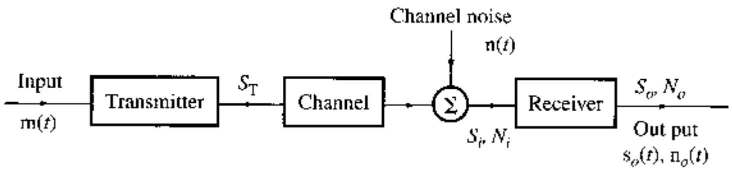


Figure 9.17 Communication system model.

and N_o, respectively. In analog systems, the quality of the received signal is determined by S_o/N_o, the output SNR. Hence, we shall focus our attention on this figure of merit under either a fixed transmission power S_T or for a given S_i.

In baseband systems, the signal is transmitted directly without any modulation. This mode of communication is suitable over a pair of twisted wires or coaxial cables. It is mainly used in short-haul links. For a baseband system, the transmitter and the receiver are ideal baseband filters. The ideal low-pass transmitter limits the input signal spectrum to a given bandwidth, whereas the low-pass receiver eliminates the out-of-band noise and other channel interference. (More elaborate transmitter and receiver filters can be used, as shown in the next section.)

The baseband signal $m(t)$ is assumed to be a zero mean, wide-sense stationary random process band-limited to B Hz. We consider the case of ideal low-pass (or baseband) filters with bandwidth B at the transmitter and the receiver (Fig. 9.17). The channel is assumed to be distortionless. The power, or the mean square value, of $m(t)$ is $\overline{m^2}$, given by

$$S_i = \overline{m^2} = 2\int_0^B S_m((f))\,df \tag{9.49}$$

For this case,

$$S_o = S_i \tag{9.50a}$$

and

$$N_o = 2\int_0^B S_n(f)\,df \tag{9.50b}$$

where $S_n(f)$ is the PSD of the channel noise. For the case of a white noise, $S_n(f) = \mathcal{N}/2$, and

$$N_o = 2\int_0^B \frac{\mathcal{N}}{2}\,df = \mathcal{N}B \tag{9.50c}$$

and

$$\frac{S_o}{N_o} = \frac{S_i}{\mathcal{N}B} \tag{9.50d}$$

We define a parameter γ as

$$\gamma = \frac{S_i}{\mathcal{N}B} \tag{9.51}$$

From Eqs. (9.50d) and (9.51) we have

$$\frac{S_o}{N_o} = \gamma \tag{9.52}$$

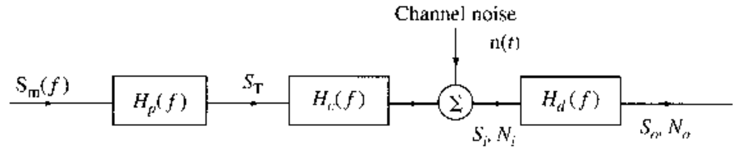


Figure 9.18 Optimum preemphasis and deemphasis filters in baseband systems.

The parameter γ is directly proportional to S_i and, therefore, directly proportional to S_T. Hence, a given S_T (or S_i) implies a given γ. Equation (9.52) is precisely the result we are looking for. It gives the receiver output SNR for a given S_T (or S_i).

The value of the SNR in Eq. (9.52) often serves as a benchmark against which the output SNRs of other modulation systems are measured in practice.

9.8 APPLICATION: OPTIMUM PREEMPHASIS-DEEMPHASIS SYSTEMS

It is possible to increase the output SNR by deliberate distortion of the transmitted signal (preemphasis) and the corresponding compensation (deemphasis) at the receiver. For an intuitive understanding of this process, consider a case of white channel noise and a signal $m(t)$ whose PSD decreases with frequency. In this case, we can boost the high-frequency components of $m(t)$ at the transmitter (preemphasis). Because the signal has relatively less power at high frequencies, this preemphasis will require only a small increase in transmitted power.* At the receiver, the high-frequency components are attenuated (or deemphasized) to undo the preemphasis at the transmitter. This will restore the useful signal to its original form. The channel noise receives an entirely different treatment. Because the noise is added after the transmitter, it does not undergo preemphasis. At the receiver, however, it does undergo deemphasis (i.e., attenuation of high-frequency components). Thus, at the receiver output, the signal power is restored but the noise power is reduced. The output SNR is therefore increased.

In this section, we consider a baseband system. The extension of preemphasis and deemphasis to modulated systems is straightforward. A baseband system with a preemphasis filter $H_p(f)$ at the transmitter and the corresponding complementary deemphasis filter $H_d(f)$ at the receiver is shown in Fig. 9.18. The channel transfer function is $H_c(f)$, and the PSD of the input signal $m(t)$ is $S_m(f)$. We shall determine the optimum preemphasis-deemphasis (PDE) filters $H_p(f)$ and $H_d(f)$ required for distortionless transmission of the signal $m(t)$.

For distortionless transmission,

$$|H_p(f)H_c(f)H_d(f)| = G \quad \text{(a constant)} \tag{9.53a}$$

and

$$\theta_p(f) + \theta_c(f) + \theta_d(f) = -2\pi f t_d \tag{9.53b}$$

We want to maximize the output SNR, S_o/N_o, for a given transmitted power S_T.

Referring to Fig. 9.18, we have

$$S_T = \int_{-\infty}^{\infty} S_m(f)|H_p(f)|^2\, df \tag{9.54a}$$

* Actually, the transmitted power is maintained constant by attenuating the preemphasized signal slightly.

Because $H_p(f)H_c(f)H_d(f) = G\exp(-j2\pi f t_d)$, the signal power S_o at the receiver output is

$$S_o = G^2 \int_{-\infty}^{\infty} S_m(f)\,df \tag{9.54b}$$

The noise power N_o at the receiver output is

$$N_o = \int_{-\infty}^{\infty} S_n(f)|H_d(f)|^2\,df \tag{9.54c}$$

Thus,

$$\frac{S_o}{N_o} = \frac{G^2 \int_{-\infty}^{\infty} S_m(f)\,df}{\int_{-\infty}^{\infty} S_n(f)|H_d(f)|^2\,df} \tag{9.55}$$

We wish to maximize this ratio subject to the condition in Eq. (9.54a) with S_T as a given constant. Applying this power limitation makes the design of $H_p(f)$ a well-posed problem, for otherwise filters with larger gains will always be better. We can include this constraint by multiplying the numerator and the denominator of the right-hand side of Eq. (9.55) by the left-hand side and the right-hand side, respectively, of Eq. (9.54a). This gives

$$\frac{S_o}{N_o} = \frac{G^2 S_T \int_{-\infty}^{\infty} S_m(f)\,df}{\int_{-\infty}^{\infty} S_n(f)|H_d(f)|^2\,df \int_{-\infty}^{\infty} S_m(f)|H_p(f)|^2\,df} \tag{9.56}$$

The numerator of the right-hand side of Eq. (9.56) is fixed and *unaffected* by the PDE filters. Hence, to maximize S_o/N_o, we need only minimize the denominator of the right-hand side of Eq. (9.56). To do this, we use the Cauchy-Schwarz inequality (Appendix B),

$$\int_{-\infty}^{\infty} S_m(f)|H_p(f)|^2\,df \int_{-\infty}^{\infty} S_n(f)|H_d(f)|^2\,df \geq \left| \int_{-\infty}^{\infty} [S_m(f)S_n(f)]^{1/2}|H_p(f)H_d(f)|\,df \right|^2 \tag{9.57}$$

The equality holds if and only if

$$S_m(f)|H_p(f)|^2 = K^2 S_n(f)|H_d(f)|^2 \tag{9.58}$$

where K is an arbitrary constant. Thus to maximize S_o/N_o, Eq. (9.58) must be satisfied. Substitution of Eq. (9.53a) into Eq. (9.58) yields

$$|H_p(f)|^2_{\text{opt}} = GK \frac{\sqrt{S_n(f)/S_m(f)}}{|H_c(f)|} \tag{9.59a}$$

$$|H_d(f)|^2_{\text{opt}} = \frac{G}{K} \frac{\sqrt{S_m(f)/S_n(f)}}{|H_c(f)|} \tag{9.59b}$$

The constant K is found by substituting Eq. (9.59a) into the power constraint of Eq. (9.54a) as

$$K = \frac{S_T}{G \int_{-\infty}^{\infty} [\sqrt{S_m(f)S_n(f)}/|H_c(f)|]\,df} \tag{9.59c}$$

Substitution of this value of K into Eqs. (9.59a) and (9.59b) yields

$$|H_p(f)|^2_{\text{opt}} = \frac{S_T\sqrt{S_n(f)/S_m(f)}}{|H_c(f)| \int_{-\infty}^{\infty}[\sqrt{S_m(f)S_n(f)}/|H_c(f)|]\,df} \tag{9.60a}$$

$$|H_d(f)|^2_{\text{opt}} = \frac{G^2 \int_{-\infty}^{\infty}[\sqrt{S_m(f)S_n(f)}/|H_c(f)|]\,df}{S_T|H_c(f)|\sqrt{S_n(f)/S_m(f)}} \tag{9.60b}$$

The output SNR under optimum conditions is given by Eq. (9.56) with its denominator replaced by the right-hand side of Eq. (9.57). Finally, substituting $|H_p(f)H_d(f)| = G/|H_c(f)|$ leads to

$$\left(\frac{S_o}{N_o}\right)_{\text{opt}} = \frac{S_T \int_{-\infty}^{\infty} S_m(f)\,df}{\left(\int_{-\infty}^{\infty}[\sqrt{S_m(f)S_n(f)}/|H_c(f)|]\,df\right)^2} \tag{9.60c}$$

Equations (9.60a) and (9.60b) give the magnitudes of the optimum filters $H_p(f)$ and $H_d(f)$. The phase functions must be chosen to satisfy the condition of distortionless transmission [Eq. (9.53b)].

Observe that the preemphasis filter in Eq. (9.59a) boosts frequency components where the signal is weak and suppresses frequency components where the signal is strong. The deemphasis filter in Eq. (9.59b) does exactly the opposite. Thus, the signal is unchanged but the noise is reduced.

Example 9.12 Consider the case with $\alpha = 1400\pi$,

$$S_m(f) = \begin{cases} \frac{C}{(2\pi f)^2+\alpha^2} & |f| \le 4000 \\ 0 & |f| \ge 4000 \end{cases} \tag{9.61a}$$

The channel noise is white with PSD

$$S_n(f) = \frac{\mathcal{N}}{2} \tag{9.61b}$$

The channel is assumed to be ideal [$H_c(f) = 1$ and $G = 1$] over the band of interest (0–4000 Hz).

Without preemphasis-deemphasis, we have

$$\begin{aligned} S_o &= \int_{-4000}^{4000} S_m(f)\,df \\ &= 2\int_0^{4000} \frac{C}{(2\pi f)^2+\alpha^2}\,df \qquad \alpha = 1400\pi \\ &= 10^{-4}C \end{aligned}$$

Also, because $G = 1$, the transmitted power $S_T = S_o$,

$$S_o = S_T = 10^{-4}C$$

and the noise power without preemphasis-deemphasis is

$$N_o = \mathcal{N}B = 4000\mathcal{N}$$

Therefore,

$$\frac{S_o}{N_o} = 2.5 \times 10^{-8}\frac{C}{\mathcal{N}} \tag{9.62}$$

The optimum transmitting and receiving filters are given [Eqs. (9.60a) and (9.60b)] by

$$|H_p(f)|^2 = \frac{10^{-4}\sqrt{(2\pi f)^2+\alpha^2}}{\int_{-\infty}^{\infty}\left(1/\sqrt{(2\pi f)^2+\alpha^2}\right)df} = \frac{1.286\sqrt{(2\pi f)^2+\alpha^2}}{10^4} \qquad |f| \le 4000 \tag{9.63a}$$

$$|H_d(f)|^2 = \frac{10^4\int_{-\infty}^{\infty}\left(1/\sqrt{(2\pi f)^2+\alpha^2}\right)df}{\sqrt{(2\pi f)^2+\alpha^2}} = \frac{0.778\times 10^4}{\sqrt{(2\pi f)^2+\alpha^2}} \qquad |f| \le 4000 \tag{9.63b}$$

The output SNR using optimum preemphasis and deemphasis in found from Eq. (9.60c) as

$$\left(\frac{S_o}{N_o}\right)_{\text{opt}} = \frac{(10^{-4}C)^2}{(\mathcal{N}C/2)\left[\int_{-4000}^{4000}\left[1/\sqrt{4\pi^2 f^2+(1400\pi)^2}\right]df\right]^2}$$
$$= 3.3\times 10^{-8}\frac{C}{\mathcal{N}} \tag{9.64}$$

Comparison of Eq. (9.62) with Eq. (9.64) shows that preemphasis-deemphasis has increased the output SNR by a factor of 1.32.

9.9 BANDPASS RANDOM PROCESSES

If the PSD of a random process is confined to a certain passband (Fig. 9.19), the process is a **bandpass** random process. Bandpass random processes can be used effectively to model modulated communication signals and bandpass noises. Just as a bandpass signal can be represented in terms of quadrature components [see Eq. (3.39)], we can express a bandpass random process $x(t)$ in terms of quadrature components as follows:

$$x(t) = x_c(t)\cos\omega_c t + x_s(t)\sin\omega_c t \tag{9.65}$$

In this representation, $x_c(t)$ is known as the **in-phase** component and $x_s(t)$ is known as the **quadrature** component of the bandpass random process.

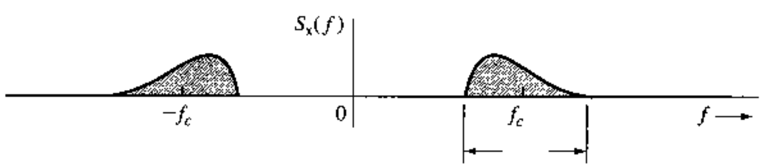


Figure 9.19 PSD of a bandpass random process.

This can be proven by considering the system in Fig. 9.20a, where $H_0(f)$ is an ideal low-pass filter (Fig. 9.20b) with unit impulse response $h_0(t)$. First we show that the system in Fig. 9.20a is an ideal bandpass filter with the transfer function $H(f)$ shown in Fig. 9.20c. This can be conveniently done by computing the response $h(t)$ to the unit impulse input $\delta(t)$. Because the system contains time-varying multipliers, however, we must also test whether it is a time-varying or a time-invariant system. It is therefore appropriate to consider the system response to an input $\delta(t-\alpha)$. This is an impulse at $t = \alpha$. Using the fact that [see Eq. (2.10b)] $f(t)\,\delta(t-\alpha) = f(\alpha)\delta(t-\alpha)$, we can express the signals at various points as follows:

Signal at a_1: $\cos(\omega_c\alpha+\theta)\,\delta(t-\alpha)$

a_2: $\sin(\omega_c\alpha+\theta)\,\delta(t-\alpha)$

b_1: $\cos(\omega_c\alpha+\theta)h_0(t-\alpha)$

b_2: $\sin(\omega_c\alpha+\theta)h_0(t-\alpha)$

c_1: $\cos(\omega_c\alpha+\theta)\cos(\omega_c t+\theta)h_0(t-\alpha)$

c_2: $\sin(\omega_c\alpha+\theta)\sin(\omega_c t+\theta)h_0(t-\alpha)$

d: $h_0(t-\alpha)\,[\cos(\omega_c\alpha+\theta)\cos(\omega_c t+\theta)+\sin(\omega_c\alpha+\theta)\sin(\omega_c t+\theta)]$

$$= 2h_0(t-\alpha)\cos[\omega_c(t-\alpha)]$$

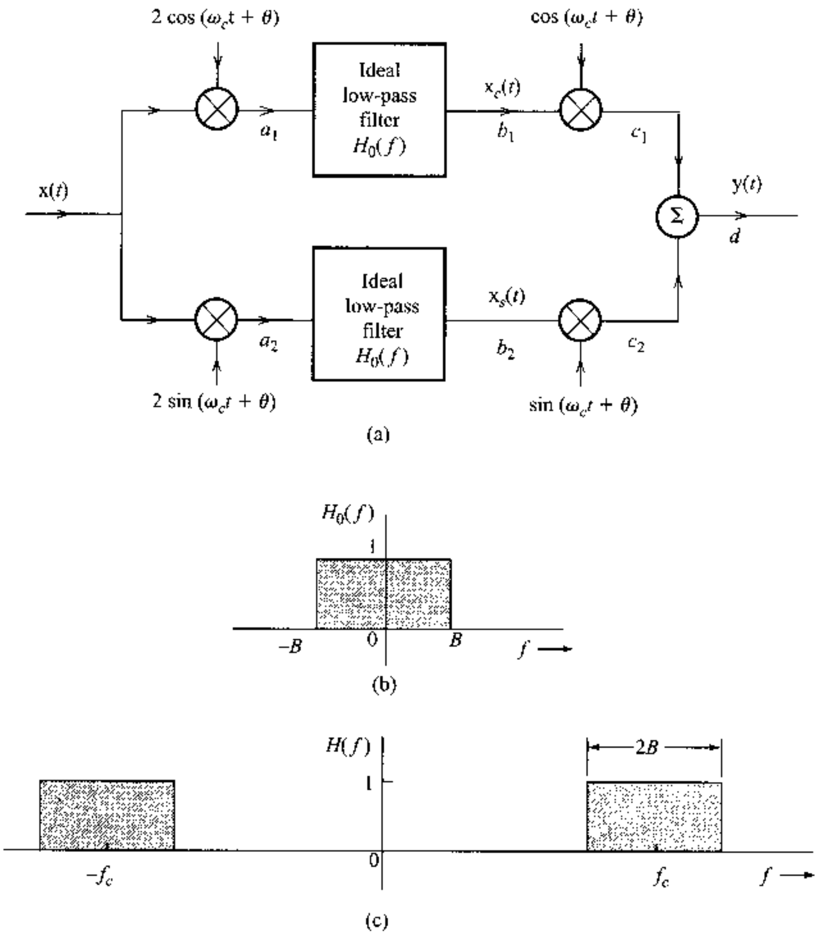


Figure 9.20 (a) Equivalent circuit of an ideal bandpass filter. (b) Ideal low-pass filter frequency response. (c) Ideal bandpass filter frequency response.

Thus, the system response to the input $\delta(t-\alpha)$ is $2h_0(t-\alpha)\cos[\omega_c(t-\alpha)]$. Clearly, this means that the underlying system is linear time invariant, with impulse response

$$h(t) = 2h_0(t)\cos\omega_c t$$

and transfer function

$$H(f) = H_0(f+f_c) + H_0(f-f_c)$$

The transfer function $H(f)$ (Fig. 9.20c) represents an ideal bandpass filter.

If we apply the bandpass process $x(t)$ (Fig. 9.19) to the input of this system, the output $y(t)$ at d will remain the same as $x(t)$. Hence, the output PSD will be the same as the input PSD

$$|H(f)|^2 S_x(f) = S_x(f)$$

If the processes at points b_1 and b_2 (low-pass filter outputs) are denoted by $x_c(t)$ and $x_s(t)$, respectively, then the output $x(t)$ can be written as

$$x(t) = x_c(t)\cos(\omega_c t + \theta) + x_s(t)\sin(\omega_c t + \theta) \tag{9.66}$$

where $x_c(t)$ and $x_s(t)$ are low-pass random processes band-limited to B Hz (because they are the outputs of low-pass filters of bandwidth B). Because Eq. (9.66) is valid for any value of θ, by substituting $\theta = 0$, we get the desired representation in Eq. (9.65).

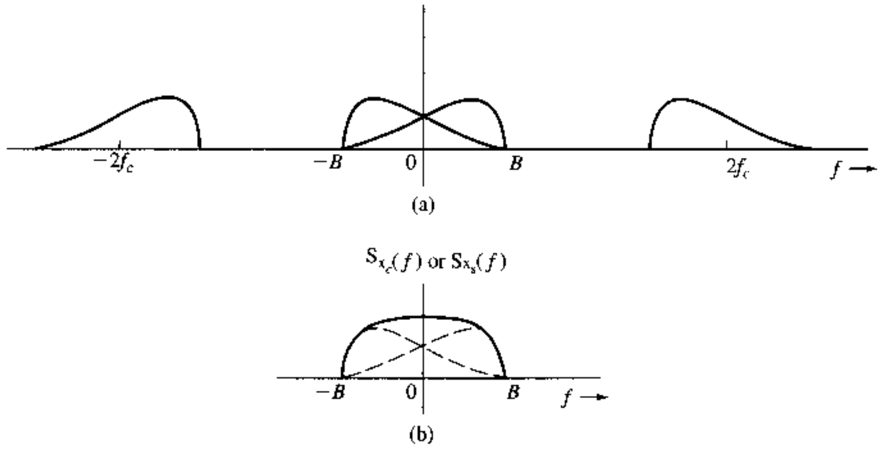
To characterize $x_c(t)$ and $x_s(t)$, consider once again Fig. 9.20a with the input $x(t)$. Let θ be an RV uniformly distributed over the range $(0, 2\pi)$, that is, for a sample function, θ is equally likely to take on any value in the range $(0, 2\pi)$. In this case $x(t)$ is represented as in Eq. (9.66). We observe that $x_c(t)$ is obtained by multiplying $x(t)$ by $2\cos(\omega_c t + \theta)$, and then passing the result through a low-pass filter. The PSD of $2x(t)\cos(\omega_c t + \theta)$ is [see Eq. (9.22b)]

$$4 \times \frac{1}{4}[S_x(f+f_c) + S_x(f-f_c)]$$

This PSD is $S_x(f)$ shifted up and down by f_c, as shown in Fig. 9.21a. When this is passed through a low-pass filter, the resulting PSD of $x_c(t)$ is as shown in Fig. 9.21b. It is clear that

$$S_{x_c}(f) = \begin{cases} S_x(f+f_c) + S_x(f-f_c) & |f| \le B \\ 0 & |f| > B \end{cases} \tag{9.67a}$$

Figure 9.21 Derivation of PSDs of quadrature components of a bandpass random process.



We can obtain $S_{x_s}(f)$ in the same way. As far as the PSD is concerned, multiplication by $\cos(\omega_c t + \theta)$ or $\sin(\omega_c t + \theta)$ makes no difference [see footnote following Eq. (9.22a)], and we get

$$S_{x_c}(f) = S_{x_s}(f) = \begin{cases} S_x(f + f_c) + S_x(f - f_c), & |f| \le B \\ 0, & |f| > B \end{cases} \tag{9.67b}$$

From Figs. 9.19 and 9.21b, we make the interesting observation that the areas under the PSDs $S_x(f)$, $S_{x_c}(f)$, and $S_{x_s}(f)$ are equal. Hence, it follows that

$$\overline{x_c^2(t)} = \overline{x_s^2(t)} = \overline{x^2(t)} \tag{9.67c}$$

Thus, the mean square values (or powers) of $x_c(t)$ and $x_s(t)$ are identical to that of $x(t)$.

These results are derived by assuming Θ to be an RV. For the representation in Eq. (9.65), $\Theta = 0$, and Eqs. (9.67b) and (9.67c) may not be true. Fortunately, those equations hold even for the case of $\Theta = 0$. The proof is rather long and cumbersome and will not be given here.[1-3] It can also be shown[1-3] that

$$\overline{x_c(t)x_s(t)} = R_{x_c x_s}(0) = 0 \tag{9.68}$$

That is, the amplitudes x_c and x_s at any given instant are uncorrelated. Moreover, if $S_x(f)$ is symmetrical about ω_c (as well as $-\omega_c$), then

$$R_{x_c x_s}(\tau) = 0 \tag{9.69}$$

Example 9.13 The PSD of a bandpass white noise $n(t)$ is $\mathcal{N}/2$ (Fig. 9.22a). Represent this process in terms of quadrature components. Derive $S_{n_c}(f)$ and $S_{n_s}(f)$, and verify that $\overline{n_c^2} = \overline{n_s^2} = \overline{n^2}$.

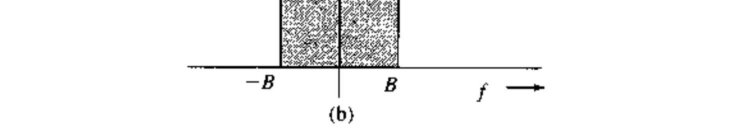


Figure 9.22 (a) PSD of a bandpass white noise process. (b) PSD of its quadrature components.

We have the expression

$$n(t) = n_c(t)\cos\omega_c t + n_s(t)\sin\omega_c t \tag{9.70}$$

where

$$S_{n_c}(f) = S_{n_s}(f) = \begin{cases} S_n(f + f_c) + S_n(f - f_c) & |f| \le B \\ 0 & |f| > B \end{cases}$$

It follows from this equation and from Fig. 9.22 that

$$S_{n_c}(f) = S_{n_s}(f) = \begin{cases} \mathcal{N} & |f| \leq B \\ 0 & |f| > B \end{cases} \tag{9.71}$$

Also,

$$\overline{n^2} = 2\int_{f_c-B}^{f_c+B} \frac{\mathcal{N}}{2}\, df = 2\mathcal{N}B \tag{9.72a}$$

From Fig. 9.22b it follows that

$$\overline{n_c^2} = \overline{n_s^2} = 2\int_0^B \mathcal{N}\, df = 2\mathcal{N}B \tag{9.72b}$$

Hence,

$$\overline{n_c^2} = \overline{n_s^2} = \overline{n^2} = 2\mathcal{N}B \tag{9.72c}$$

Nonuniqueness of the Quadrature Representation

No unique center frequency exists for a bandpass signal. For the spectrum in Fig. 9.23a, for example, we may consider the spectrum to have a bandwidth $2B$ centered at ω_c. The same spectrum can be considered to have a bandwidth $2B'$ centered at ω_1, as also shown in Fig. 9.23a. The quadrature representation [Eq. (9.65)] is also possible for center frequency ω_1:

$$x(t) = x_{c_1}(t)\cos\omega_1 t + x_{s_1}(t)\sin\omega_1 t$$

where

$$S_{x_{c_1}}(f) = S_{x_{s_1}}(f) = \begin{cases} S_x(f+f_1) + S_x(f-f_1) & |f| \leq B' \\ 0 & |f| > B' \end{cases} \tag{9.73}$$

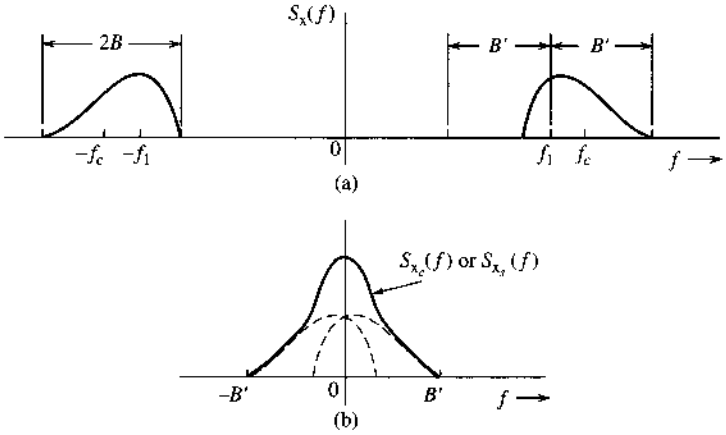


Figure 9.23 Nonunique nature of quadrature component representation of a bandpass process.

This is shown in Fig. 9.23b. Thus, the quadrature representation of a bandpass process is not unique. An infinite number of possible choices exist for the center frequency, and **corresponding to each center frequency is a distinct quadrature representation.**

Example 9.14 A bandpass white noise PSD of an SSB channel (lower sideband) is shown in Fig. 9.24a. Represent this signal in terms of quadrature components with the carrier frequency ω_c.

The true center frequency of this PSD is not ω_c; but we can still use ω_c as the center frequency, as discussed earlier,

$$\mathrm{n}(t) = \mathrm{n}_c(t)\cos\omega_c t + \mathrm{n}_s(t)\sin\omega_c t \tag{9.74}$$

The PSD $S_{\mathrm{n}_c}(f)$ or $S_{\mathrm{n}_s}(f)$ obtained by shifting $S_{\mathrm{n}}(f)$ up and down by f_c [see Eq. (9.73)] is shown in Fig. 9.24b,

$$S_{\mathrm{n}_c}(f) = S_{\mathrm{n}_s}(f) = \begin{cases} \frac{\mathcal{N}}{2} & |f| \le B \\ 0 & |f| > B \end{cases} \tag{9.75}$$

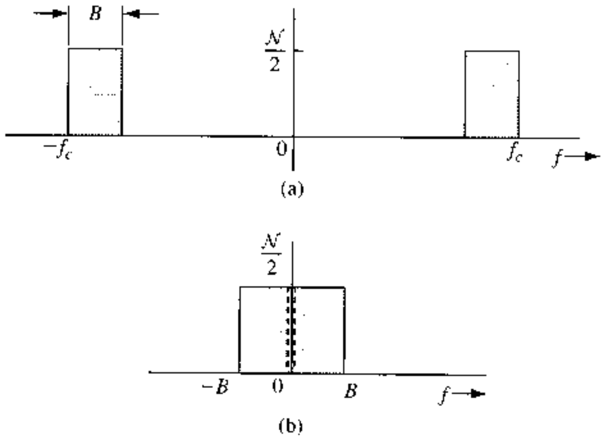


Figure 9.24 A possible form of quadrature component representation of noise in SSB.

From Fig. 9.24a it follows that

$$\overline{\mathrm{n}^2} = \mathcal{N}B \tag{9.76a}$$

Similarly, from Fig. 9.24b we have

$$\overline{\mathrm{n}_c^2} = \overline{\mathrm{n}_s^2} = \mathcal{N}B \tag{9.76b}$$

Hence,

$$\overline{\mathrm{n}_c^2} = \overline{\mathrm{n}_s^2} = \overline{\mathrm{n}^2} = \mathcal{N}B \tag{9.76c}$$

Bandpass "White" Gaussian Random Process

Thus far we have avoided defining a Gaussian random process. The Gaussian random process is perhaps the single most important random process in the area of communication. A careful and unhurried discussion, however, is beyond our scope. All we need to know here is that an RV $\mathrm{x}(t)$ formed by sample function amplitudes at instant t of a Gaussian process is Gaussian, with a PDF of the form of Eq. (8.39).

A Gaussian random process with a uniform PSD is called a white Gaussian random process. The term *bandpass "white" Gaussian process* is actually a misnomer. However, it is a popular notion to represent a random process $\mathrm{n}(t)$ with uniform PSD $\mathcal{N}/2$ centered at ω_c and with a bandwidth $2B$ (Fig. 9.22a). Utilizing the quadrature representation, it can be expressed as

$$\mathrm{n}(t) = \mathrm{n}_c(t) \cos \omega_c t + \mathrm{n}_s(t) \sin \omega_c t \tag{9.77}$$

where, from Eq. (9.71), we have

$$S_{\mathrm{n}_c}(f) = S_{\mathrm{n}_s}(f) = \begin{cases} \mathcal{N} & |f| \leq B \\ 0 & |f| > B \end{cases}$$

Also, from Eq. (9.72c),

$$\overline{\mathrm{n}_c^2} = \overline{\mathrm{n}_s^2} = \overline{\mathrm{n}^2} = 2\mathcal{N}B \tag{9.78}$$

The bandpass signal can also be expressed in polar form [see Eq. (3.40)]:

$$\mathrm{n}(t) = \mathrm{E}(t) \cos (\omega_c t + \Theta) \tag{9.79a}$$

where the random envelope and random phase are defined by

$$\mathrm{E}(t) = \sqrt{\mathrm{n}_c^2(t) + \mathrm{n}_s^2(t)} \tag{9.79b}$$

$$\Theta(t) = -\tan^{-1} \frac{\mathrm{n}_s(t)}{\mathrm{n}_c(t)} \tag{9.79c}$$

The RVs $\mathrm{n}_c(t)$ and $\mathrm{n}_s(t)$ are uncorrelated [see Eq. (9.68)] Gaussian RVs with zero means and variance $2\mathcal{N}B$ [Eq. (9.78)]. Hence, their PDFs are identical:

$$p_{\mathrm{n}_c}(\alpha) = p_{\mathrm{n}_s}(\alpha) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\alpha^2/2\sigma^2} \tag{9.80a}$$

where

$$\sigma^2 = 2\mathcal{N}B \tag{9.80b}$$

It has been shown in Prob. 8.2-10 that if two Gaussian RVs are uncorrelated, they are independent. In such a case, as shown in Example 8.17, $\mathrm{E}(t)$ has a Rayleigh density

$$p_{\mathrm{E}}(E) = \frac{E}{\sigma^2} e^{-E^2/2\sigma^2} u(E), \qquad \sigma^2 = 2\mathcal{N}B \tag{9.81}$$

and Θ in Eq. (9.79a) is uniformly distributed over $(0, 2\pi)$.

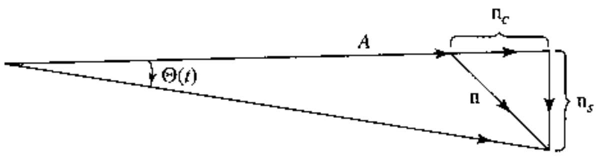


Figure 9.25 Phasor representation of a sinusoid and a narrowband Gaussian noise.

Sinusoidal Signal in Noise

Another case of interest is a sinusoid plus a narrowband Gaussian noise. If $A\cos(\omega_c t+\varphi)$ is a sinusoid mixed with $n(t)$, a Gaussian bandpass noise centered at ω_c, then the sum $y(t)$ is given by

$$y(t) = A\cos(\omega_c t+\varphi) + n(t)$$

Using Eq. (9.66) to represent the bandpass noise, we have

$$y(t) = [A + n_c(t)]\cos(\omega_c t+\varphi) + n_s(t)\sin(\omega_c t+\varphi) \tag{9.82a}$$

$$= E(t)\cos[\omega_c t + \Theta(t) + \varphi] \tag{9.82b}$$

where $E(t)$ is the envelope $[E(t) > 0]$ and $\Theta(t)$ is the angle shown in Fig. 9.25,

$$E(t) = \sqrt{[A + n_c(t)]^2 + n_s^2(t)} \tag{9.83a}$$

$$\Theta(t) = -\tan^{-1}\frac{n_s(t)}{A + n_c(t)} \tag{9.83b}$$

Both $n_c(t)$ and $n_s(t)$ are Gaussian, with variance σ^2. For white Gaussian noise, $\sigma^2 = 2\mathcal{N}B$ [Eq. (9.80b)]. Arguing in a manner analogous to that used in deriving Eq. (8.57), and observing that

$$\begin{aligned} n_c^2 + n_s^2 &= E^2 - A^2 - 2An_c \\ &= E^2 - 2A(A + n_c) + A^2 \\ &= E^2 - 2AE\cos\Theta(t) + A^2 \end{aligned}$$

we have

$$p_{E\Theta}(E,\theta) = \frac{E}{2\pi\sigma^2}e^{-(E^2-2AE\cos\theta+A^2)/2\sigma^2} \tag{9.84}$$

where σ^2 is the variance of n_c (or n_s) and is equal to $2\mathcal{N}B$ for white noise. From Eq. (9.84) we have

$$\begin{aligned} p_E(E) &= \int_{-\pi}^{\pi} p_{E\Theta}(E,\theta)\,d\theta \\ &= \frac{E}{\sigma^2}e^{-(E^2+A^2)/2\sigma^2}\left[\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{(AE/\sigma^2)\cos\theta}\,d\theta\right] \end{aligned} \tag{9.85}$$

The bracketed term on the right-hand side of Eq. (9.85) defines $I_0(AE/\sigma^2)$, where I_0 is the **modified zero-order Bessel function** of the first kind. Thus,

$$p_E(E) = \frac{E}{\sigma^2}e^{-(E^2+A^2)/2\sigma^2}I_0\left(\frac{AE}{\sigma^2}\right) \tag{9.86a}$$

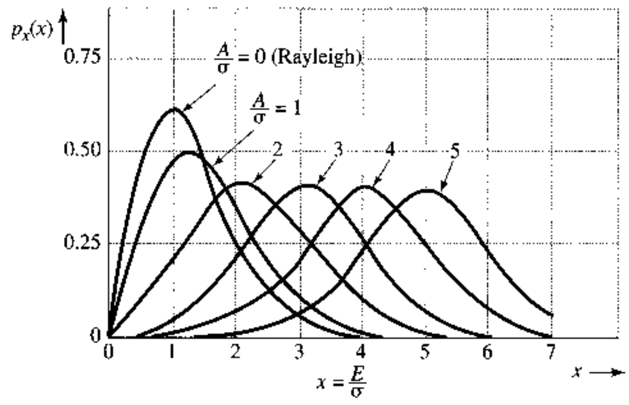


Figure 9.26
Rican PDF.

This is known as the **Rice density**, or **Ricean density**. For a large sinusoidal signal ($A \gg \sigma$), it can be shown that[4]

$$I_0\left(\frac{AE}{\sigma^2}\right) \simeq \sqrt{\frac{\sigma^2}{2\pi AE}}e^{AE/\sigma^2}$$

and

$$P_E(E) \simeq \sqrt{\frac{E}{2\pi A\sigma^2}}e^{-(E-A)^2/2\sigma^2} \tag{9.86b}$$

Because $A \gg \sigma$, $E \simeq A$, and $p_E(E)$ in Eq. (9.86b) is very nearly a Gaussian density with mean A and variance σ,

$$p_E(E) \simeq \frac{1}{\sigma\sqrt{2\pi}}e^{-(E-A)^2/2\sigma^2} \tag{9.86c}$$

Figure 9.26 shows the PDF of the normalized RV E/σ. Note that for $A/\sigma = 0$, we obtain the Rayleigh density.

From the joint PDF $p_{E\Theta}(E, \theta)$, we can also obtain $p_\Theta(\theta)$, the PDF of the phase Θ, by integrating the joint PDF with respect to E,

$$p_\Theta(\theta) = \int_0^\infty p_{E\Theta}(E, \theta)\, dE$$

Although the integration is straightforward, there are a number of involved steps, and for this reason it will not be repeated here. The final result is

$$p_\Theta(\theta) = \frac{1}{2\pi}e^{-A^2/2\sigma^2}\left\{1 + \frac{A}{\sigma}\sqrt{2\pi}\cos\theta e^{A^2\cos^2\theta/2\sigma^2}\left[1 - Q\left(\frac{A\cos\theta}{\sigma}\right)\right]\right\} \tag{9.86d}$$

REFERENCES

1. B. P. Lathi, *An Introduction to Random Signals and Communication Theory*, International Textbook Co., Scranton, PA, 1968.
2. J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering*, Wiley, New York, 1965.

3. A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 2nd ed., McGraw-Hill, New York, 1984.
4. S. O. Rice, "Mathematical Analysis of Random Noise," *Bell. Syst. Tech. J.*, vol. 23, pp. 282–332, July 1944; vol. 24, pp. 46–156, Jan. 1945.

PROBLEMS

9.1-1 **(a)** Sketch the ensemble of the random process

$$\mathrm{x}(t) = \mathrm{a} \cos(\omega_c t + \Theta)$$

where ω_c and Θ are constants and a is an RV uniformly distributed in the range $(-A, A)$.

(b) Just by observing the ensemble, determine whether this is a stationary or a nonstationary process. Give your reasons.

9.1-2 Repeat part **(a)** of Prob. 9.1-1 if a and Θ are constants but ω_c is an RV uniformly distributed in the range $(0, 100)$.

9.1-3 **(a)** Sketch the ensemble of the random process

$$\mathrm{x}(t) = \mathrm{a}t + b$$

where b is a constant and a is an RV uniformly distributed in the range $(-2, 2)$.

(b) Just by observing the ensemble, state whether this is a stationary or a nonstationary process.

9.1-4 Determine $\overline{\mathrm{x}(t)}$ and $R_{\mathrm{x}}(t_1, t_2)$ for the random process in Prob. 9.1-1, and determine whether this is a wide-sense stationary process.

9.1-5 Repeat Prob. 9.1-4 for the process $\mathrm{x}(t)$ in Prob. 9.1-2.

9.1-6 Repeat Prob. 9.1-4 for the process $\mathrm{x}(t)$ in Prob. 9.1-3.

9.1-7 Given a random process $\mathrm{x}(t) = \mathrm{k}t$, where k is an RV uniformly distributed in the range $(-1, 1)$.

(a) Sketch the ensemble of this process.

(b) Determine $\overline{\mathrm{x}(t)}$.

(c) Determine $R_{\mathrm{x}}(t_1, t_2)$.

(d) Is the process wide-sense stationary?

(e) Is the process ergodic?

(f) If the process is wide-sense stationary, what is its power P_s [that is, its mean square value $\overline{\mathrm{x}^2(t)}$?

9.1-8 Repeat Prob. 9.1-7 for the random process

$$\mathrm{x}(t) = \mathrm{a} \cos(\omega_c t + \Theta)$$

where ω_c is a constant and a and Θ are independent RVs uniformly distributed in the ranges $(-1, 1)$ and $(0, 2\pi)$, respectively.

9.2-1 For each of the following functions, state whether it can be a valid PSD of a real random process.

(a) $\dfrac{(2\pi f)^2}{(2\pi f)^2 + 16}$

(b) $\dfrac{1}{(2\pi f)^2 - 16}$

(c) $\dfrac{(2\pi f)}{(2\pi f)^2 + 16}$

(d) $\delta(2\pi f) + \dfrac{1}{(2\pi f)^2 + 16}$

(e) $\delta[2\pi(f + f_0)] - \delta[2\pi(f - f_0)]$

(f) $j[\delta(f + f_0) + \delta(f - f_0)]$

(g) $\dfrac{j(2\pi f)^2}{(2\pi f)^2 + 16}$

9.2-2 Show that for a wide-sense stationary process $x(t)$,

(a) $R_x(0) \geq |R_x(\tau)| \qquad \tau \neq 0$

Hint: $\overline{(x_1 \pm x_2)^2} = \overline{x_1^2} + \overline{x_2^2} \pm \overline{2x_1x_2} \geq 0$. Let $x_1 = x(t_1)$ and $x_2 = x(t_2)$.

(b) $\lim_{\tau \to \infty} R_x(\tau) = \bar{x}^2$

Hint: As $\tau \to \infty$, x_1 and x_2 tend to become independent.

9.2-3 Show that if the PSD of a random process $x(t)$ is band-limited, and if

$$R_x\left(\frac{n}{2B}\right) = \begin{cases} 1 & n = 0 \\ 0 & n = \pm 1, \pm 2, \pm 3, \ldots \end{cases}$$

then the minimum bandwidth process $x(t)$ that can exhibit this autocorrelation function is a white band-limited process; that is, $S_x(f) = k\,\Pi(f/2W)$.

Hint: Use the sampling theorem to reconstruct $R_x(\tau)$.

9.2-4 For the random binary process in Example 9.5 (Fig. 9.9a), determine $R_x(\tau)$ and $S_x(f)$ if the probability of transition (from 1 to −1 or vice versa) at each node is p instead of 0.5.

9.2-5 A wide-sense stationary white process $m(t)$ band-limited to B Hz is sampled at the Nyquist rate. Each sample is transmitted by a basic pulse $p(t)$ multiplied by the sample value. This is a PAM signal. Show that the PSD of the PAM signal is $2BR_m(0)|P(f)|^2$.

Hint: Use Eq. (9.31). Show that Nyquist samples a_k and a_{k+n} $(n \geq 1)$ are uncorrelated.

9.2-6 A duobinary line code proposed by Lender is a ternary scheme similar to bipolar that requires only half the bandwidth of the latter. In this code, **0** is transmitted by no pulse, and **1** is transmitted by pulse $p(t)$ or $-p(t)$ using the following rule: A **1** is encoded by the same pulse as that used to encode the preceding **1** if the two **1**s are separated by an even number of **0**s. It is encoded by the negative of the pulse used to encode the preceding **1** if the two **1**s are separated by an odd number of **0**s. Random binary digits are transmitted every T_b seconds. Assuming $P(\mathbf{0}) = P(\mathbf{1}) = 0.5$, show that

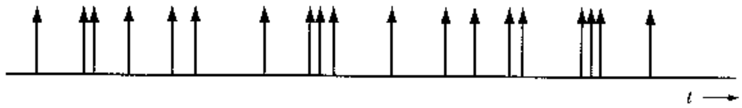
$$S_y(f) = \frac{|P(f)|^2}{T_b} \cos^2(\pi f T_b)$$

Find $S_y(f)$ if $p(t)$, the basic pulse used, is a half-width rectangular pulse $\Pi(2t/T_b)$.

9.2-7 Determine $S_y(f)$ for polar signaling if $P(\mathbf{1}) = Q$ and $P(\mathbf{0}) = 1 - Q$.

9.2-8 An impulse noise $x(t)$ can be modeled by a sequence of unit impulses located at random instants (Fig. P9.2-8). There are an average of α impulses per second, and the location of any impulse is independent of the locations of other impulses. Show that $R_x(\tau) = \alpha\,\delta(\tau) + \alpha^2$.

Figure P.9.2-8



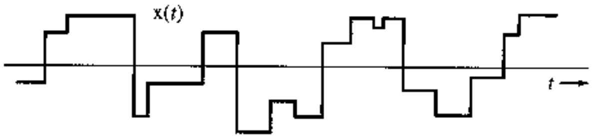
9.2-9 Repeat Prob. 9.2-8 if the impulses are equally likely to be positive and negative.

9.2-10 A sample function of a random process $x(t)$ is shown in Fig. P9.2-10. The signal $x(t)$ changes abruptly in amplitude at random instants. There are an average of β amplitude changes (or shifts) per second. The probability that there will be no amplitude shift in τ seconds is given by $P_0(\tau) = e^{-\beta\tau}$. The amplitude after a shift is independent of the amplitude before the shift. The amplitudes are randomly distributed, with a PDF $p_x(x)$. Show that

$$R_x(\tau) = \overline{x^2} e^{-\beta|\tau|} \quad \text{and} \quad S_x(f) = \frac{2\beta\overline{x^2}}{\beta^2 + (2\pi f)^2}$$

This process represents a model for thermal noise.[1]

Figure P.9.2-10



9.3-1 Show that for jointly wide-sense stationary, real, random processes $x(t)$ and $y(t)$,

$$|R_{xy}(\tau)| \leq [R_x(0)R_y(0)]^{1/2}$$

Hint: For any real number a, $\overline{(ax - y)^2} \geq 0$.

9.3-2 If $x(t)$ and $y(t)$ are two incoherent random processes, and two new processes $u(t)$ and $v(t)$ are formed as follows:

$$u(t) = 2x(t) - y(t) \qquad v(t) = x(t) + 3y(t)$$

find $R_u(\tau)$, $R_v(\tau)$, $R_{uv}(\tau)$, and $R_{vu}(\tau)$ in terms of $R_x(\tau)$ and $R_y(\tau)$.

9.3-3 Two random processes $x(t)$ and $y(t)$ are

$$x(t) = A\cos(\omega_0 t + \varphi) \quad \text{and} \quad y(t) = B\sin(n\omega_0 t + n\varphi + \psi)$$

where n = integer $\neq 0$ and A, B, ψ, and ω_0 are constants and φ is an RV uniformly distributed in the range $(0, 2\pi)$. Show that the two processes are incoherent.

9.3-4 A sample signal is a periodic random process $x(t)$ shown in Fig. P9.3-4. The initial delay b where the first pulse begins is an RV uniformly distributed in the range $(0, T_b)$.

(a) Show that the sample signal can be written as

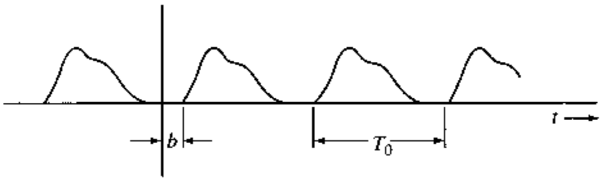
$$x(t) = C_0 + \sum_{n=1}^{\infty} C_n \cos[n\omega_0(t - b) + \theta_n]$$

by first finding its trigonometric Fourier series when b = 0.

(b) Show that

$$R_x(\tau) = C_0^2 + \frac{1}{2}\sum_{n=1}^{\infty} C_n^2 \cos n\omega_0\tau \qquad \omega_0 = \frac{2\pi}{T_0}$$

Figure P.9.3-4

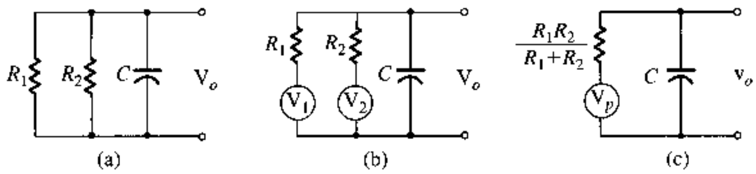


9.4-1 A simple *RC* circuit has two resistors R_1 and R_2 in parallel (Fig. P9.4-1a). Calculate the rms value of the thermal noise voltage v_o across the capacitor in two ways:

(a) Consider resistors R_1 and R_2 as two separate resistors, with respective thermal noise voltages of PSD $2kTR_1$ and $2kTR_2$ (Fig. P9.4-1b). Note that the two sources are independent.

(b) Consider the parallel combination of R_1 and R_2 as a single resistor of value $R_1R_2/(R_1+R_2)$, with its thermal-noise voltage source of PSD $2kTR_1R_2/(R_1+R_2)$ (Fig. P9.4-1c). Comment.

Figure P.9.4-1



9.4-2 Show that $R_{xy}(\tau)$, the cross-correlation function of the input process x(*t*) and the output process y(*t*) in Fig. 9.12, is

$$R_{xy}(\tau) = h(\tau) * R_x(\tau) \qquad \text{and} \qquad S_{xy}(f) = H(f)S_x(f)$$

Hence, show that for the thermal noise n(*t*) and the output $v_o(t)$ in Fig. 9.13 (Example 9.9),

$$S_{nv_o}(f) = \frac{2kTR}{1 + j2\pi fRC} \qquad \text{and} \qquad R_{nv_o}(\tau) = \frac{2kT}{C}e^{-\tau/RC}u(\tau)$$

9.4-3 A shot noise is similar to impulse noise described in Prob. 9.2-8 except that instead of random impulses, we have pulses of finite width. If we replace each impulse in Fig. P9.2-8 by a pulse $h(t)$ whose width is large in comparison to $1/\alpha$, so that there is a considerable overlapping of pulses, we get shot noise. The result of pulse overlapping is that the signal looks like a continuous random signal, as shown in Fig. P9.4-3.

(a) Derive the autocorrelation function and the PSD of such a random process.

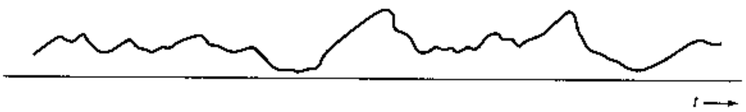
Hint: Shot noise results from passing impulse noise through a suitable filter. First derive the PSD of the shot noise and then obtain the autocorrelation function from the PSD. The answers will be in terms of α, $h(t)$, or $H(f)$.

(b) The shot noise in transistors can be modeled by

$$h(t) = \frac{q}{T} e^{-t/T} u(t)$$

where q is the charge on an electron and T is the electron transit time. Determine and sketch the autocorrelation function and the PSD of the transistor shot noise.

Figure P.9.4-3



9.6-1 A signal process m(t) is mixed with a channel noise n(t). The respective PSDs are

$$S_m(f) = \frac{6}{9 + (2\pi f)^2} \quad \text{and} \quad S_n(f) = 6$$

(a) Find the optimum Wiener-Hopf filter.

(b) Sketch its unit impulse response.

(c) Estimate the amount of delay necessary to make this filter closely realizable (causal).

(d) Compute the noise power at the input and the output of the filter.

9.6-2 Repeat Prob. 9.6-1 if

$$S_m(f) = \frac{4}{4 + (2\pi f)^2} \quad \text{and} \quad S_n(f) = \frac{32}{64 + (2\pi f)^2}$$

9.7-1 A message signal m(t) with

$$S_m(f) = \left[\frac{\alpha^2}{(2\pi f)^2 + \alpha^2}\right]^2 \qquad (\alpha = 3000\pi)$$

DSB-SC modulates a carrier of 100 kHz. Assume an ideal channel with $H_c(f) = 10^{-3}$] and the channel noise PSD $S_n(f) = 2 \times 10^{-9}$. The transmitted power is required to be 1 kW, and $G = 10^{-2}$].

(a) Determine transfer functions of optimum preemphasis and deemphasis filters.

(b) Determine the output signal power, the noise power, and the output SNR.

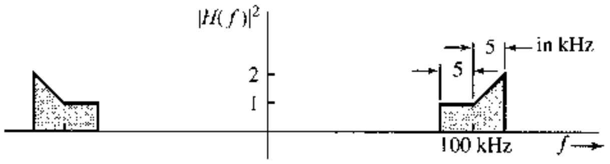
(c) Determine γ at the demodulator input.

9.7-2 Repeat Prob. 9.7-1 for the SSB (USB) case.

9.7-3 It was shown in the text that when the baseband m(t) is band-limited with a uniform PSD, PM and FM have identical performance from the SNR point of view. For such m(t), show that optimum PDE filters in angle modulation can improve the output SNR by a factor of 4/3 (or 1.3 dB) only. Find the optimum PDE filter transfer functions.

9.8-1 A white process of PSD $\mathcal{N}/2$ is transmitted through a bandpass filter $H(f)$ (Fig. P9.8-1). Represent the filter output $n(t)$ in terms of quadrature components, and determine $S_{n_c}(f)$, $S_{n_s}(f)$, $\overline{n_c^2}$, $\overline{n_s^2}$, and $\overline{n^2}$ when the center frequency used in this representation is 100 kHz (i.e., $f_c = 100 \times 10^3$).

Figure P.9.8-1



9.8-2 Repeat Prob. 9.8-1 if the center frequency f_c used in the representation is not a true center frequency. Consider three cases: **(a)** $f_c = 105$ kHz; **(b)** $f_c = 95$ kHz; **(c)** $f_c = 120$ kHz.

9.8-3 A random process $x(t)$ with the PSD shown in Fig. P9.8-3a is passed through a bandpass filter (Fig. P9.8-3b). Determine the PSDs and mean square values of the quadrature components of the output process. Assume the center frequency in the representation to be 0.5 MHz.

Figure P.9.8-3