CHAPTER 1

Basic Descriptions and Properties

COPYRIGHTED MATERIAL

This first chapter gives basic descriptions and properties of deterministic data and random data to provide a physical understanding for later material in this book. Simple classification ideas are used to explain differences between stationary random data, ergodic random data, and nonstationary random data. Fundamental statistical functions are defined by words alone for analyzing the amplitude, time, and frequency domain properties of single stationary random records and pairs of stationary random records. An introduction is presented on various types of input/output linear system problems solved in this book, as well as necessary error analysis criteria to design experiments and evaluate measurements.

1.1 DETERMINISTIC VERSUS RANDOM DATA

Any observed data representing a physical phenomenon can be broadly classified as being either deterministic or nondeterministic. Deterministic data are those that can be described by an explicit mathematical relationship. For example, consider a rigid body that is suspended from a fixed foundation by a linear spring, as shown in Figure 1.1. Let m be the mass of the body (assumed to be inelastic) and k be the spring constant of the spring (assumed to be massless). Suppose the body is displaced from its position of equilibrium by a distance X and released at time $t = 0$. From either basic laws of mechanics or repeated observations, it can be established that the following relationship will apply:

$$x(t) = X \cos\sqrt{\frac{k}{m}}\,t \quad t \geq 0 \tag{1.1}$$

Equation (1.1) defines the exact location of the body at any instant of time in the future. Hence, the physical data representing the motion of the mass are deterministic.

Random Data: Analysis and Measurement Procedures, Fourth Edition. By Julius S. Bendat and Allan G. Piersol
Copyright © 2010 John Wiley & Sons, Inc.

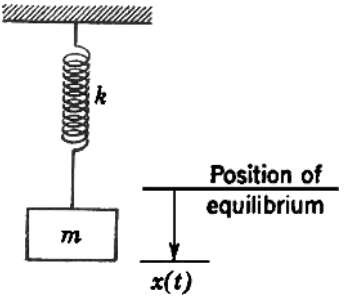

Figure 1.1 Simple spring mass system.

There are many physical phenomena in practice that produce data that can be represented with reasonable accuracy by explicit mathematical relationships. For example, the motion of a satellite in orbit about the earth, the potential across a condenser as it discharges through a resistor, the vibration response of an unbalanced rotating machine, and the temperature of water as heat is applied are all basically deterministic. However, there are many other physical phenomena that produce data that are not deterministic. For example, the height of waves in a confused sea, the acoustic pressures generated by air rushing through a pipe, and the electrical output of a noise generator represent data that cannot be described by explicit mathematical relationships. There is no way to predict an exact value at a future instant of time. These data are random in character and must be described in terms of probability statements and statistical averages rather than by explicit equations.

The classification of various physical data as being either deterministic or random might be debated in many cases. For example, it might be argued that no physical data in practice can be truly deterministic because there is always a possibility that some unforeseen event in the future might influence the phenomenon producing the data in a manner that was not originally considered. On the other hand, it might be argued that no physical data are truly random, because an exact mathematical description might be possible if a sufficient knowledge of the basic mechanisms of the phenomenon producing the data were available. In practical terms, the decision of whether physical data are deterministic or random is usually based on the ability to reproduce the data by controlled experiments. If an experiment producing specific data of interest can be repeated many times with identical results (within the limits of experimental error), then the data can generally be considered deterministic. If an experiment cannot be designed that will produce identical results when the experiment is repeated, then the data must usually be considered random in nature.

Various special classifications of deterministic and random data will now be discussed. Note that the classifications are selected from an analysis viewpoint and do not necessarily represent the most suitable classifications from other possible viewpoints. Further note that physical data are usually thought of as being functions of time and will be discussed in such terms for convenience. Any other variable, however, can replace time, as required.

1.2 CLASSIFICATIONS OF DETERMINISTIC DATA

Data representing deterministic phenomena can be categorized as being either periodic or nonperiodic. Periodic data can be further categorized as being either sinusoidal or complex periodic. Nonperiodic data can be further categorized as being either "almost-periodic" or transient. These various classifications of deterministic data are schematically illustrated in Figure 1.2. Of course, any combination of these forms may also occur. For purposes of review, each of these types of deterministic data, along with physical examples, will be briefly discussed.

1.2.1 Sinusoidal Periodic Data

Sinusoidal data are those types of periodic data that can be defined mathematically by a time-varying function of the form

$$x(t) = X \sin(2\pi f_0 t + \theta) \tag{1.2}$$

where

$X =$ amplitude
$f_0 =$ cyclic frequency in cycles per unit time
$\theta =$ initial phase angle with respect to the time origin in radians
$x(t) =$ instantaneous value at time t

The sinusoidal time history described by Equation (1.2) is usually referred to as a sine wave. When analyzing sinusoidal data in practice, the phase angle θ is often ignored. For this case,

$$x(t) = X \sin 2\pi f_0 t \tag{1.3}$$

Equation (1.3) can be pictured by a time history plot or by an amplitude–frequency plot (frequency spectrum), as illustrated in Figure 1.3.

The time interval required for one full fluctuation or cycle of sinusoidal data is called the period T_p. The number of cycles per unit time is called the frequency f_0.

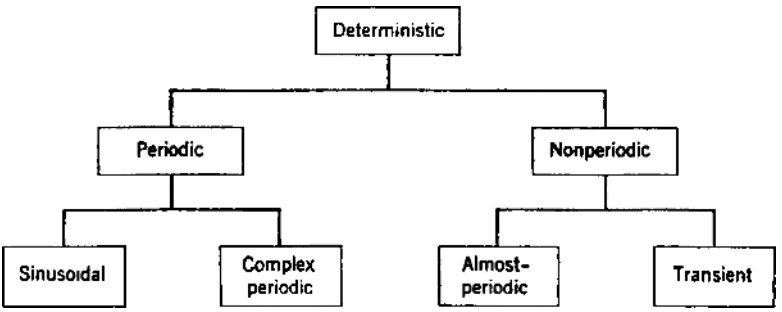

Figure 1.2 Classification of deterministic data.

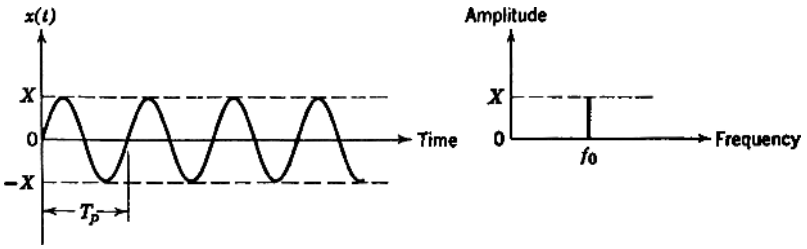

Figure 1.3 Time history and spectrum of sinusoidal data.

The frequency and period are related by

$$T_p = \frac{1}{f_0} \tag{1.4}$$

Note that the frequency spectrum in Figure 1.3 is composed of an amplitude component at a specific frequency, as opposed to a continuous plot of amplitude versus frequency. Such spectra are called *discrete spectra* or *line spectra.*

There are many examples of physical phenomena that produce approximately sinusoidal data in practice. The voltage output of an electrical alternator is one example; the vibratory motion of an unbalanced rotating weight is another. Sinusoidal data represent one of the simplest forms of time-varying data from the analysis viewpoint.

1.2.2 Complex Periodic Data

Complex periodic data are those types of periodic data that can be defined mathematically by a time-varying function whose waveform exactly repeats itself at regular intervals such that

$$x(t) = x(t \pm nT_p) \quad n = 1, 2, 3, \ldots \tag{1.5}$$

As for sinusoidal data, the time interval required for one full fluctuation is called the *period* T_p. The number of cycles per unit time is called the *fundamental frequency* f_1. A special case for complex periodic data is clearly sinusoidal data, where $f_1 = f_0$.

With few exceptions in practice, complex periodic data may be expanded into a Fourier series according to the following formula:

$$x(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos 2\pi n f_1 t + b_n \sin 2\pi n f_1 t) \tag{1.6}$$

where

$$f_1 = \frac{1}{T_p}$$

$$a_n = \frac{2}{T_p} \int_0^{T_p} x(t) \cos 2\pi n f_1 t \, dt \quad n = 0, 1, 2, \ldots$$

$$b_n = \frac{2}{T_p} \int_0^{T_p} x(t) \sin 2\pi n f_1 t \, dt \quad n = 1, 2, 3, \ldots$$

An alternative way to express the Fourier series for complex periodic data is

$$x(t) = X_0 + \sum_{n=1}^{\infty} X_n \cos(2\pi n f_1 t - \theta_n) \tag{1.7}$$

where

$$\begin{aligned} X_0 &= a_0/2 \\ X_n &= \sqrt{a_n^2 + b_n^2} \qquad n = 1, 2, 3, \ldots \\ \theta_n &= \tan^{-1}(b_n/a_n) \quad n = 1, 2, 3, \ldots \end{aligned}$$

In words, Equation (1.7) says that complex periodic data consist of a static component X_0 and an infinite number of sinusoidal components called harmonics, which have amplitudes X_n and phases θ_n. The frequencies of the harmonic components are all integral multiples of f_1.

When analyzing periodic data in practice, the phase angles θ_n are often ignored. For this case, Equation (1.7) can be characterized by a discrete spectrum, as illustrated in Figure 1.4. Sometimes, complex periodic data will include only a few components. In other cases, the fundamental component may be absent. For example, suppose a periodic time history is formed by mixing three sine waves that have frequencies of 60, 75, and 100 Hz. The highest common divisor is 5 Hz, so the period of the resulting periodic data is $T_p = 0.2$ s. Hence, when expanded into a Fourier series, all values of X_n are zero except for $n = 12$, $n = 15$, and $n = 20$.

Physical phenomena that produce complex periodic data are far more common than those that produce simple sinusoidal data. In fact, the classification of data as being sinusoidal is often only an approximation for data that are actually complex. For example, the voltage output from an electrical alternator may actually display, under careful inspection, some small contributions at higher harmonic frequencies. In other cases, intense harmonic components may be present in periodic physical data. For example, the vibration response of a multicyclinder reciprocating engine will usually display considerable harmonic content.

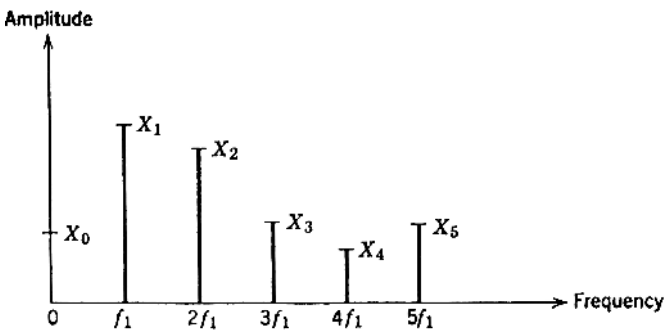

Figure 1.4 Spectrum of complex periodic data.

1.2.3 Almost-Periodic Data

In Section 1.2.2, it is noted that periodic data can generally be reduced to a series of sine waves with commensurately related frequencies. Conversely, the data formed by summing two or more commensurately related sine waves will be periodic. However, the data formed by summing two or more sine waves with arbitrary frequencies generally will not be periodic. Specifically, the sum of two or more sine waves will be periodic only when the ratios of all possible pairs of frequencies form rational numbers. This indicates that a fundamental period exists that will satisfy the requirements of Equation (1.5). Hence,

$$x(t) = X_1 \sin(2t + \theta_1) + X_2 \sin(3t + \theta_2) + X_3 \sin(7t + \theta_3)$$

is periodic because $\frac{2}{3}$, $\frac{2}{7}$, and $\frac{3}{7}$ are rational numbers (the fundamental period is $T_p = 1$). On the other hand,

$$x(t) = X_1 \sin(2t + \theta_1) + X_2 \sin(3t + \theta_2) + X_3 \sin(\sqrt{50}t + \theta_3)$$

is not periodic because $2/\sqrt{50}$ and $3/\sqrt{50}$ are not rational numbers (the fundamental period is infinitely long). The resulting time history in this case will have an almost-periodic character, but the requirements of Equation (1.5) will not be satisfied for any finite value of T_p.

Based on these discussions, almost-periodic data are those types of nonperiodic data that can be defined mathematically by a time-varying function of the form

$$x(t) = \sum_{n=1}^{\infty} X_n \sin(2\pi f_n t + \theta_n) \tag{1.8}$$

where $f_n/f_m \neq$ rational number in all cases. Physical phenomena producing almost-periodic data frequently occur in practice when the effects of two or more unrelated periodic phenomena are mixed. A good example is the vibration response in a multiple-engine propeller airplane when the engines are out of synchronization.

An important property of almost-periodic data is as follows. If the phase angles θ_n are ignored, Equation (1.8) can be characterized by a discrete frequency spectrum similar to that for complex periodic data. The only difference is that the frequencies of the components are not related by rational numbers, as illustrated in Figure 1.5.

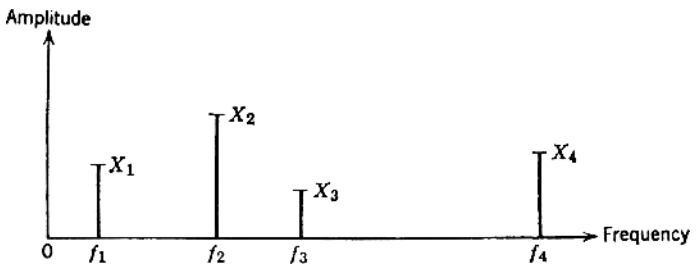

Figure 1.5 Spectrum of almost-periodic data.

$$x(t) = \begin{cases} A\,e^{-at} & t \geq 0 \\ 0 & t < 0 \end{cases}$$

$$x(t) = \begin{cases} A\,e^{-at}\cos bt & t \geq 0 \\ 0 & t < 0 \end{cases}$$

$$x(t) = \begin{cases} A & c \geq t \geq 0 \\ 0 & c < t < 0 \end{cases}$$

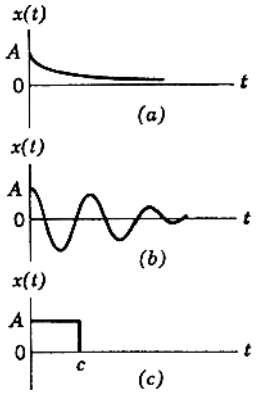

Figure 1.6 Illustrations of transient data.

1.2.4 Transient Nonperiodic Data

Transient data are defined as all nonperiodic data other than the almost-periodic data discussed in Section 1.2.3. In other words, transient data include all data not previously discussed that can be described by some suitable time-varying function. Three simple examples of transient data are given in Figure 1.6.

Physical phenomena that produce transient data are numerous and diverse. For example, the data in Figure 1.6(*a*) could represent the temperature of water in a kettle (relative to room temperature) after the flame is turned off. The data in Figure 1.6(*b*) might represent the free vibration of a damped mechanical system after an excitation force is removed. The data in Figure 1.6(*c*) could represent the stress in an end-loaded cable that breaks at time *c*.

An important characteristic of transient data, as opposed to periodic and almost-periodic data, is that a discrete spectral representation is not possible A *continuous* spectral representation for transient data can be obtained in most cases, however, from a Fourier transform given by

$$X(f) = \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft}dt \tag{1.9}$$

The Fourier transform $X(f)$ is generally a complex number that can be expressed in complex polar notation as

$$X(f) = |X(f)|e^{-j\theta(f)}$$

Here, $|X(f)|$ is the magnitude of $X(f)$ and $\theta(f)$ is the argument. In terms of the magnitude $|X(f)|$, continuous spectra of the three transient time histories in Figure 1.6 are as presented in Figure 1.7. Modern procedures for the digital computation of Fourier series and finite Fourier transforms are detailed in Chapter 11.

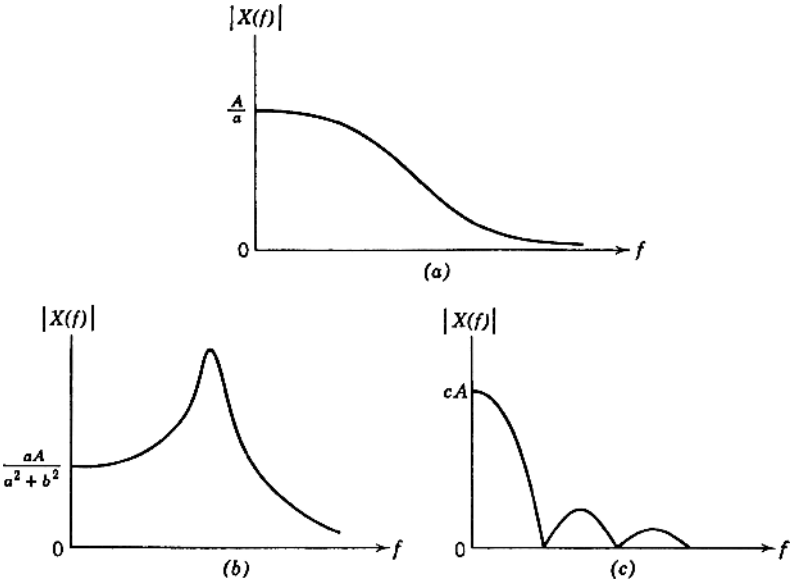

Figure 1.7 Spectra of transient data.

1.3 CLASSIFICATIONS OF RANDOM DATA

As discussed earlier, data representing a random physical phenomenon cannot be described by an explicit mathematical relationship because each observation of the phenomenon will be unique. In other words, any given observation will represent only one of many possible results that might have occurred. For example, assume the output voltage from a thermal noise generator is recorded as a function of time. A specific voltage time history record will be obtained, as shown in Figure 1.8. If a second thermal noise generator of identical construction and assembly is operated simultaneously, however, a different voltage time history record would result. In fact, every thermal noise generator that might be constructed would produce a different voltage time history record, as illustrated in Figure 1.8. Hence, the voltage time history for any one generator is merely one example of an infinitely large number of time histories that might have occurred.

A single time history representing a random phenomenon is called a *sample function* (or a *sample record* when observed over a finite time interval). The collection of all possible sample functions that the random phenomenon might have produced is called a *random process* or a *stochastic process.* Hence, a sample record of data for a random physical phenomenon may be thought of as one physical realization of a random process.

Random processes may be categorized as being either stationary or nonstationary. Stationary random processes may be further categorized as being either ergodic or nonergodic. Nonstationary random processes may be further categorized in terms of

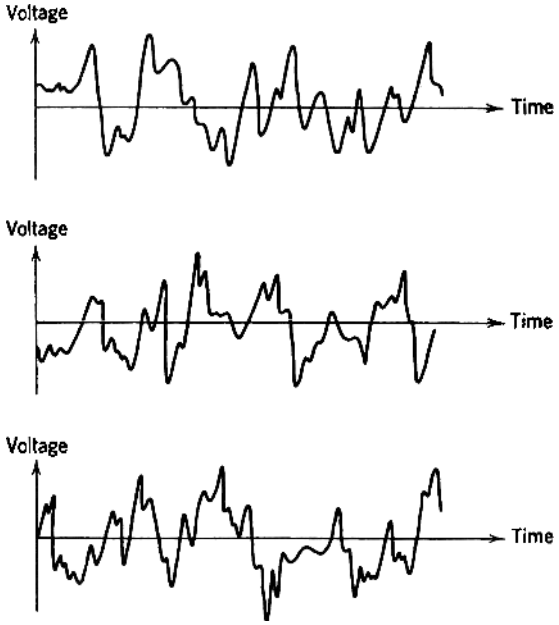

Figure 1.8 Sample records of thermal noise generator outputs.

specific types of nonstationary properties. These various classifications of random processes are schematically illustrated in Figure 1.9. The meaning and physical significance of these various types of random processes will now be discussed in broad terms. More analytical definitions and developments are presented in Chapters 5 and 12.

1.3.1 Stationary Random Data

When a physical phenomenon is considered in terms of a random process, the properties of the phenomenon can hypothetically be described at any instant of time by computing

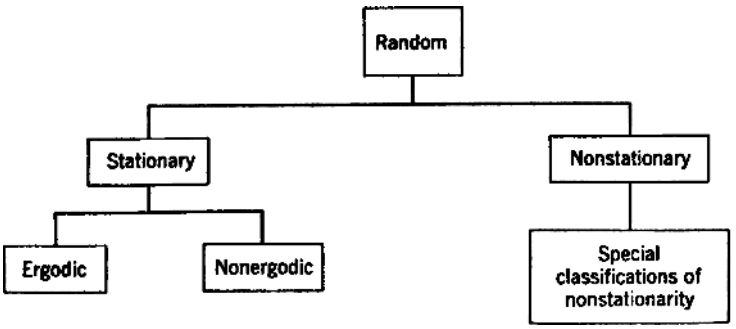

Figure 1.9 Classifications of random data.

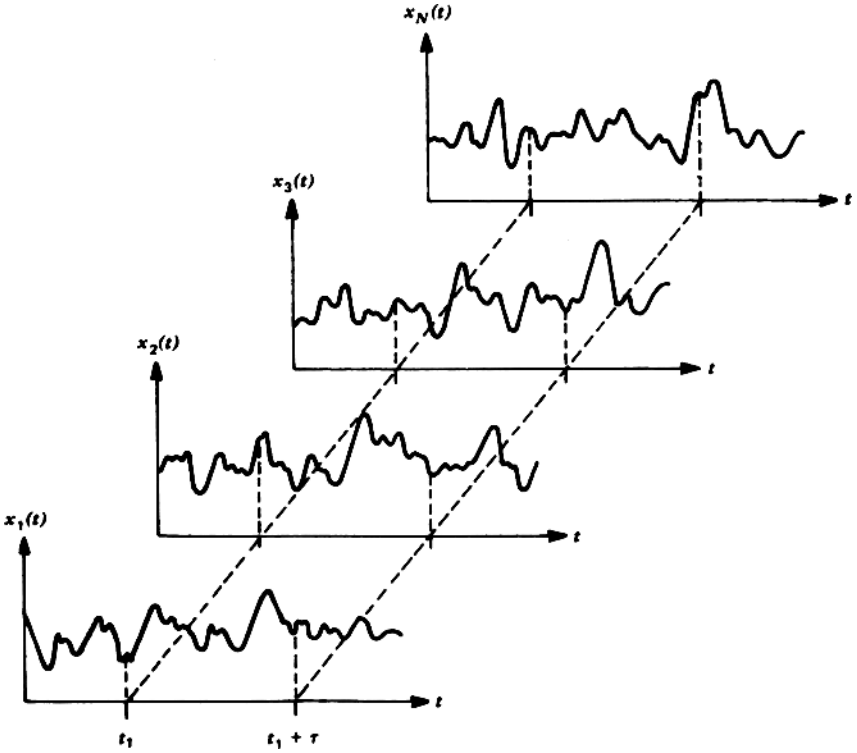

Figure 1.10 Ensemble of time history records defining a random process.

average values over the collection of sample functions that describe the random process. For example, consider the collection of sample functions (also called the *ensemble*) that forms the random process illustrated in Figure 1.10. The *mean value* (first moment) of the random process at some t_1 can be computed by taking the instantaneous value of each sample function of the ensemble at time t_1, summing the values, and dividing by the number of sample functions. In a similar manner, a correlation (joint moment) between the values of the random process at two different times (called the *autocorrelation function*) can be computed by taking the ensemble average of the product of instantaneous values at two times, t_1 and $t_1 + \tau$. That is, for the random process $\{x(t)\}$, where the symbol $\{\}$ is used to denote an ensemble of sample functions, the mean value $\mu_x(t_1)$ and the autocorrelation function $R_{xx}(t_1, t_1 + \tau)$ are given by

$$\mu_x(t_1) = \lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} x_k(t_1) \tag{1.10a}$$

$$R_{xx}(t_1, t_1 + \tau) = \lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} x_k(t_1) x_k(t_1 + \tau) \tag{1.10b}$$

where the final summation assumes that each sample function is equally likely.

For the general case where $\mu_x(t_1)$ and $R_{xx}(t_1, t_1 + \tau)$ defined in Equation (1.10) vary as time t_1 varies, the random process $\{x(t)\}$ is said to be *nonstationary.* For the special case where $\mu_x(t_1)$ and $R_{xx}(t_1, t_1 + \tau)$ do not vary as time t_1 varies, the random process $\{x(t)\}$ is said to be *weakly stationary* or stationary in the wide sense. For weakly stationary random processes, the mean value is a constant and the autocorrelation function is dependent only on the time displacement τ. That is, $\mu_x(t_1) = \mu_x$ and $R_{xx}(t_1, t_1 + \tau) = R_{xx}(\tau)$.

An infinite collection of higher order moments and joint moments of the random process $\{x(t)\}$ could also be computed to establish a complete family of probability distribution functions describing the process. For the special case where all possible moments and joint moments are time invariant, the random process $\{x(t)\}$ is said to be *strongly stationary* or stationary in the strict sense. For many practical applications, verification of weak stationarity will justify an assumption of strong stationarity.

1.3.2 Ergodic Random Data

In Section 1.3.1, it is noted how the properties of a random process can be determined by computing ensemble averages at specific instants of time. In most cases, however, it is also possible to describe the properties of a stationary random process by computing time averages over specific sample functions in the ensemble. For example, consider the kth sample function of the random process illustrated in Figure 1.10. The mean value $\mu_x(k)$ and the autocorrelation function $R_{xx}(\tau, k)$ of the kth sample function are given by

$$\mu_x(k) = \lim_{T \to \infty} \frac{1}{T} \int_0^T x_k(t)\, dt \tag{1.11a}$$

$$R_{xx}(\tau, k) = \lim_{T \to \infty} \frac{1}{T} \int_0^T x_k(t) x_k(t + \tau)\, dt \tag{1.11b}$$

If the random process $\{x(t)\}$ is stationary, and $\mu_x(k)$ and $R_{xx}(\tau, k)$ defined in Equation (1.11) do not differ when computed over different sample functions, the random process is said to be *ergodic.* For ergodic random processes, the time-averaged mean value and autocorrelation function (as well as all other time-averaged properties) are equal to the corresponding ensemble-averaged values. That is, $\mu_x(k) = \mu_x$ and $R_{xx}(\tau, k) = R_{xx}(\tau)$. Note that only stationary random processes can be ergodic.

Ergodic random processes are clearly an important class of random processes since all properties of ergodic random processes can be determined by performing time averages over a single sample function. Fortunately, in practice, random data representing stationary physical phenomena are generally ergodic. It is for this reason that the properties of stationary random phenomena can be measured properly, in most cases, from a single observed time history record. A full development of the properties of ergodic random processes is presented in Chapter 5.

1.3.3 Nonstationary Random Data

Nonstationary random processes include all random processes that do not meet the requirements for stationary defined in Section 1.3.1. Unless further restrictions are imposed, the properties of a nonstationary random process are generally time-varying functions that can be determined only by performing instantaneous averages over the ensemble of sample functions forming the process. In practice, it is often not feasible to obtain a sufficient number of sample records to permit the accurate measurement of properties by ensemble averaging. This fact has tended to impede the development of practical techniques for measuring and analyzing nonstationary random data.

In many cases, the nonstationary random data produced by actual physical phenomena can be classified into special categories of nonstationarity that simplify the measurement and analysis problem. For example, some types of random data might be described by a nonstationary random process $\{x(t)\}$, where each sample function is given by $x(t) = a(t)u(t)$. Here, $u(t)$ is a sample function from a stationary random process $\{u(t)\}$ and $a(t)$ is a deterministic multiplication factor. In other words, the data might be represented by a nonstationary random process consisting of sample functions with a common deterministic time trend. If nonstationary random data fit a specific model of this type, ensemble averaging is not always needed to describe the data. The various desired properties can sometimes be estimated from a single sample record, as is true for ergodic stationary data. These matters are discussed in detail in Chapter 12.

1.3.4 Stationary Sample Records

The concept of stationarity, as defined and discussed in Section 1.3.1, relates to the ensemble-averaged properties of a random process. In practice, however, data in the form of individual time history records of a random phenomenon are frequently referred to as being stationary or nonstationary. A slightly different interpretation of stationarity is involved here. When a single time history record is referred to as being stationary, it is generally meant that the properties computed over short time intervals do not vary significantly from one interval to the next. The word *significantly* is used here to mean that observed variations are greater than would be expected due to normal statistical sampling variations.

To help clarify this point, consider a single sample record $x_k(t)$ obtained from the kth sample function of a random process $\{x(t)\}$. Assume a mean value and an autocorrelation function are obtained by time averaging over a short interval T with a starting time of t_1 as follows:

$$\mu_x(t_1, k) = \frac{1}{T}\int_{t_1}^{t_1+T} x_k(t)\, dt \tag{1.12a}$$

$$R_{xx}(t_1, t_1+\tau, k) = \frac{1}{T}\int_{t_1}^{t_1+T} x_k(t)x_k(t+\tau)\, dr \tag{1.12b}$$

For the general case where the sample properties defined in Equation (1.12) vary significantly as the starting time t_1 varies, the individual sample record is said to be nonstationary. For the special case where the sample properties defined in Equation (1.12) do not vary significantly as the starting time t_1 varies, the sample record is said to be stationary. Note that a sample record obtained from an ergodic random process will be stationary. Furthermore, sample records from most physically interesting nonstationary random processes will be nonstationary. Hence, if an ergodic assumption is justified (as it is for most actual stationary physical phenomena), verification of stationarity for a single sample record will effectively justify an assumption of stationarity and ergodicity for the random process from which the sample record is obtained. Tests for stationarity of individual sample records are discussed in Chapters 4 and 10.

1.4 ANALYSIS OF RANDOM DATA

The analysis of random data involves different considerations from the deterministic data discussed in Section 1.2. In particular, because no explicit mathematical equation can be written for the time histories produced by a random phenomenon, statistical procedure must be used to define the descriptive properties of the data. Nevertheless, well-defined input/output relations exist for random data, which are fundamental to a wide range of applications. In such applications, however, an understanding and control of the statistical errors associated with the computed data properties and input/output relationships is essential.

1.4.1 Basic Descriptive Properties

Basic statistical properties of importance for describing single stationary random records are

1. Mean and mean square values
2. Probability density functions
3. Autocorrelation functions
4. Autospectral density functions

For the present discussion, it is instructive to define these quantities by words alone, without the use of mathematical equations. After this has been done, they will be illustrated for special cases of interest.

The mean value μ_x and the variance σ_x^2 for a stationary record represent the central tendency and dispersion, respectively, of the data. The mean square value ψ_x^2, which equals the variance plus the square of the mean, constitutes a measure of the combined central tendency and dispersion. The mean value is estimated by simply computing the average of all data values in the record. The mean square value is similarly estimated by computing the average of the squared data values. By first subtracting the

mean value estimate from all the data values, the mean square value computation yields a variance estimate.

The probability density function $p(x)$ for a stationary record represents the rate of change of probability with data value. The function $p(x)$ is generally estimated by computing the probability that the instantaneous value of the single record will be in a particular narrow amplitude range centered at various data values, and then dividing by the amplitude range.The total area under the probability density function over all data values will be unity because this merely indicates the certainty of the fact that the data values must fall between $-\infty$ and $+\infty$. The partial area under the probability density function from $-\infty$ to some given value x represents the probability distribution function, denoted by $P(x)$. The area under the probability density function between any two values x_1 and x_2, given by $P(x_2) - P(x_1)$, defines the probability that any future data values at a randomly selected time will fall within this amplitude interval. Probability density and distribution functions are fully discussed in Chapters 3 and 4.

The autocorrelation function $R_{xx}(\tau)$ for a stationary record is a measure of time-related properties in the data that are separated by fixed time delays. It can be estimated by delaying the record relative to itself by some fixed time delay τ, then multiplying the original record with the delayed record, and finally averaging the resulting product values over the available record length or over some desired portion of this record length. The procedure is repeated for all time delays of interest.

The autospectral (also called *power* spectral) density function $G_{xx}(f)$ for a stationary record represents the rate of change of mean square value with frequency. It is estimated by computing the mean square value in a narrow frequency band at various center frequencies, and then dividing by the frequency band. The total area under the autospectral density function over all frequencies will be the total mean square value of the record. The partial area under the autospectral density function from f_1 to f_2 represents the mean square value of the record associated with that frequency range. Autocorrelation and autospectral density functions are developed in Chapter 5.

Four typical time histories of a sine wave, sine wave plus random noise, narrow bandwidth random noise, and wide bandwidth random noise are shown in Figure 1.11. Theoretical plots of their probability density functions, autocorrelation functions, and autospectral density functions are shown in Figures 1.12, 1.13, and 1.14, respectively. Equations for all of these plots are given in Chapter 5, together with other theoretical formulas.

For pairs of random records from two different stationary random processes, joint statistical properties of importance are

1. Joint probability density functions
2. Cross-correlation functions
3. Cross-spectral density functions
4. Frequency response functions
5. Coherence functions

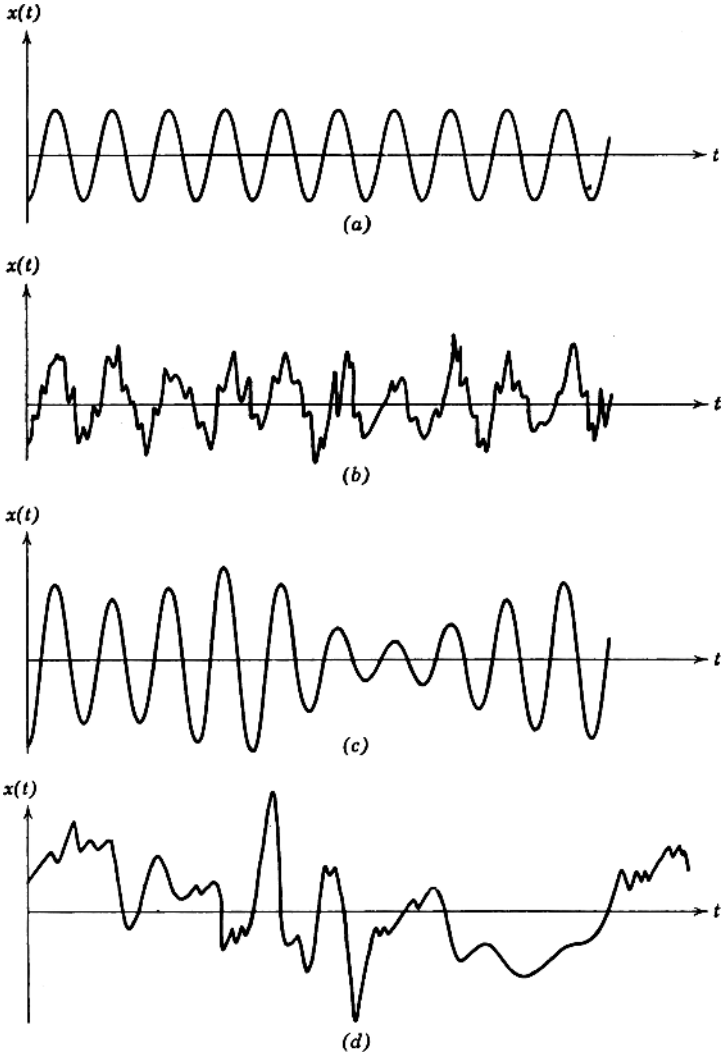

Figure 1.11 Four special time histories. (*a*) Sine wave. (*b*) Sine wave plus random noise. (*c*) Narrow bandwidth random noise. (*d*) Wide bandwidth random noise.

The first three functions measure fundamental properties shared by the pair of records in the amplitude, time, or frequency domains. From knowledge of the cross-spectral density function between the pair of records, as well as their individual autospectral density functions, one can compute theoretical linear frequency response functions (gain factors and phase factors) between the two records. Here, the two records are treated as a single-input/single-output problem. The coherence function is a measure of the accuracy of the assumed linear input/output model and can also be computed from the measured autospectral and cross-spectral density functions. Detailed discussions of these topics appear in Chapters 5, 6, and 7.

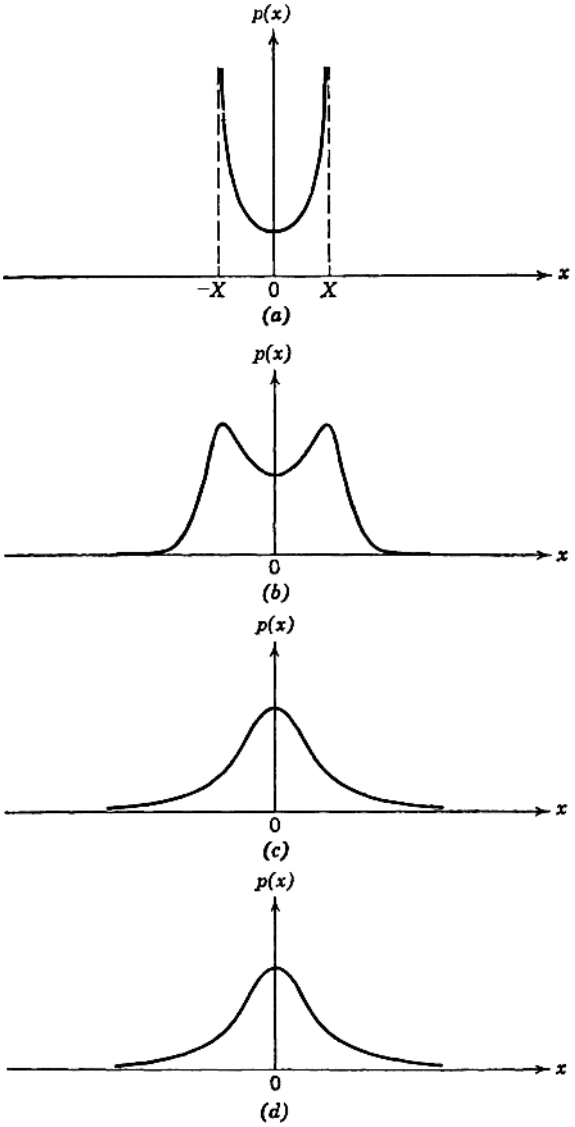

Figure 1.12 Probability density function plots. (*a*) Sine wave. (*b*) Sine wave plus random noise. (*c*) Narrow bandwidth random noise. (*d*) Wide bandwidth random noise.

Common applications of probability density and distribution functions, beyond a basic probabilistic description of data values, include

1. Evaluation of normality
2. Detection of data acquisition errors
3. Indication of nonlinear effects
4. Analysis of extreme values

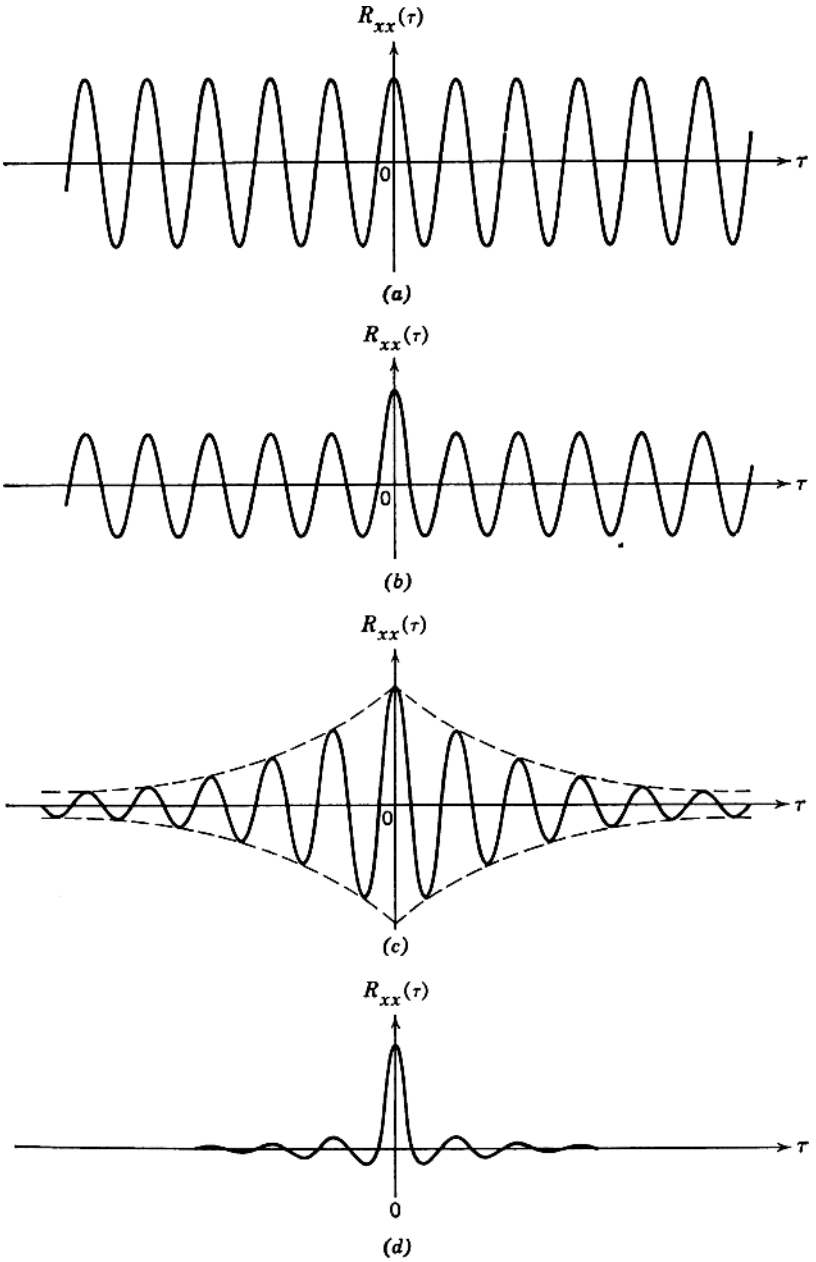

Figure 1.13 Autocorrelation function plots. (*a*) Sine wave. (*b*) Sine wave plus random noise. (*c*) Narrow bandwidth random noise. (*d*) Wide bandwidth random noise.

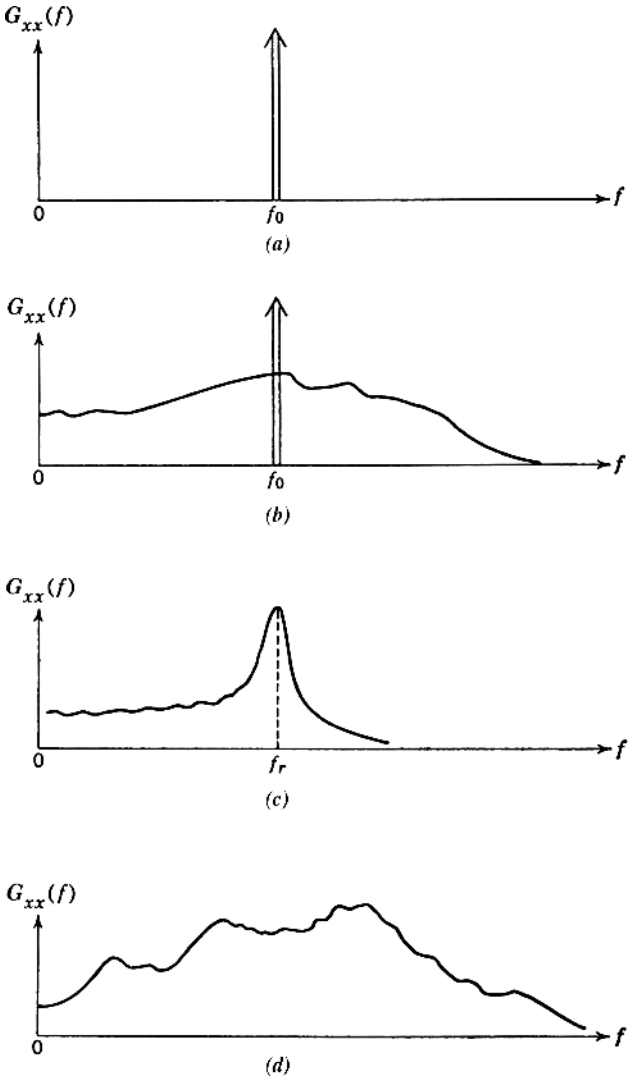

Figure 1.14 Autospectral density function plots. (*a*) Sine wave. (*b*) Sine wave plus random noise. (*c*) Narrow bandwidth random noise. (*d*) Wide bandwidth random noise.

The primary applications of correlation measurements include

1. Detection of periodicities
2. Prediction of signals in noise
3. Measurement of time delays
4. Location of disturbing sources
5. Identification of propagation paths and velocities

Typical applications of spectral density functions include

1. Determination of system properties from input data and output data
2. Prediction of output data from input data and system properties
3. Identification of input data from output data and system properties
4. Specifications of dynamic data for test programs
5. Identification of energy and noise sources
6. Optimum linear prediction and filtering

1.4.2 Input/Output Relations

Input/output cases of common interest can usually be considered as combinations of one or more of the following linear system models:

1. Single-input/single-output model
2. Single-input/multiple-output model
3. Multiple-input/single-output model
4. Multiple-input/multiple-output model

In all cases, there may be one or more parallel transmission paths with different time delays between each input point and output point. For multiple-input cases, the various inputs may or may not be correlated with each other. Special analysis techniques are required when nonstationary data are involved, as treated in Chapter 12, or when systems are nonlinear, as treated in Chapter 14.

A simple single-input/single-output model is shown in Figure 1.15. Here, $x(t)$ and $y(t)$ are the measured input and output stationary random records, and $n(t)$ is the unmeasured extraneous output noise. The quantity $H_{xy}(f)$ is the frequency response function of a constant-parameter linear system between $x(t)$ and $y(t)$. Figure 1.16 shows a single-input/multiple-output model that is a simple extension of Figure 1.15, where an input $x(t)$ produces many outputs $y_i(t)$, $i = 1, 2, 3, \ldots$. Any output $y_i(t)$ is the result of $x(t)$ passing through a constant-parameter linear system described by the frequency response function $H_{xi}(f)$. The noise terms $n_i(t)$ represent unmeasured extraneous output noise at the different outputs. It is clear that Figure 1.16 can be considered as a combination of separate single-input/single-output models.

Appropriate procedures for solving single-input models are developed in Chapter 6 using measured autospectral and cross-spectral density functions. Ordinary coher-

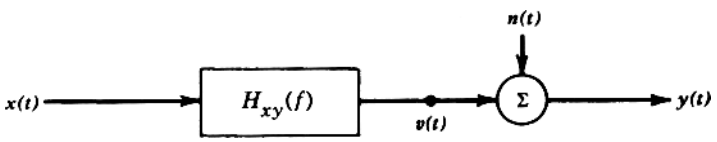

Figure 1.15 Single-input/single-output system with output noise.

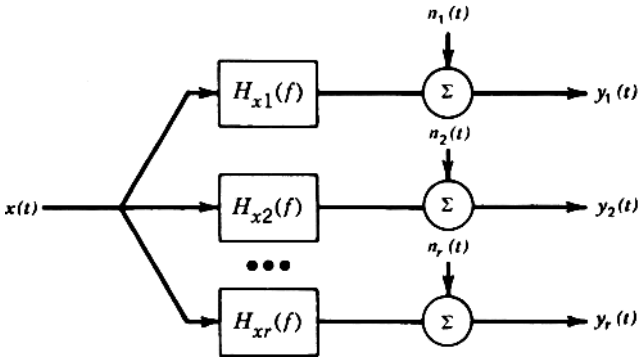

Figure 1.16 Single-input/multiple-output system with output noise.

ence functions are defined, which play a key role in both system-identification and source-identification problems. To determine both the gain factor and the phase factor of a desired frequency response function, it is always necessary to measure the cross-spectral density function between the input and output points. A good estimate of the gain factor alone can be obtained from measurements of the input and output autospectral density functions only if there is negligible input and output extraneous noise.

For a well-defined single-input/single-output model where the data are stationary, the system is linear and has constant parameters, and there is no extraneous noise at either the input or output point, the ordinary coherence function will be identically unity for all frequencies. Any deviation from these ideal conditions will cause the coherence function to be less than unity. In practice, measured coherence functions will often be less than unity and are important in determining the statistical confidence in frequency response function measurements.

Extensions of these ideas can be carried out for general multiple-input/multiple-output problems, which require the definition and proper interpretation of multiple coherence functions and partial coherence functions. These general situations can be considered as combinations of a set of multiple-input/single-output models for a given set of stationary inputs and for different constant-parameter linear systems, as shown in Figure 1.17. Modern procedures for solving multiple-input/output problems are developed in Chapter 7 using conditioned (residual) spectral density functions. These procedures are extensions of classical regression techniques discussed in Chapter 4. In particular, the output autospectral density function in Figure 1.17 is decomposed to show how much of this output spectrum at any frequency is due to any input conditioned on other inputs in a prescribed order.

Basic statistical principles to evaluate random data properties are covered in Chapter 4. Error analysis formulas for bias errors and random errors are developed in Chapters 8 and 9 for various estimates made in analyzing single random records and multiple random records. Included are random error formulas for estimates of frequency response functions (both gain factors and phase factors) and estimates of

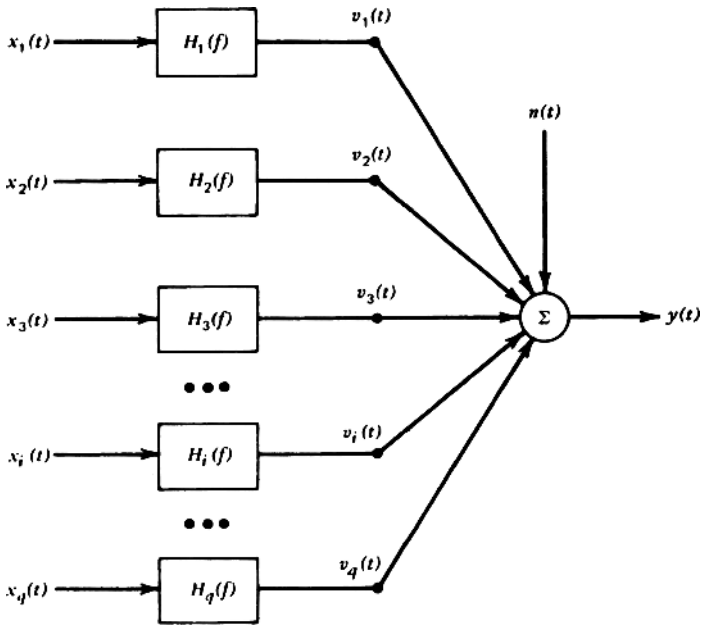

Figure 1.17 Multiple-input/single-output system with output noise.

coherence functions (ordinary, multiple, or partial). These computations are easy to apply and should be performed to obtain proper interpretations of measured results.

1.4.3 Error Analysis Criteria

Some error analysis criteria for measured quantities will now be defined as background for the material in Chapters 8 and 9. Let a hat (^) symbol over a quantity ϕ, namely, $\hat{\phi}$, denote an estimate of this quantity. The quantity $\hat{\phi}$ will be an estimate of ϕ based on a finite time interval or a finite number of sample points.

Conceptually, suppose $\hat{\phi}$ can be estimated many times by repeating an experiment or some measurement program. Then, the expected value of $\hat{\phi}$, denoted by $E[\hat{\phi}]$, is something one can estimate. For example, if an experiment is repeated many times to yield results $\hat{\phi}_i$, $i = 1, 2, \ldots, N$, then

$$E[\hat{\phi}] = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \hat{\phi}_i \tag{1.13}$$

This expected value may or may not equal the true value ϕ. If it does, the estimate $\hat{\phi}$ is said to be unbiased. Otherwise, it is said to be biased. The *bias* of the estimate, denoted $b[\hat{\phi}]$, is equal to the expected value of the estimate minus the true value—that is,

$$b[\hat{\phi}] = E[\hat{\phi}] - \phi \tag{1.14}$$

It follows that the bias error is a systematic error that always occurs with the same magnitude in the same direction when measurements are repeated under identical circumstances.

The *variance* of the estimate, denoted by $\mathrm{Var}[\hat{\phi}]$, is defined as the expected value of the squared differences from the mean value. In equation form,

$$\mathrm{Var}[\hat{\phi}] = E[(\hat{\phi} - E[\hat{\phi}])^2] \tag{1.15}$$

The variance describes the *random error* of the estimate—that is, that portion of the error that is not systematic and can occur in either direction with different magnitudes from one measurement to another.

An assessment of the total estimation error is given by the *mean square error,* which is defined as the expected value of the squared differences from the *true* value. The mean square error of $\hat{\phi}$ is indicated by

$$\text{mean square error}[\hat{\phi}] = E[(\hat{\phi} - \phi)^2] \tag{1.16}$$

It is easy to verify that

$$E[(\hat{\phi} - \phi)^2] = \mathrm{Var}[\hat{\phi}] + (b[\hat{\phi}])^2 \tag{1.17}$$

In words, the mean square error is equal to the variance plus the square of the bias. If the bias is zero or negligible, then the mean square error and variance are equivalent.

Figure 1.18 illustrates the meaning of the bias (systematic) error and the variance (random) error for the case of testing two guns for possible purchase by shooting each gun at a target. In Figure 1.18(*a*), gun A has a large bias error and small variance error. In Figure 1.18(*b*), gun B has a small bias error but large variance error. As shown, gun A will never hit the target, whereas gun B will occasionally hit the target. Nevertheless, most people would prefer to buy gun A because the bias error can be removed (assuming one knows it is present) by adjusting the sights of the gun, but the random

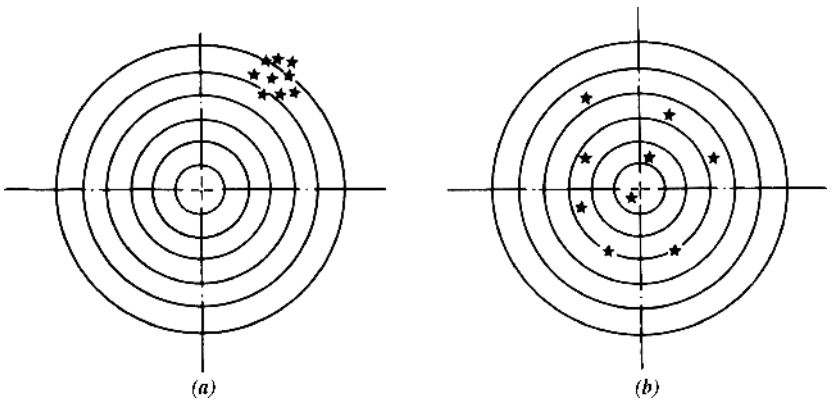

Figure 1.18 Random and bias errors in gun shoots at a target. (*a*) Gun A: large bias error and small random error. (*b*) Gun B: small bias error and large random error.

error cannot be removed. Hence, gun A provides the potential for a smaller mean square error.

A final important quantity is the *normalized rms error* of the estimate, denoted by $\varepsilon[\hat{\phi}]$. This error is a dimensionless quantity that is equal to the square root of the mean square error divided by the true value (assumed, of course, to be different from zero). Symbolically,

$$\varepsilon[\hat{\phi}] = \frac{\sqrt{E[(\hat{\phi}-\phi)^2]}}{\phi} \tag{1.18}$$

In practice, one should try to make the normalized rms error as small as possible. This will help to guarantee that an arbitrary estimate $\hat{\phi}$ will lie close to the true value ϕ.

1.4.4 Data Analysis Procedures

Recommended data analysis procedures are discussed in more detail in Chapters 10–14. Chapter 10 deals with data acquisition problems, including data collection, storage, conversion, and qualification. General steps are outlined for proper data analysis of individual records and multiple records, as would be needed for different applications. Digital data analysis techniques discussed in Chapter 11 involve computational procedures to perform trend removal, digital filtering, Fourier series, and fast Fourier transforms on discrete time series data representing sample records from stationary (ergodic) random data. Digital formulas are developed to compute estimates of probability density functions, correlation functions, and spectral density functions for individual records and for associated joint records. Further detailed digital procedures are stated to obtain estimates of all of the quantities described in Chapters 6 and 7 to solve various types of single-input/output problems and multiple-input/output problems. Chapter 12 is devoted to separate methods for nonstationary data analysis, and Chapter 13 develops Hilbert transform techniques. Chapter 14 discusses models for nonlinear system analysis.

PROBLEMS

1.1 Determine the period of the function defined by

$$x(t) = \sin 11t + \sin 12t$$

1.2 For the following functions, which are periodic and which are nonperiodic?

(a) $x(t) = 3 \sin t + 2 \sin 2t + \sin 3t$.

(b) $x(t) = 3 \sin t + 2 \sin 2t + \sin \pi t$.

(c) $x(t) = 3 \sin 4t + 2 \sin 5t + \sin 6t$.

(d) $x(t) = e^{-t} \sin t$.

1.3 If a stationary random process $\{x(t)\}$ has a mean value of μ_x, what is the limiting value of the autocorrelation function $R_{xx}(\tau)$ as the time delay τ becomes long?

1.4 An estimate is known to have a mean square error of 0.25 and a bias error of 0.40. Determine the variance of the estimate.

1.5 In Problem 1.4, if the quantity being estimated has a true value of $\phi = 5$, what is the normalized rms error of the estimate?

In Problems 1.6–1.9 state which properties are always true.

1.6 A stationary random process must

(a) be discrete.

(b) be continuous.

(c) be ergodic.

(d) have ensemble-averaged properties that are independent of time.

(e) have time-averaged properties that are equal to the ensemble-averaged properties.

1.7 An ergodic random process must

(a) be discrete.

(b) be continuous.

(c) be stationary.

(d) have ensemble-averaged properties that are independent of time.

(e) have time-averaged properties that are equal to the ensemble-averaged properties.

1.8 A single sample function can be used to find all statistical properties of a random process if the process is

(a) deterministic.

(b) ergodic.

(c) stationary.

(d) all of the above.

1.9 The autocorrelation function of a stationary random process

(a) must decrease as $|\tau|$ increases.

(b) is a function of the time difference only.

(c) must approach a constant as $|\tau|$ increases.

(d) must always be nonnegative.

1.10 How do the answers to Problem 1.9 change if the stationary random process is mixed with a periodic process?