# OPTIMAL CONTROL AND ESTIMATION

Robert F. Stengel

*Department of Mechanical and Aerospace Engineering*
*Princeton University, Princeton, New Jersey*

DOVER PUBLICATIONS, INC.
New York

where $k_x$ and $d_{\dot{x}}$ are evaluated at the nominal operating or equilibrium point. However, experimentation may reveal decidedly nonlinear characteristics, as shown in Figs. 2.3-3*b* and *c*. In addition to curvature, these sketches illustrate limiting, discontinuity, "stiction," asymmetry, and hysteresis in the forces, all of which can occur in practice. Although not shown, it also is possible for the damper to exert a spring force and for the spring to provide damping. Deterministic curve-fitting procedures are required to portray $k(x)$ and $d(\dot{x})$ over the expected ranges of $x$ and $\dot{x}$ for integration of Eq. 2.3-92.

The two most commonly used approaches to deterministic representation of nonlinear functions for digital computation are polynomial and tabular approximation. In the first approach, the original nonlinear function is replaced by an additive series of simpler functions that fits the original function with arbitrarily small error. The function then is evaluated by computing the polynomial. In the second approach, values of the original function are tabulated at given values of the independent variable(s), and the function is evaluated between these points by interpolation. The two approaches can be combined, providing a polynomial fit between tabulated points. Whichever method is used, the instantaneous value of the independent variable becomes the *argument* of the function evaluation. For example, in performing a modified Euler integration of Eq. 2.3-92, $k(x_1)$ should be computed as $k[x_1(t_{k-1})]$ in calculating $\Delta\mathbf{x}_1$ and as $k[x_1(t_{k-1}) + \Delta x_1]$ in calculating $\Delta\mathbf{x}_2$.

## 2.4 RANDOM VARIABLES, SEQUENCES, AND PROCESSES

### Scalar Random Variables

A scalar variable $x$ may take different values each time it is sampled; if these values bear no deterministic (or fixed) relationship to the sampling process, $x$ is a *scalar random variable*. Suppose a candy store clerk scoops jelly beans, and $x$ represents the number in one scoop. The number will vary from one scoop (or *event*) to the next, so $x$ is a random variable.

The statistics of $x$ could be determined by careful experiment. Given $N$ events, the frequency of occurrence for specific values of $x$ provides an estimate of the probability that a given result will be obtained in a future sampling:

$$\Pr(x_i) = n_i/N, \qquad i = 1 \text{ to } I \tag{2.4-1}$$

In the example, $n_i$ is the number of scoops containing $x_i$ jelly beans and $I$ is the number of different values of $x_i$ obtained in the experiments. Of course, the probability estimates determined in this way are statistically significant only if $N$ is very large; however, even when $N$ is small, the total number of



jelly beans is $\sum_{i=1}^{I} x_i n_i$, and

$$\sum_{i=1}^{I} \Pr(x_i) = \frac{1}{N}\sum_{i=1}^{I} n_i = 1 \tag{2.4-2}$$

By its definition, $\Pr(x_i)$ cannot be negative, and its value is between 0 and 1.

$\Pr(x_i)$, which is also called a *probability mass function*, is the appropriate statistic when $x$ takes discrete values. If $x$ is a continuous variable, the probability mass function must be redefined. Now the clerk is dispensing juice instead of jelly beans, and $x$ represents the volume in each glass. Assume that the device for measuring volume has a resolution of $\pm\Delta x/2$; then $n_i$ represents the number of glasses containing $(x_i \pm \Delta x/2)$ milliliters of juice, and the preceding equations become

$$\Pr\left(x_i \pm \frac{\Delta x}{2}\right) = \frac{n_i}{N}, \qquad i = 1 \text{ to } I \tag{2.4-3}$$

$$\sum_{i=1}^{I} \Pr\left(x_i \pm \frac{\Delta x}{2}\right) = 1 \tag{2,4-4}$$

To this point, it has been assumed that the amplitude of the scalar variable is restricted to a finite set of values (i.e., that it is quantized). The development can be extended to continuous probability distributions by considering what happens as $\Delta x$ becomes vanishingly small and $N$ becomes arbitrarily large to accommodate the decreasing number of measurements in each interval. First, the *probability density function* for discrete sampling is defined as follows:

$$\text{pr}(x_i) = \frac{\Pr(x_i \pm \Delta x/2)}{\Delta x} \tag{2.4-5a}$$

or

$$\Pr(x_i \pm \Delta x/2) = \text{pr}(x_i)\Delta x \tag{2.4-5b}$$

Then $\Delta x$ decreases as $N$ and $I$ increase, and in the limit,

$$\sum_{i=1}^{I} \text{pr}(x_i)\Delta x \xrightarrow[\Delta x \to 0]{I \to \infty} \int_{x_{\min}}^{x_{\max}} \text{pr}(x)\,dx = 1 \tag{2.4-6}$$

More generally, $x$ may have no finite limits, so the limits of the summation and the integral are expanded for both continuous and discrete measures of

probability:

$$\sum_{i=1}^{\infty} \Pr(x_i) = 1 \tag{2.4-7}$$

$$\int_{-\infty}^{\infty} \text{pr}(x)\, dx = 1 \tag{2.4-8}$$

The average or *expected value* of $x$ is readily determined from $\Pr(x_i)$ or $\text{pr}(x)$. Denoting the expected value by $E(x)$ or $\mathscr{E}(x)$,

$$E(x) = \sum_{i=1}^{\infty} x_i \Pr(x_i) = \bar{x} \tag{2.4-9}$$

for a discrete random variable, and

$$\mathscr{E}(x) = \int_{-\infty}^{\infty} x\, \text{pr}(x)\, dx = \bar{x} \tag{2.4-10}$$

for a continuous random variable. $E(x)$ or $\mathscr{E}(x)$ also is called the *first moment about the origin* or the *mean value* of $x$, and it is denoted by $\bar{x}$. The mean value has the same units as $x$. Taking the expected value of a discrete variable is a linear operation because

$$E(ax) = aE(x) \tag{2.4-11}$$

$$E\left[\sum_{k=1}^{K} x_k\right] = \sum_{k=1}^{K} E(x_k) \tag{2.4-12}$$

and the same is true for continuous variables.

*Higher moments* provide measures of the variability of $x$, and the $n^{\text{th}}$ moments of discrete and continuous variables are defined by similar expressions:

$$E(x^n) = \sum_{i=1}^{\infty} x_i^n \Pr(x_i) \tag{2.4-13}$$

$$\mathscr{E}(x^n) = \int_{-\infty}^{\infty} x^n\, \text{pr}(x)\, dx \tag{2.4-14}$$

Higher moments about the origin reflect not only variation about the mean but variation in the mean value itself. The former taken alone is a more useful measure of variation, leading to the *nth central moments* for discrete and continuous variables:



$$E[(x - \bar{x})^n] = \sum_{i=1}^{\infty} [(x - \bar{x})^n \Pr(x_i)] \tag{2.4-15}$$

$$\mathscr{E}[(x - \bar{x})^n] = \int_{-\infty}^{\infty} (x - \bar{x})^n\, \text{pr}(x)\, dx \tag{2.4-16}$$

The first central moment is zero in either case, while the second central moment defines the *variance* of $x$:

$$E[x - E(x)] = E(x - \bar{x}) = 0 \tag{2.4-17}$$

$$\mathscr{E}[x - \mathscr{E}(x)] = \mathscr{E}(x - \bar{x}) = 0 \tag{2.4-18}$$

$$E[(x - \bar{x})^2] = \sigma^2 \tag{2.4-19}$$

$$\mathscr{E}[(x - \bar{x})^2] = \sigma^2 \tag{2.4-20}$$

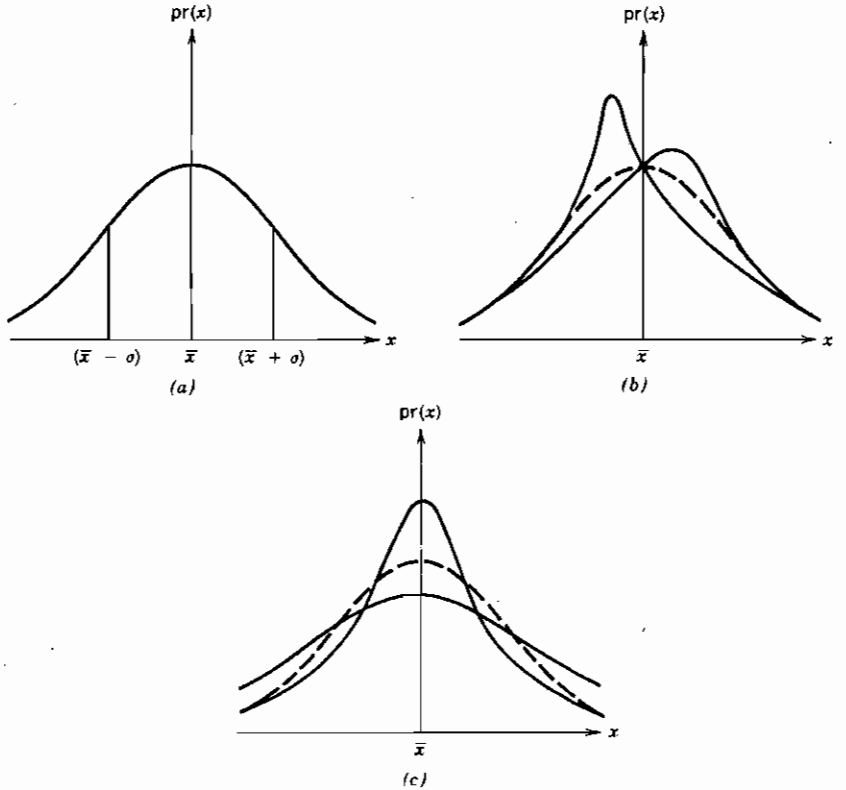


**FIGURE 2.4-1** Gaussian and non-Gaussian probability distributions. (*a*) Gaussian distribution; (*b*) skewed distributions; (*c*) distributions with non-Gaussian kurtosis.

The *standard deviation* of $x$ is $\sigma$, and it has the same units as $x$. (Note that the mean and standard deviation of a discrete random variable need not equal one of the discrete values.) It will be seen that the mean and standard deviation are sufficient to define the Gaussian probability distribution (Fig. 2.4-1*a*) described later. In the remainder, expected values of both discrete and continuous variables will be signified by $E(\cdot)$.

The higher central moments reflect important attributes of non-Gaussian distributions; the third moment indicates *skewness* of the distribution (Fig. 2.4-1*b*), while the fourth moment (or *kurtosis*) measures the steepness of the distribution's peak (Fig. 2.4-1*c*). The Gaussian distribution is symmetric about the mean so its third central moment is zero. Its kurtosis is not zero, but the distribution is defined without reference to the fourth moment's value.

### Groups of Random Variables

When two random variables are sampled, the probabilities that each takes a certain value may be interrelated. Suppose that the aforementioned jelly beans come in two colors (orange and black), that $N$ scoopfuls are obtained, and that $x_i$ and $y_j$ represent the numbers of each color in each scoopful. The unconditional, individual probabilities can be estimated as before:

$$\Pr(x_i) = n(x_i)/N, \qquad i = 1 \text{ to } I \tag{2.4-21}$$

$$\Pr(y_j) = n(y_j)/N, \qquad j = 1 \text{ to } J \tag{2.4-22}$$

The true probabilities are obtained only as $N \to \infty$, but for any $N$,

$$\sum_{i=1}^{I} \Pr(x_i) = \frac{1}{N} \sum_{i=1}^{I} n(x_i) = 1 \tag{2.4-23}$$

$$\sum_{j=1}^{J} \Pr(y_j) = \frac{1}{N} \sum_{j=1}^{J} n(y_j) = 1 \tag{2.4-24}$$

Now we are interested in knowing the *joint probability* that $x_i$ and $y_j$ occur simultaneously, signified by $\Pr(x_i, y_j)$. If the orange and black jelly beans were kept in separate barrels and each was sampled in separate scoops, then the probabilities would be unrelated, and the joint probability would be

$$\Pr(x_i, y_j) = \Pr(x_i) \Pr(y_j) \tag{2.4-25}$$

If they were kept in the same barrel and a single scoop was used, then the numbers in a single event would be related, and the joint probability would



be estimated by the following:

$$\Pr(x_i, y_j) = \frac{n(x_i)}{N} \text{ and } \frac{n(y_j)}{N}, \qquad i = 1 \text{ to } I, \quad j = 1 \text{ to } J \tag{2.4-26}$$

Whether or not $x_i$ and $y_j$ are related, the sum over all possible values must equal one:

$$\sum_{i=1}^{I} \sum_{j=1}^{J} \Pr(x_i, y_j) = 1 \tag{2.4-27}$$

The *conditional probability* of $x_i$ given $y_j$ is expressed by $\Pr(x_i|y_j)$, and it is simply

$$\Pr(x_i|y_j) = \frac{\Pr(x_i, y_j)}{\Pr(y_j)} \tag{2.4-28}$$

Similarly, the conditional probability of $y_j$ given $x_i$ is

$$\Pr(y_j|x_i) = \frac{\Pr(x_i, y_j)}{\Pr(x_i)} \tag{2.4-29}$$

If $x$ and $y$ are unrelated (Eq. 2.4-25), then the conditional and unconditional probabilities are the same; that is,

$$\Pr(x_i|y_j) = \Pr(x_i) \tag{2.4-30}$$

$$\Pr(y_j|x_i) = \Pr(y_j) \tag{2.4-31}$$

Rearranging Eqs. 2.4-28 and 2.4-29,

$$\Pr(x_i|y_j) \Pr(y_j) = \Pr(y_j|x_i) \Pr(x_i) \tag{2.4-32}$$

leading to alternate expressions for *Bayes's Rule*:*

$$\Pr(x_i|y_j) = \frac{\Pr(y_j|x_i) \Pr(x_i)}{\Pr(y_j)} \tag{2.4-33}$$

$$\Pr(y_j|x_i) = \frac{\Pr(x_i|y_j) \Pr(y_j)}{\Pr(x_i)} \tag{2.4-34}$$

*English clergyman Thomas Bayes (1702–1761) established the basis for statistical inference in a brief piece entitled "Essay Towards Solving a Problem in the Doctrine of Chances."

We will have need of such relationships in developing multiple-model estimation algorithms in Section 4.7. Note that the probability of $x_i$ is

$$\Pr(x_i) = \sum_{j=1}^{J} [\Pr(x_i|y_j) \Pr(y_j)] \tag{2.4-35}$$

and if $\Pr(y_j) = 1$, then

$$\Pr(x_i, y_j) = \Pr(x_i|y_j) = \Pr(x_i) \tag{2.4-36}$$

Analogous expressions hold for $y_j$.

Following the discrete results, conditional and joint probability density functions can be developed for continuous random variables. The two probability density functions must satisfy

$$\int_{-\infty}^{\infty} \text{pr}(x)dx = 1 \tag{2.4-37}$$

$$\int_{-\infty}^{\infty} \text{pr}(y)dy = 1 \tag{2.4-38}$$

and the *joint probability density function* must satisfy

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \text{pr}(x, y)dx\,dy = 1 \tag{2.4-39}$$

The notations $\text{pr}(x)$, $\text{pr}(y)$, and $\text{pr}(x, y)$ should be read "probability density function of $x$, of $y$, and of $x$ and $y$," respectively, signifying possibly different mathematical functions. The *conditional probability density function* is then defined as

$$\text{pr}(x|y) = \frac{\text{pr}(x, y)}{\text{pr}(y)} \tag{2.4-40}$$

and Bayes's Rule (Eq. 2.4-33) also holds true for probability density functions:

$$\text{pr}(x|y)\Delta x = \frac{\text{pr}(y|x)\Delta y\ \text{pr}(x)\Delta x}{\text{pr}(y)\Delta y} \tag{2.4-41}$$

Canceling the incremental values,

$$\text{pr}(x|y) = \frac{\text{pr}(y|x)\,\text{pr}(x)}{\text{pr}(y)} \tag{2.4-42}$$



It is possible to consider the conditional probabilities of variables with continuous and discrete distributions at the same time. If there are a discrete number of $x_i$ and a continuous distribution of $y_j$, the previous results imply that

$$\Pr(x_k|y) = \frac{\text{pr}(y|x_k)\Delta y \Pr(x_k)}{\text{pr}(y)\Delta y} = \frac{\text{pr}(y|x_k)\Pr(x_k)}{\text{pr}(y)}$$

$$= \frac{\text{pr}(y|x_k)\Pr(x_k)}{\sum_{i=1}^{I} \text{pr}(y|x_i)\Pr(x_i)} \tag{2.4-43}$$

where $k$ is a particular value in the interval $[1, I]$.

If $y$ is a function of $x$, its expected value and probability distribution can be determined using the probability distribution of $x$. We already have seen examples of the expected values of specific functions of $x$ in the discussion of higher moments; given $y = f(x)$, the previous equations suggest that

$$E[f(x)] = \int_{-\infty}^{\infty} f(x)\,\text{pr}(x)\,dx \tag{2.4-44}$$

Assuming that $f(x)$ is a continuous function, $y$ can be expressed as a Taylor series expansion of $x$,

$$y(x) = y(x_o) + \Delta y = f(x_o) + \left[\frac{\partial f(x)}{\partial x}\right]_{x=x_o} \Delta x + \cdots \tag{2.4-45}$$

so a first-order equation for the perturbation in $y$ due to a variation in $x$ is

$$\Delta y = \left.\frac{\partial f(x)}{\partial x}\right|_{x=x_o} \Delta x = \frac{\partial f(x_o)}{\partial x}\Delta x \tag{2.4-46}$$

Now the probability that $y$ lies in $\Delta y$ [referenced to $y(x_o)$] must be the same as the probability that $x$ is in $\Delta x$; hence

$$\text{pr}[y(x_o)]\Delta y = \text{pr}[y(x_o)]\frac{\partial f(x_o)}{\partial x}\Delta x$$

$$= \text{pr}(x_o)\Delta x \tag{2.4-47a}$$

or

$$\text{pr}[y(x_o)] = \frac{\text{pr}(x_o)}{|\partial f(x_o)/\partial x|} \tag{2.4-47b}$$

This equation assumes that there is a monotonic relationship between $y$ and $x$ (and between $x$ and $y$), but multivalued functions often occur. In such case, Eq. 2.4-47b must be modified by a suitable multiplier. If

$$y = f(x) = A \sin(\omega t + x) \tag{2.4-48}$$

where $x$ is a random phase angle between 0 and $2\pi$, $y$ takes the same value twice in the range of $x$. Let $x$ have a *uniform* (or *rectangular*) *distribution*,

$$\text{pr}(x) = \frac{1}{2\pi}, \qquad 0 \le x \le 2\pi \tag{2.4-49}$$

Then the sensitivity of $y$ to variations in $x$ is

$$\frac{\partial f}{\partial x} = A\cos(\omega t + x) = A\sqrt{1 - \sin^2(\omega t + x)}$$
$$= \sqrt{A^2 - y^2} \tag{2.4-50}$$

Multiplying Eq. 2.4-47b by two, the *probability density function of a sine wave* with amplitude $A$ and random phase angle is

$$\text{pr}(y) = \begin{cases} \dfrac{1}{\pi\sqrt{A^2 - y^2}}, & y \le A \\ 0, & y > A \end{cases} \tag{2.4-51}$$

Note that the distribution is independent of the natural frequency $\omega$ of the sine wave.

A distinctly different situation arises when the probability distribution of the *sum* of two or more independent random variables must be determined. If $z$ is the sum of independent variables,

$$z = x + y \tag{2.4-52}$$

then the conditional probability density function of $z$ given $x$ is

$$\text{pr}(z|x) = \text{pr}(y) = \text{pr}(z - x) \tag{2.4-53}$$

and the unconditional probability density function of $z$ is

$$\text{pr}(z) = \int_{-\infty}^{\infty} \text{pr}(z|x)\,\text{pr}(x)\,dx = \int_{-\infty}^{\infty} \text{pr}(z - x)\,\text{pr}(x)\,dx \tag{2.4-54}$$

This equation is known as a *convolution* or *superposition integral* (introduced in a different context in Section 2.3); pr($z$) results from an integral smoothing of component probability density functions. Furthermore, from Eq. 2.4-12, the mean value and variance of $z$ are



$$\bar{z} = \bar{x} + \bar{y} \tag{2.4-55}$$

$$\sigma_z^2 = \sigma_x^2 + \sigma_y^2 \tag{2.4-56}$$

and these results (Eqs. 2.4-54 to 2.4-56) can be extended for sums of greater numbers of independent variables:

$$z = \sum_{i=1}^{I} a_i x_i \tag{2.4-57}$$

$$\bar{z} = \sum_{i=1}^{I} a_i \bar{x}_i \tag{2.4-58}$$

$$\sigma_z^2 = \sum_{i=1}^{I} a_i^2 \sigma_{x_i}^2 \tag{2.4-59}$$

The *Central Limit Theorem* states that the probability density function of the sum approaches a Gaussian distribution as the number grows, even when the component distributions are markedly non-Gaussian. This finding is fortuitous for estimation problems, which often involve such sums, because the *Gaussian distribution* is specified by only two quantities: its mean value, $\bar{x}$, and its standard deviation, $\sigma_x$ (or variance, $\sigma_x^2$).* If $x$ is a Gaussian random variable, its probability density function is

$$\text{pr}(x) = \frac{1}{(2\pi)^{1/2}\sigma_x} e^{-(x-\bar{x})^2/2\sigma_x^2} \tag{2.4-60}$$

Evaluating $E(x)$ and $E[(x - \bar{x})^2]$ for this distribution will confirm the previous statement.

## Scalar Random Sequences and Processes

A *scalar random sequence* consists of a group of related scalar random variables that occur at discrete points in time or space. The variables themselves can have continuous or discrete amplitude distributions, but they are defined at distinct points. In the remainder of the section, it will be assumed that the variables have continuous amplitudes and that the sampling variable represents time increments; hence the random sequence $x_1$, $x_2$, $x_3$ . . . is accompanied by the probability *density* sequence pr($x_1$), pr($x_2$), and so on. A *scalar random process* is continuous in time as well as amplitude, so

*In the multivariate case, these two quantities are the mean value *vector* and the covariance *matrix*, as described later in this section.

$x(t)$ is accompanied by $\text{pr}[x(t)]$. The distinction between processes and sequences becomes important in the latter portion of this section; in the next few paragraphs, the comments on processes apply to both.

A specific "random" process actually is deterministic, in the sense that it happens (or happened) with probability one; however, there is no deterministic way of accurately predicting the future process, and the "random" fluctuations of past measurements may mask an underlying process that really is systematic. In such instance, it is useful to think of the process as one member of an *ensemble of sequences*, each of which characterizes the same phenomenon without having the same time history. The statistics of $x$ derive from estimates of the probability density function across the entire ensemble, and they may vary in time (i.e., they may be *nonstationary*).

An infinity of ensemble members would be required to compute precise statistics for a nonstationary random process (Fig. 2.4-2). In some few cases,

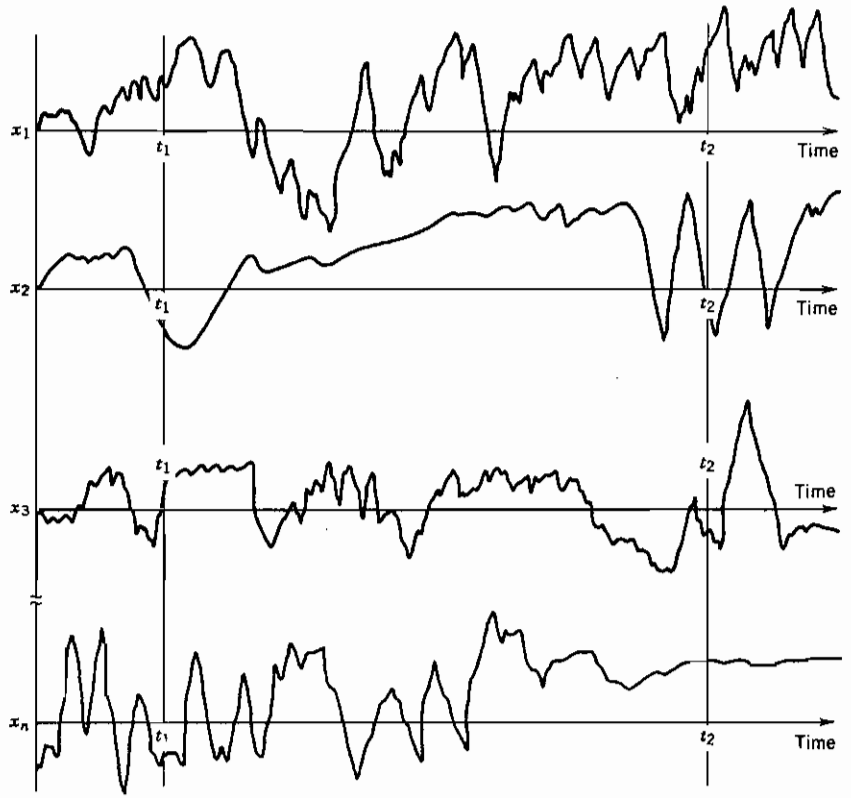


**FIGURE 2.4-2** Nonstationary ensemble of random sequences.



enough members of an ensemble are available to provide experimental nonstationary statistics for a random variable; however, nonstationary statistics usually derive from assumptions or from mathematical inferences, such as, optimal estimation algorithms that do exhibit systematic behavior and that model the propagation of the random variable's probability density function in time.

A random process is *stationary* if its ensemble statistics do not vary in time (Fig. 2.4-3). The probability distribution of a stationary process computed at $t_1$ has the same mean value and central moments as a probability distribution computed at $t_2$. In both the stationary and nonstationary cases, the probability density function can be defined as

$$\text{pr}[x(t)] = \lim_{\substack{M\to\infty \\ \Delta x\to 0}} \frac{1}{M} \sum_{m=1}^{M} \frac{n[x(t) \pm \Delta x/2]}{\Delta x} \tag{2.4-61}$$

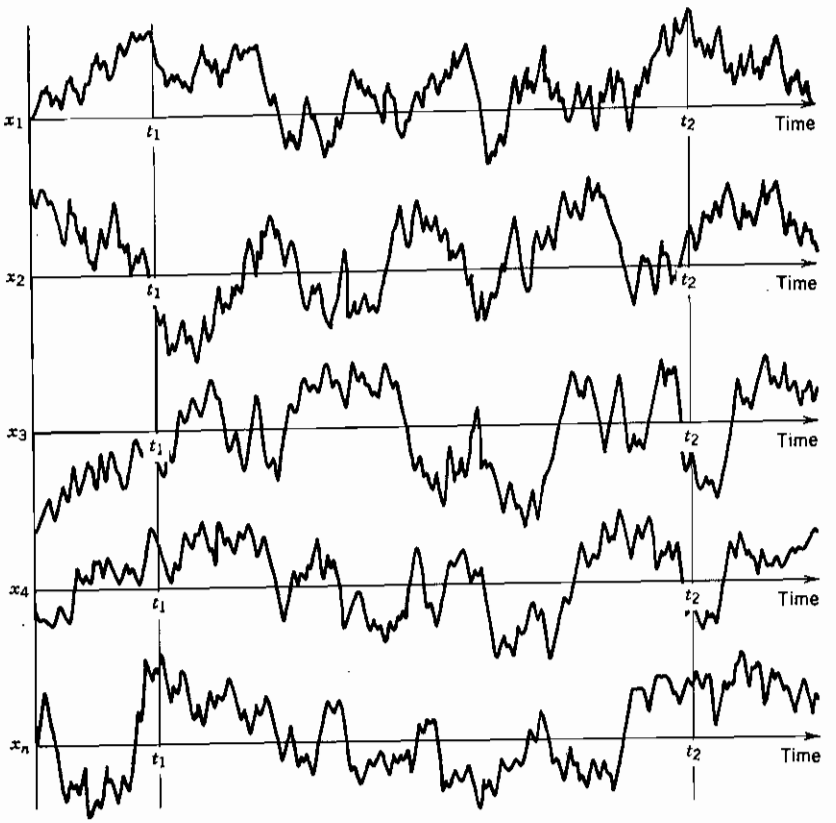


**FIGURE 2.4-3** Stationary ensemble of ergodic sequences.

where $M$ is the number of ensemble members and $n[x(t) \pm \Delta x/2]$ is the number of events in the $\Delta x$ band centered at $x(t)$. For a stationary random process, $\text{pr}[x(t)]$ is invariant with $t$; for a nonstationary process, it is not.

A stationary random process, $\chi(t)$, can generate a nonstationary random process, $x(t)$, if the former is multiplied by a deterministic, time-varying coefficient, $a(t)$:

$$x(t) = a(t)\chi(t) \tag{2.4-62}$$

Assuming that $\chi(t)$ is a Gaussian random variable with fixed mean $\bar{\chi}$ and standard deviation $\sigma_\chi$, the *nonstationary Gaussian probability density function* is readily described as

$$\text{pr}[x(t)] = \frac{1}{(2\pi)^{1/2}\sigma_x(t)} e^{-[x(t)-\bar{x}(t)]^2/2\sigma_x^2(t)} \tag{2.4-63}$$

with

$$\bar{x}(t) = a(t)\bar{\chi} \tag{2.4-64}$$

$$\sigma_x^2 = a^2(t)\sigma_\chi^2 \tag{2.4-65}$$

Equation 2.4-61 defines what is sometimes called the *first* probability density function of $x$, where the *second* probability density function is the *joint* probability density function of $x(t_1)$ and $x(t_2)$, for arbitrary $t_1$ and $t_2$:

$$\text{pr}[x(t_1), x(t_2)] = \lim_{\substack{M\to\infty \\ \Delta x\to 0}} \frac{1}{M} \sum_{m=1}^{M} \left\{ \frac{n[x(t_1) \pm \Delta x/2]}{\Delta x^2} \text{ and } \frac{n[x(t_2) \pm \Delta x/2]}{\Delta x^2} \right\} \tag{2.4-66}$$

For a nonstationary process, $\text{pr}[x(t_1), x(t_2)]$ depends on the choice of $t_1$ and $t_2$; however, the second probability density function of a stationary process depends only on the *difference* between $t_1$ and $t_2$,

$$\begin{aligned} \text{pr}[x(t_1), x(t_2)] &= \text{pr}[x(t_1), x(t_1 + \Delta t)] \\ &= \text{pr}[x(t), x(t + \Delta t)] \end{aligned} \tag{2.4-67}$$

where $t$ and $\Delta t$ are arbitrary. The nonstationary process's first probability density function (p.d.f.) varies in time, and its second p.d.f. is distributed differently along the $x(t_1)$ and $x(t_2)$ axes (Fig. 2.4-4*a*). The stationary process's first p.d.f. is the same at all times, and its distribution along the $x(t)$ and $x(t+\Delta t)$ axes for any $t$ and $\Delta t$ are the same (Fig. 2.4-4*b*).

A special case of the stationary process plays an important role in practical application. The probability densities could be calculated for samples of a single random process taken at different times (i.e., using *time*

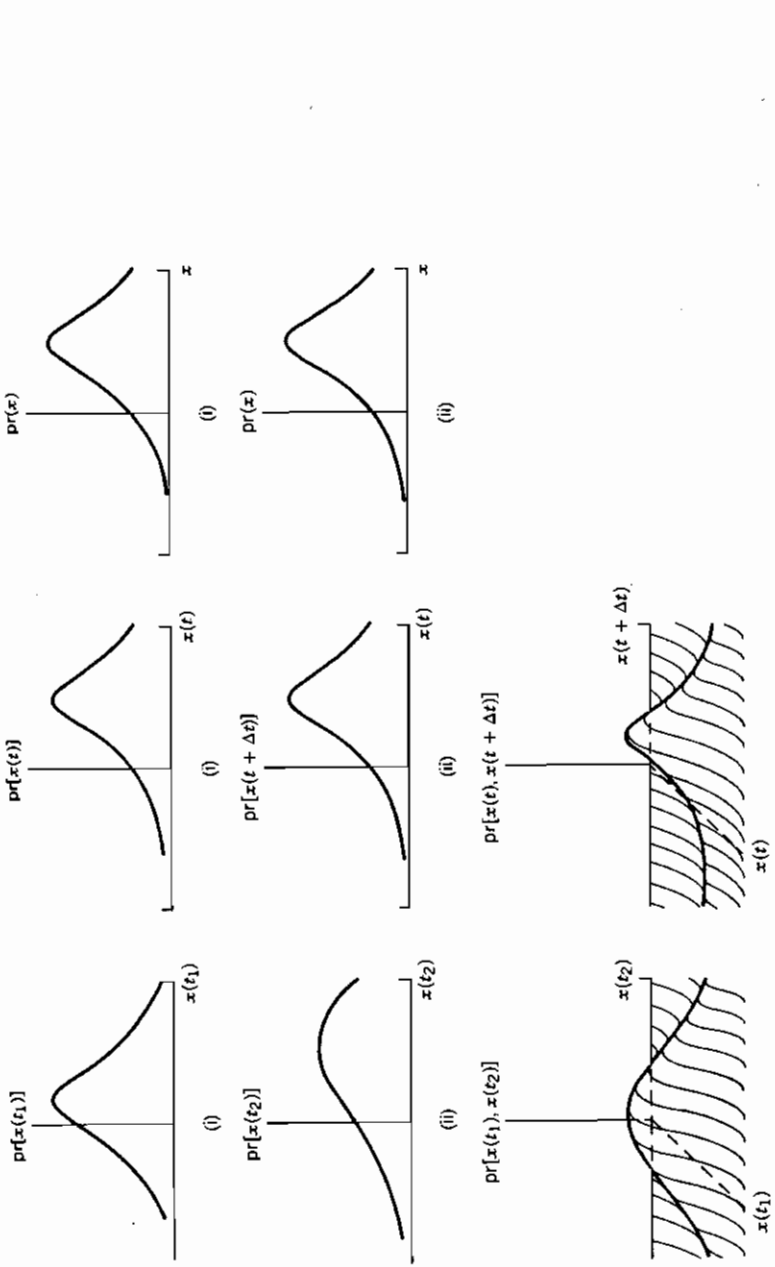


**FIGURE 2.4-4** Probability density functions (p.d.f.) for nonstationary, stationary, and ergodic processes. (*a*) Nonstationary process: (i) first p.d.f. at $t_1$; (ii) first p.d.f. at $t_2$; (iii) second p.d.f. at $t_1$ and $t_2$. (*b*) Stationary process: (i) first p.d.f. at $t$; (ii) first p.d.f. at $t+\Delta t$; (iii) second p.d.f. at $t$ and $t+\Delta t$. (*c*) Ergodic process: (i) ensemble-averaged p.d.f.; (ii) time-averaged p.d.f.

*averaging* rather than *ensemble averaging*, see Fig. 2.4-4*c*). If these values are identical to probability densities computed for the ensemble, then the stationary process is said to be *ergodic*. The time-based first and second probability density functions can be calculated as

$$\text{pr}[x(t)] = \text{pr}(x) = \lim_{\substack{M\to\infty \\ \Delta x\to 0}} \frac{1}{M} \sum_{m=1}^{M} \frac{n(x \pm \Delta x/2)}{\Delta x} \tag{2.4-68}$$

$$\text{pr}[x(t), x(t+\Delta t)] = \lim_{\substack{M\to\infty \\ \Delta x\to 0}} \sum_{m=1}^{M} \left\{ \frac{n[x(t) \pm \Delta x/2]}{\Delta x^2} \text{ and } \frac{n[x(t+\Delta t) \pm \Delta x/2]}{\Delta x^2} \right\} \tag{2.4-69}$$

where $M$ is the number of samples taken at different times on a single record. For an ergodic process, the results of these calculations are identical to those derived from Eqs. 2.4-61 and 2.4-67. Examples of stationary, nonergodic processes include ensembles whose members have constant but different mean values and/or standard deviations.

### Correlation and Covariance Functions

The second probability density function is used in the definition of a joint expected value of $x(t_1)$ and $x(t_2)$ that is called the *autocorrelation function* of $x$. If $x$ is defined as having continuous amplitude, the autocorrelation function for a nonstationary random sequence or process is expressed as

$$E[x(t_1), x(t_2)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x(t_1)x(t_2)\,\text{pr}[x(t_1), x(t_2)]\,dx(t_1)\,dx(t_2)$$
$$= \psi[x(t_1), x(t_2)] \tag{2.4-70}$$

The similarity to higher moments is apparent, and it is sometimes more useful to subtract out mean values, leading to the *autocovariance function*:

$$E\{[x(t_1) - \bar{x}(t_1)][x(t_2) - \bar{x}(t_2)]\} = \phi[\tilde{x}(t_1), \tilde{x}(t_2)]$$
$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [x(t_1) - \bar{x}(t_1)][x(t_2) - \bar{x}(t_2)]\,\text{pr}[x(t_1), x(t_2)]\,dx(t_1)\,dx(t_2) \tag{2.4-71}$$

with $\tilde{x}(t) \triangleq x(t) - \bar{x}(t)$. When $t_1 = t_2$,

$$E\{[x(t_1) - \bar{x}(t_1)][x(t_2) - \bar{x}(t_2)]\} = E\{[x(t_1) - \bar{x}(t_1)]^2\}$$
$$= \sigma_x^2 \tag{2.4-72}$$

which is the *variance* of $x$ at $t_1$. The autocorrelation and autocovariance functions are the same when the mean values are zero.



Given two random processes (or sequences), $x(t)$ and $y(t)$, the *cross-correlation* and *cross-covariance functions* are similarly defined. The latter is

$$E\{[x(t_1) - \bar{x}(t_1)][y(t_2) - \bar{y}(t_2)]\} = \phi[\tilde{x}(t_1), \tilde{y}(t_2)]$$
$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [x(t_1) - \bar{x}(t_1)][y(t_2) - \bar{y}(t_2)]\,\text{pr}[x(t_1), y(t_2)]\,dx(t_1)\,dy(t_2) \tag{2.4-73}$$

with $\tilde{y}(t) \triangleq y(t) - \bar{y}(t)$, and the *covariance* of $x$ and $y$ is given by the value of this equation for $t_1 = t_2$:

$$E\{[x(t_1) - \bar{x}(t_1)][y(t_1) - \bar{y}(t_1)]\} = \sigma_{xy}(t_1) \tag{2.4-74}$$

The sample mean $x_m$ of any variable $x$ can be estimated as

$$x_m = \frac{1}{N} \sum_{n=1}^{N} x_n \tag{2.4-75}$$

and as $N \to \infty$, $x_m \to \bar{x}$. The sample covariance $s_{xy}$ of any two variables $x$ and $y$ can be obtained from

$$s_{xy} = \frac{1}{N-1} \sum_{n=1}^{N} (x_n - x_m)(y_n - y_m) \tag{2.4-76}$$

and as $N \to \infty$, $s_{xy} \to \sigma_{xy}$. These equations could be used to evaluate the auto- and cross-covariance functions of nonstationary and stationary random processes using ensemble data.

For stationary processes with $\Delta t = t_2 - t_1$ (Fig. 2.4-5*a*),

$$\phi[\tilde{x}(t_1), \tilde{x}(t_2)] = \phi[\tilde{x}(t_1), \tilde{x}(t_1 + \Delta t)]$$
$$= \phi_{xx}(\Delta t) \tag{2.4-77}$$

$$\phi[\tilde{x}(t_1), \tilde{y}(t_2)] = \phi[\tilde{x}(t_1), \tilde{y}(t_1 + \Delta t)]$$
$$= \phi_{xy}(\Delta t) \tag{2.4-78}$$

Because the ordering of $x$ in Eq. 2.4-77 is immaterial, $\phi_{xx}(\Delta t) = \phi_{xx}(-\Delta t)$, so the autocovariance function is symmetric about the origin. The cross-covariance function is not symmetric because different variables are sampled at the two times separated by $\Delta t$. The covariance functions have the following additional properties:

$$\phi_{xx}(0) \geq \phi_{xx}(\Delta t) \tag{2.4-79}$$

$$\phi_{yy}(0) \geq \phi_{yy}(\Delta t) \tag{2.4-80}$$

$$\phi_{xx}(0)\phi_{yy}(0) \geq [\phi_{xy}(\Delta t)]^2 \tag{2.4-81}$$

If the random sequences or processes are ergodic, their covariance functions can be computed from time averages of individual time histories. Let $x_i$ and $y_j$ be the $i^{th}$ and $j^{th}$ samples in ergodic, zero-mean random sequences. Their auto- and cross-covariance functions can be computed as

$$\phi_{xx}(k) = E(x_n x_{n+k}) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} x_n x_{n+k} \tag{2.4-82}$$

$$\phi_{yy}(k) = E(y_n y_{n+k}) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} y_n y_{n+k} \tag{2.4-83}$$

$$\phi_{xy}(k) = E(x_n y_{n+k}) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} x_n y_{n+k} \tag{2.4-84}$$

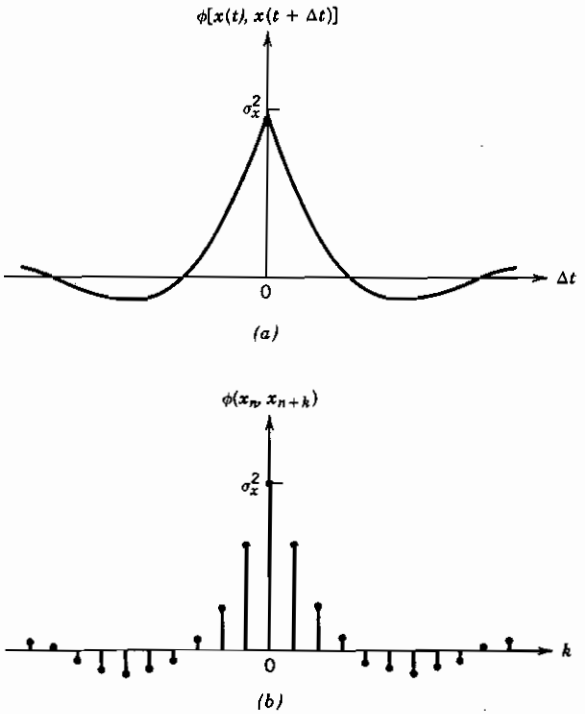


**FIGURE 2.4-5** Autocovariance functions of random processes. (*a*) autocovariance function of a stationary random process; (*b*) autocovariance function of a stationary random sequence.



With sampling interval $\tau$, the "lag" $\Delta t$ is $k\tau$, and the covariance functions are defined only for discrete increments of the time lag (Fig. 2.4-5*b*). As these sequences are stationary, the lower limit of summation could be taken as $-N$, and each sum then would be divided by $2N-1$.

The auto- and cross-covariance functions of ergodic, zero-mean random processes are computed using integration rather than summation:

$$\phi_{xx}(\Delta t) = E[x(t)x(t+\Delta t)] = \lim_{T\to\infty} \frac{1}{T} \int_0^T x(t)x(t+\Delta t)\, dt \tag{2.4-85}$$

$$\phi_{yy}(\Delta t) = E[y(t)y(t+\Delta t)] = \lim_{T\to\infty} \frac{1}{T} \int_0^T y(t)y(t+\Delta t)\, dt \tag{2.4-86}$$

$$\phi_{xy}(\Delta t) = E[x(t)y(t+\Delta t)] = \lim_{T\to\infty} \frac{1}{T} \int_0^T x(t)y(t+\Delta t)\, dt \tag{2.4-87}$$

The covariance functions of random processes are continuous functions of the lag, and they are expressed as *convolution integrals* of the dependent variables. As in the discrete case, the lower limit of integration could be taken as $-T$; then the integral would be divided by $2T$ rather than $T$.

Two autocovariance functions are of particular interest. The first deals with a variable that is totally uncorrelated with itself from one instant to the next, and it must be handled differently for sequences and processes. If $\phi_{xx}(k)$ or $\phi_{xx}(\Delta t)$ is zero except when its argument is zero, the random sequence or process is called *white noise*, in analogy to the wide-band thermal or "shot" noise produced in electronic amplifiers. As will be seen, "white" connotes equal power at all frequencies in the power spectral density function. For the white random *sequence*, the zero value is the variance of the process:

$$\phi_{xx}(0) = \sigma_x^2 \tag{2.4-88}$$

The notion that the autocovariance function of the white random *process* is zero except at zero lag causes a problem, in that $\Delta t$ can become infinitesimally small, yet $\phi_{xx}(\Delta t \neq 0) = 0$. The problem is resolved by defining the autocovariance function of continuous white noise to be

$$\phi_{xx}(\Delta t) = \phi_{xx}(0)\delta(\Delta t) \tag{2.4-89}$$

Here $\delta(\Delta t)$ is the Dirac delta function, which has these characteristics:

$$\delta(\Delta t) = \begin{cases} \infty, & \Delta t = 0 \\ 0, & \Delta t \neq 0 \end{cases} \tag{2.4-90}$$

$$\lim_{\Delta t_1 \to 0} \int_{-\Delta t_1}^{\Delta t_1} \delta(\Delta t) d(\Delta t) = 1 \tag{2.4-91}$$

The value of $\phi_{xx}(0)$ will be deduced in the following section.

A further expression of the independence of white noise derives from the conditional probability density function for sequential samples of the process, which is

$$\text{pr}[x(t) \mid x(t+\Delta t)] = \text{pr}[x(t)] \tag{2.4-92}$$

for any $\Delta t$. In other words, the amplitudes of contiguous values of $x$ are unrelated, even from a probabilistic standpoint.

The second autocovariance function of interest belongs to the Markov sequence,* which is discussed further in Section 4.2. The Markov sequence provides the simplest example of *colored noise*, with decreasing power at increasing frequencies. Let $w_i$ be a zero-mean white noise sequence with covariance $\sigma_w^2$, and assume that $w_i$ is the input to a first-order difference equation with sampling index $i$:

$$x_{i+1} = ax_i + \sqrt{1-a^2}\, w_i, \qquad 0 \leq a \leq 1 \tag{2.4-93}$$

Squaring and taking the expected values,

$$E(x_{i+1}^2) = a^2 E(x_i^2) + 2a\sqrt{1-a^2}\, E(x_i w_i) + (1-a^2) E(w_i^2) \tag{2.4-94}$$

$E(x_i w_i) = 0$ because $w_i$ has no instantaneous effect on $x_i$. In *stochastic equilibrium*, Eq. 2.4-94 reaches steady state [i.e., $E(x_{i+1}^2) = E(x_i^2)$], leaving

$$E(x_i^2) = \sigma_x^2 = \sigma_w^2 \tag{2.4-95}$$

Substituting Eq. 2.4-93 in Eq. 2.4-82,

$$E(x_i x_{i+1}) = aE(x_i^2) = a\sigma_x^2 \tag{2.4-96}$$

so the autocovariance function of the Markov sequence (Fig. 2.4-6*a*) is

$$\phi_{xx}(k) = a^{|k|}\sigma_x^2 = a^{|k|}\sigma_w^2 \tag{2.4-97}$$

*Together with Chebyshev and Lyapunov, Russian mathematician Andrei Markov (1856–1922) laid the foundations of modern probability theory, studying sequences of variables and developing the methods of moments.



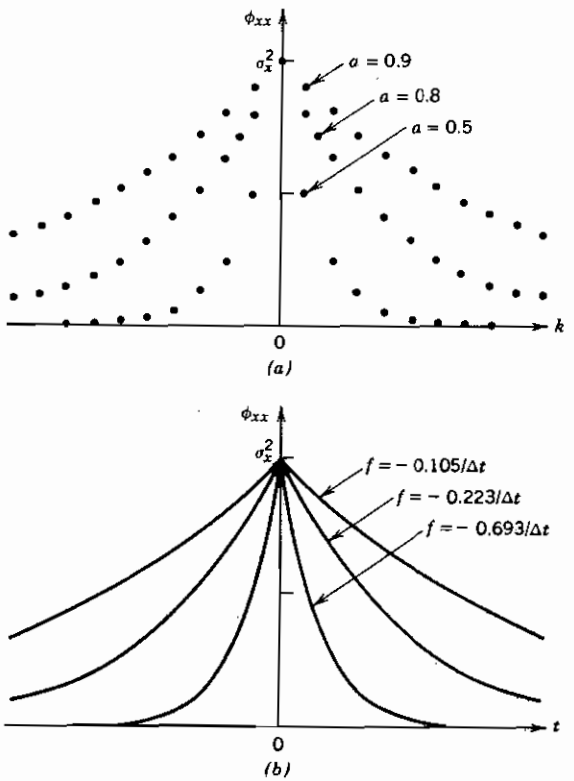


**FIGURE 2.4-6** Autocovariance functions of Markov sequences and processes. (Inverse correlation time, $f = -\ln a/\Delta t$.) (*a*) Autocovariance functions of Markov sequences; (*b*) autocovariance functions of Markov processes.

Similar equations can be derived for the Markov process. Using the limiting arguments of Section 2.3, as $\Delta t \to 0$, Eq. 2.4-93 and 2.4-94 become

$$\dot{x}(t) = fx(t) - fw(t), \qquad f < 0 \tag{2.4-98}$$

$$\frac{d[\sigma_x^2(t)]}{dt} = 2f[\sigma_x^2(t) - \sigma_w^2] \tag{2.4-99}$$

where $a = e^{f\Delta t}$ and $\sigma_x^2(t) = E[x^2(t)]$. In stochastic equilibrium, $\sigma_x^2 = \sigma_w^2$. From Eq. 2.3-81b and 2.4-85,

$$\begin{aligned} \phi_{xx}(\Delta t) &= E[x(t)x(t+\Delta t)] \\ &= E\left[e^{f\Delta t}x^2(t) + x(t)\int_t^{t+\Delta t} e^{f(\Delta t - \tau)} fw\, d\tau\right] \\ &= e^{f|\Delta t|}\sigma_x^2 = e^{f|\Delta t|}\sigma_w^2 \end{aligned} \tag{2.4-100}$$

The expected value of $x(t)$ times the forcing integral is zero, and the absolute value of $\Delta t$ accounts for the symmetry of the autocovariance function. In this form, $-1/f$ is called the *correlation time* of the Markov process (Fig. 2.4-6*b*). Even though the Markov process is driven by white noise, it does not suffer from the white noise's singularity problem. The exponential autocorrelation function corresponds to "band limiting," which is presented in greater detail later.

A time-varying Markov process could be defined with $f = f(t)$ and constant $\sigma_w^2$. This process would be nonstationary, but its statistics could be derived without reference to entire ensembles (Section 4.2).

### Fourier Series and Integrals

Dynamic processes occur during a period of time, and it is most natural to express their governing equations as differential or difference equations which use time as the independent variable. Such equations are said to be written in the *time domain*, and they are the most general mathematical models for physical phenomena, admitting nonlinear and nonstationary effects with relative ease. However, solutions and insight are sometimes difficult to capture with time-domain formulations. The difficulties are reduced somewhat if prior knowledge indicates that solutions can be described by canonical functions, such as exponentials, sinusoids, or other waveforms. Having chosen a set of functions, it remains only to express the solution as some combination of these functions, with suitable multipliers to fit these functions to the problem at hand.

If we choose sinusoids as the canonical functions, then the solutions will be expressed in terms of amplitude, phase angle, and frequency of sine waves, and the equations will be written in the *frequency domain*. Frequency-domain methods are not restricted to stochastic problems, and our review begins by considering the application of Fourier series to strictly periodic signals.* These results are then extended to continuing aperiodic signals (including random processes), culminating in expressions for power spectral density functions, which portray the "frequency content" of signals, and cross-spectral density functions, which indicate the correlation and phasing of two signals as functions of frequency. Although nonstationary methods can be developed, Fourier analysis is most readily applied to stationary processes.

If $x(t)$ is periodic and continuous in time, repeating itself every $T$ sec, then it can be characterized precisely in any time interval $[-T/2, T/2]$ by

*Baron Jean B. J. Fourier (1768–1830) was a professor of mathematics at the Ecole Polytechnique in Paris, later accompanying Napoleon I to Egypt and serving as prefect of the department of Isere in France.



an infinite series of sines and cosines,

$$x(t) = \frac{a(0)}{2} + \sum_{n=1}^{\infty} [a(n\omega_o)\cos n\omega_o t + b(n\omega_o)\sin n\omega_o t] \quad (2.4\text{-}101)$$

where $a(n\omega_o)$ and $b(n\omega_o)$ are chosen to fit $x(t)$. The fundamental frequency, $\omega_o$, is based upon the period $T$:

$$\omega_o = \frac{2\pi}{T}\,(\text{rad/s}) \quad (2.4\text{-}102)$$

$$a(n\omega_o) = \frac{2}{T}\int_{-T/2}^{T/2} x(t)\cos n\omega_o t\,dt, \qquad n = 0 \text{ to } \infty \quad (2.4\text{-}103)$$

$$b(n\omega_o) = \frac{2}{T}\int_{-T/2}^{T/2} x(t)\sin n\omega_o t\,dt, \qquad n = 0 \text{ to } \infty \quad (2.4\text{-}104)$$

This is the basic definition of a *Fourier series*. Using a trigonometric identity, Eq. 2.4-101 can be written in terms of amplitudes $c(n\omega_o)$ and phase angles $\xi(n\omega_o)$,

$$x(t) = \frac{a(0)}{2} + \sum_{n=1}^{\infty} c(n\omega_o)\cos[n\omega_o t + \xi(n\omega_o)], \quad (2.4\text{-}105)$$

with

$$c(n\omega_o) = \sqrt{a^2(n\omega_o) + b^2(n\omega_o)}, \qquad n = 1 \text{ to } \infty \quad (2.4\text{-}106)$$

$$\xi(n\omega_o) = \tan^{-1}\left[\frac{-b(n\omega_o)}{a(n\omega_o)}\right], \qquad n = 1 \text{ to } \infty \quad (2.4\text{-}107)$$

The Fourier series also ean be expressed in complex-variable form as

$$x(t) = \frac{1}{T}\sum_{n=-\infty}^{\infty} g(n\omega_o)\, e^{jn\omega_o t} \quad (2.4\text{-}108)$$

where $j = \sqrt{-1}$, $g$ is a complex number,

$$g(n\omega_o) = \begin{cases} T[a(n\omega_o) + jb(n\omega_o)]/2, & 0 \le n \le \infty \\ T[a(-n\omega_o) - jb(-n\omega_o)]/2, & -\infty \le n \le -1 \end{cases} \quad (2.4\text{-}109)$$

and $b(0)$ is zero. The complex coefficients can be written directly, using Eq. 2.4-103 and 2.4-104:

$$g(n\omega_o) = \int_{-T/2}^{T/2} x(t)\, e^{-jn\omega_o t}\,dt \quad (2.4\text{-}110)$$