APPENDIX H

# INTRODUCTION TO PROBABILITY AND RANDOM PROCESSES

This appendix is not intended to be a definitive dissertation on the subject of random processes. The major concepts, definitions, and results which are employed in the text are stated here with little discussion and no proof. The reader who requires a more complete presentation of this material is referred to any one of several excellent books on the subject: among them Davenport and Root (Ref. 2), Laning and Battin (Ref. 3), and Lee (Ref. 4). Possibly the most important function served by this appendix is the definition of the notation and of certain conventions used in the text.

## PROBABILITY

Consider an event $E$ which is a possible outcome of a random experiment. We denote by $P(E)$ the *probability* of this event, and think of it intuitively as the limit, as the number of trials becomes large, of the ratio of the number of times $E$ occurred to the number of times the experiment was tried. The joint event that $A$ and $B$ and $C$, etc., occurred is denoted by $ABC \cdots$, and the probability of this joint event, by $P(ABC\cdots)$. If these events $A$, $B$, $C$, etc., are mutually *independent*, which means that the occurrence of any one of them bears no relation to the occurrence of any other, the probability of the joint event is the product of the probabilities of the simple events. That is,

$$P(ABC\cdots) = P(A)P(B)P(C)\cdots \tag{H-1}$$

if the events $A$, $B$, $C$, etc., are mutually independent. Actually, the mathematical definition of independence is the reverse of this statement, but the result of consequence is that independence of events and the multiplicative property of probabilities go together.

## RANDOM VARIABLES

A random variable $X$ is in simplest terms a variable which takes on values at random; it may be thought of as a function of the outcomes of some random experiment. The manner of specifying the probability with which different values are taken by the random variable is by the *probability distribution function* $F(x)$, which is defined by

$$F(x) = P(X \leq x) \tag{H-2}$$

or by the *probability density function* $f(x)$, which is defined by

$$f(x) = \frac{dF(x)}{dx} \tag{H-3}$$

The inverse of the defining relation for the probability density function is

$$F(x) = \int_{-\infty}^{x} f(u)\, du \tag{H-4}$$

An evident characteristic of any probability distribution or density function is

$$F(\infty) = \int_{-\infty}^{\infty} f(u)\, du = 1 \tag{H-5}$$

From the definition, the interpretation of $f(x)$ as the density of probability of the event that $X$ takes a value in the vicinity of $x$ is clear.

$$\begin{aligned} f(x) &= \lim_{dx \to 0} \frac{F(x + dx) - F(x)}{dx} \\ &= \lim_{dx \to 0} \frac{P(x < X \leq x + dx)}{dx} \end{aligned} \tag{H-6}$$

This function is finite if the probability that $X$ takes a value in the infinitesimal interval between $x$ and $x + dx$ (the interval closed on the right) is an infinitesimal of order $dx$. This is usually true of random variables which take values over a continuous range. If, however, $X$ takes a set of discrete values $x_i$ with nonzero probabilities $p_i$, $f(x)$ is infinite at these values of $x$. This is accommodated by a set of delta functions weighted by the appropriate probabilities.

$$f(x) = \sum_i p_i\, \delta(x - x_i) \tag{H-7}$$

A suitable definition of the *delta function*, $\delta(x)$, for the present purpose is a function which is zero everywhere except at $x = 0$, and infinite at that point in such a way that the integral of the function across the singularity is unity. An important property of the delta function which follows from this definition is

$$\int_{-\infty}^{\infty} G(x)\,\delta(x - x_0)\,dx = G(x_0) \qquad \text{(H-8)}$$

if $G(x)$ is a finite-valued function which is continuous at $x = x_0$.

A random variable may take values over a continuous range and, in addition, take a discrete set of values with nonzero probability. The resulting probability density function includes both a finite function of $x$ and an additive set of probability-weighted delta functions; such a distribution is called *mixed.*

The simultaneous consideration of more than one random variable is often necessary or useful. In the case of two, the probability of the occurrence of pairs of values in a given range is prescribed by the *joint probability distribution function.*

$$F_2(x,y) = P(X \leq x \text{ and } Y \leq y) \qquad \text{(H-9)}$$

where $X$ and $Y$ are the two random variables under consideration. The corresponding *joint probability density function* is

$$f_2(x,y) = \frac{\partial^2 F_2(x,y)}{\partial x\,\partial y} \qquad \text{(H-10)}$$

It is clear that the individual probability distribution and density functions for $X$ and $Y$ can be derived from the joint distribution and density functions. For the distribution of $X$,

$$F(x) = F_2(x,\infty) \qquad \text{(H-11)}$$

$$f(x) = \int_{-\infty}^{\infty} f_2(x,y)\,dy \qquad \text{(H-12)}$$

Corresponding relations give the distribution of $Y$. These concepts extend directly to the description of the joint characteristics of more than two random variables.

If $X$ and $Y$ are independent, the event $X \leq x$ is independent of the event $Y \leq y$; thus the probability for the joint occurrence of these events is the product of the probabilities for the individual events. Equation (H-9) then gives

$$\begin{aligned} F_2(x,y) &= P(X \leq x \text{ and } Y \leq y) \\ &= P(X \leq x)P(Y \leq y) \\ &= F_X(x)F_Y(y) \end{aligned} \qquad \text{(H-13)}$$

From Eq. (H-10) the joint probability density function is, then,

$$f_2(x,y) = f_X(x)f_Y(y) \tag{H-14}$$

**Expectations and statistics of random variables** The *expectation* of a random variable is defined in words to be the sum of all values the random variable may take, each weighted by the probability with which the value is taken. For a random variable which takes values over a continuous range, this summation is done by integration. The probability, in the limit as $dx \to 0$, that $X$ takes a value in the infinitesimal interval of width $dx$ near $x$ is given by Eq. (H-6) to be $f(x)\,dx$. Thus the expectation of $X$, which we denote by $\bar{X}$, is

$$\bar{X} = \int_{-\infty}^{\infty} xf(x)\,dx \tag{H-15}$$

This is also called the *mean value* of $X$, or the mean of the distribution of $X$. This is a precisely defined number toward which the average of a number of observations of $X$ tends, in the probabilistic sense, as the number of observations becomes large. Equation (H-15) is the analytic definition of the expectation, or mean, of a random variable. This expression is usable for random variables having a continuous, discrete, or mixed distribution if the set of discrete values which the random variable takes is represented by impulses in $f(x)$ according to Eq. (H-7).

It is of frequent importance to find the expectation of a function of a random variable. If $Y$ is defined to be some function of the random variable $X$, say, $Y = g(X)$, then $Y$ is itself a random variable with a distribution derivable from the distribution of $X$. The expectation of $Y$ is defined by Eq. (H-15), where the probability density function for $Y$ would be used in the integral. Fortunately, this procedure can be abbreviated. The expectation of any function of $X$ can be calculated directly from the distribution of $X$ by the integral

$$\overline{g(X)} = \int_{-\infty}^{\infty} g(x)f(x)\,dx \tag{H-16}$$

An important statistical parameter descriptive of the distribution of $X$ is its *mean-squared value*. Using Eq. (H-16), the expectation of the square of $X$ is written

$$\overline{X^2} = \int_{-\infty}^{\infty} x^2f(x)\,dx \tag{H-17}$$

The *variance* of a random variable is the mean-squared deviation of the random variable from its mean; it is denoted by $\sigma^2$.

$$\begin{aligned}\sigma^2 &= \int_{-\infty}^{\infty} (x - \bar{X})^2f(x)\,dx \\ &= \overline{X^2} - \bar{X}^2\end{aligned} \tag{H-18}$$

The square root of the variance, or $\sigma$, is called the *standard deviation* of the random variable.

Other functions whose expectations we shall wish to calculate are sums and products of random variables. It is easily shown that the expectation of the sum of random variables is equal to the sum of the expectations,

$$\overline{X_1 + X_2 + \cdots + X_n} = \bar{X}_1 + \bar{X}_2 + \cdots + \bar{X}_n \tag{H-19}$$

whether or not the variables are independent, and that the expectation of the product of random variables is equal to the product of the expectations,

$$\overline{X_1 X_2 \cdots X_n} = \bar{X}_1 \bar{X}_2 \cdots \bar{X}_n \tag{H-20}$$

if the variables are independent. It is also true that the variance of the sum of random variables is equal to the sum of the variances if the variables are independent.

A very important concept is that of statistical dependence between random variables. A partial indication of the degree to which one variable is related to another is given by the *covariance*, which is the expectation of the product of the deviations of two random variables from their means.

$$\begin{aligned}\overline{(X - \bar{X})(Y - \bar{Y})} &= \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy\,(x - \bar{X})(y - \bar{Y}) f_2(x,y) \\ &= \overline{XY} - \bar{X}\bar{Y}\end{aligned} \tag{H-21}$$

This covariance, normalized by the standard deviations of $X$ and $Y$, is called the *correlation coefficient*, and is denoted $\rho$.

$$\rho = \frac{\overline{XY} - \bar{X}\bar{Y}}{\sigma_X \sigma_Y} \tag{H-22}$$

The correlation coefficient is a measure of the degree of linear dependence between $X$ and $Y$. If $X$ and $Y$ are independent, $\rho$ is zero; if $Y$ is a linear function of $X$, $\rho$ is $\pm 1$. If an attempt is made to approximate $Y$ by some linear function of $X$, the minimum possible mean-squared error in the approximation is $\sigma_y^2(1 - \rho^2)$. This provides another interpretation of $\rho$ as a measure of the degree of linear dependence between random variables.

One additional function associated with the distribution of a random variable which should be introduced is the *characteristic function*. It is defined by

$$\begin{aligned}g(t) &= \overline{\exp(jtX)} \\ &= \int_{-\infty}^{\infty} \exp(jtx) f(x)\,dx\end{aligned} \tag{H-23}$$

A property of the characteristic function which largely explains its value is that the characteristic function of a sum of independent random variables is the product of the characteristic functions of the individual variables. If the characteristic function of a random variable is known, the probability density function can be determined from

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} g(t) \exp(-jtx)\, dt \tag{H-24}$$

Notice that Eqs. (H-23) and (H-24) are in the form of a Fourier transform pair. Another useful relation is

$$\overline{X^n} = j^{-n} \left. \frac{d^n g(t)}{dt^n} \right|_{t=0} \tag{H-25}$$

**The uniform and normal probability distributions** Two specific forms of probability distribution which are referred to in the text are the uniform distribution and the normal distribution. The *uniform distribution* is characterized by a uniform (constant) probability density over some finite interval. The magnitude of the density function in this interval is the reciprocal of the interval width as required to make the integral of the function unity. This function is pictured in Fig. H-1. The *normal probability density* function, shown in Fig. H-2, has the analytic form

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-m)^2}{2\sigma^2}\right] \tag{H-26}$$

where the two parameters which define the distribution are $m$, the mean, and $\sigma$, the standard deviation. By calculating the characteristic function for a normally distributed random variable, one can immediately show that the distribution of the sum of independent normally distributed variables is also normal. Actually, this remarkable property of preservation of form of the distribution is true of the sum of normally distributed random variables whether they are independent or not. Even more remarkable is the fact that under certain circumstances the distribution of the sum of independent random variables, each having an *arbitrary* distribution, tends toward the normal distribution as the number of variables in the sum tends toward infinity. This statement, together with the conditions under which the result can be proved, is known as the *central limit theorem.* The conditions are rarely tested in practical situations, but the empirically observed fact is that a great many random variables—and especially those encountered by control-system engineers—display a distribution which closely approximates the normal. The reason for the common occurrence of normally distributed random variables is certainly stated in the central limit theorem.

Reference is made in the text to two random variables which possess a *bivariate normal distribution.* The form of the joint probability density function for such zero-mean variables is

$$f_2(x_1,x_2) = \frac{1}{2\pi\sqrt{m_{11}m_{22} - m_{12}^2}} \exp\left[-\frac{m_{22}x_1^2 - 2m_{12}x_1x_2 + m_{11}x_2^2}{2(m_{11}m_{22} - m_{12}^2)}\right]$$

where $m_{ij} = \overline{X_iX_j}$.

This can also be written in terms of statistical parameters previously defined as

$$f_2(x_1,x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{\dfrac{x_1^2}{\sigma_1^2} - 2\rho\dfrac{x_1}{\sigma_1}\dfrac{x_2}{\sigma_2} + \dfrac{x_2^2}{\sigma_2^2}}{2(1-\rho^2)}\right] \tag{H-27}$$

## RANDOM PROCESSES

A random process may be thought of as a collection, or *ensemble*, of functions of time, any one of which might be observed on any trial of an experiment. The ensemble may include a finite number, a countable infinity, or a noncountable infinity of such functions. We shall denote the ensemble of functions by $\{x(t)\}$, and any observed member of the ensemble by $x(t)$. The value of the observed member of the ensemble at a particular time, say, $t_1$, as shown in Fig. H-3, is a random variable; on repeated trials of the experiment, $x(t_1)$ takes different values at random. The probability that $x(t_1)$ takes values in a certain range is given by the probability distribution function, as it is for any random variable. In this case we show explicitly in the notation the dependence on the time of observation.

$$F(x_1,t_1) = P[x(t_1) \le x_1] \tag{H-28}$$

The corresponding probability density function is

$$f(x_1,t_1) = \frac{dF(x_1,t_1)}{dx_1} \tag{H-29}$$

These functions suffice to define, in a probabilistic sense, the range of amplitudes which the random process displays. To gain a sense of how quickly varying the members of the ensemble are likely to be, one has to observe the same member function at more than one time. The probability for the occurrence of a pair of values in certain ranges is given by the second-order joint probability distribution function

$$F_2(x_1,t_1;x_2,t_2) = P[x(t_1) \le x_1 \text{ and } x(t_2) \le x_2] \tag{H-30}$$

and the corresponding joint probability density function

$$f_2(x_1,t_1;x_2,t_2) = \frac{\partial^2 F_2(x_1,t_1;x_2,t_2)}{\partial x_1 \, \partial x_2} \tag{H-31}$$

Higher-ordered joint distribution and density functions can be defined following this pattern, but only rarely does one attempt to deal with more than the second-order statistics of random processes.

If two random processes are under consideration, the simplest distribution and density functions which give some indication of their joint statistical characteristics are the second-order functions

$$F_2(x,t_1;y,t_2) = P[x(t_1) \le x \text{ and } y(t_2) \le y] \tag{H-32}$$

$$f_2(x,t_1;y,t_2) = \frac{\partial^2 F_2(x,t_1;y,t_2)}{\partial x \, \partial y} \tag{H-33}$$

Actually, the characterization of random processes, in practice, is usually limited to even less information than that given by the second-order distribution or density functions. Only the first moments of these distributions are commonly measured. These moments are called auto- and cross-correlation functions. The *autocorrelation function* is defined as

$$\varphi_{xx}(t_1,t_2) = \overline{x(t_1)x(t_2)} = \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2 \, x_1 x_2 f_2(x_1,t_1;x_2,t_2) \tag{H-34}$$

and the *cross-correlation function* as

$$\varphi_{xy}(t_1,t_2) = \overline{x(t_1)y(t_2)} = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \, xy f_2(x,t_1;y,t_2) \tag{H-35}$$

In the case where $\overline{x(t_1)}$, $\overline{x(t_2)}$, and $\overline{y(t_2)}$ are all zero, these correlation functions are the covariances of the indicated random variables. If they are then normalized by the corresponding standard deviations, according to Eq. (H-22), they become correlation coefficients which measure on a scale from $-1$ to $+1$ the degree of linear dependence between the variables.

A *stationary random process* is one whose statistical properties are invariant in time. This implies that the first probability density function for the process, $f(x_1,t_1)$, is independent of the time of observation $t_1$. Then all the moments of this distribution, such as $\overline{x(t_1)}$ and $\overline{x(t_1)^2}$, are also independent of time; they are constants. The second probability density function is not in this case dependent on the absolute times of observation, $t_1$ and $t_2$, but still depends on the difference between them. So if $t_2$ is written as

$$t_2 = t_1 + \tau \tag{H-36}$$

$f_2(x_1,t_1;x_2,t_2)$ becomes $f_2(x_1, t_1; x_2, t_1 + \tau)$, which is independent of $t_1$, but still a function of $\tau$. The correlation functions are then functions only of the single variable $\tau$.

$$\varphi_{xx}(\tau) = \overline{x(t_1)x(t_1 + \tau)} \tag{H-37}$$

$$\varphi_{xy}(\tau) = \overline{x(t_1)y(t_1 + \tau)} \tag{H-38}$$

Both of these are independent of $t_1$ if the random processes are stationary. We note the following properties of these correlation functions:

$$\varphi_{xx}(-\tau) = \varphi_{xx}(\tau) \tag{H-39}$$

$$\varphi_{xy}(-\tau) = \varphi_{yx}(\tau) \tag{H-40}$$

One further concept associated with stationary random processes is the *ergodic hypothesis*. This hypothesis claims that any statistic calculated by averaging over all members of an ergodic ensemble at a fixed time can also be calculated by averaging over all time on a single representative member of the ensemble. The key to this notion is the word "representative." If a particular member of the ensemble is to be statistically representative of all, it must display at various points in time the full range of amplitude, rate of change of amplitude, etc., which are to be found among all the members of the ensemble. A classic example of a stationary ensemble which is not ergodic is the ensemble of constant functions. The failing in this case is that no member of the ensemble is representative of all. In practice, almost all empirical results for stationary processes are derived from tests on a single function under the assumption that the ergodic hypothesis holds. In this case the common statistics associated with a random process are written

$$\bar{x} = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t)\, dt \tag{H-41}$$

$$\overline{x^2} = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t)^2\, dt \tag{H-42}$$

$$\varphi_{xx}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t)x(t + \tau)\, dt \tag{H-43}$$

$$\varphi_{xy}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t)y(t + \tau)\, dt \tag{H-44}$$

An example of a stationary ergodic random process is the ensemble of sinusoids of given amplitude and frequency with a uniform distribution of phase. The member functions of this ensemble are all of the form

$$x(t) = A \sin(\omega t + \theta) \tag{H-45}$$

where $\theta$ is a random variable having the uniform distribution over the interval $(0,2\pi)$ radians. Any average taken over the members of this ensemble at any fixed time would find all phase angles represented with equal probability density. But the same is true of an average over all time on any one member. For this process, then, all members of the ensemble qualify as "representative." Note that any distribution of the phase angle $\theta$ other than the uniform distribution over an integral number of cycles would define a nonstationary process.

Another random process which plays a central role in the text is the *gaussian process*, which is characterized by the property that its joint probability distribution functions of all orders are multidimensional normal distributions. For a gaussian process, then, the distribution of $x(t)$ for any $t$ is the normal distribution, for which the density function is expressed by Eq. (H-26); the joint distribution of $x(t_1)$ and $x(t_2)$ for any $t_1$ and $t_2$ is the bivariate normal distribution of Eq. (H-27), and so on for the higher-ordered joint distributions. The $n$-dimensional normal distribution for zero-mean variables is specified by the elements of the $n$th-order covariance matrix, that is, by the $m_{ij} = \overline{X_i X_j}$ for $i, j = 1, 2, \ldots, n$. But in this case

$$X_i = x(t_i)$$

and

$$m_{ij} = \overline{x(t_i)x(t_j)} = \varphi_{xx}(t_i,t_j)$$

Thus all the statistics of a gaussian process are defined by the autocorrelation function for the process. This property is clearly a great boon to analytic operations.

## LINEAR SYSTEMS

The input-output relation for a linear system may be written

$$y(t) = \int_{-\infty}^{t} x(\tau)w(t,\tau)\,d\tau \tag{H-46}$$

where $x(t)$ = input function
$y(t)$ = output
$w(t,\tau)$ = system *weighting function*, the response at time $t$ to a unit impulse input at time $\tau$.

Using this relation, the statistics of the output process can be written in terms of those of the input.

$$\overline{y(t)} = \int_{-\infty}^{t} \overline{x(\tau)}\, w(t,\tau)\,d\tau \tag{H-47}$$

$$\overline{y(t)^2} = \int_{-\infty}^{t} d\tau_1\, w(t,\tau_1) \int_{-\infty}^{t} d\tau_2\, w(t,\tau_2)\varphi_{xx}(\tau_1,\tau_2) \tag{H-48}$$

$$\varphi_{yy}(t_1,t_2) = \int_{-\infty}^{t_1} d\tau_1\, w(t_1,\tau_1) \int_{-\infty}^{t_2} d\tau_2\, w(t_2,\tau_2)\varphi_{xx}(\tau_1,\tau_2) \tag{H-49}$$

$$\varphi_{xy}(t_1,t_2) = \int_{-\infty}^{t_2} w(t_2,\tau)\varphi_{xx}(t_1,\tau)\, d\tau \tag{H-50}$$

If the input process is stationary and the system time-invariant, the output process is also stationary in the steady state. These expressions then reduce to

$$y(t) = \int_{0}^{\infty} w(\tau)x(t-\tau)\, d\tau \tag{H-51}$$

$$\bar{y} = \bar{x}\int_{0}^{\infty} w(t)\, dt \tag{H-52}$$

$$\overline{y^2} = \int_{0}^{\infty} d\tau_1\, w(\tau_1) \int_{0}^{\infty} d\tau_2\, w(\tau_2)\varphi_{xx}(\tau_1-\tau_2) \tag{H-53}$$

$$\varphi_{yy}(\tau) = \int_{0}^{\infty} d\tau_1\, w(\tau_1) \int_{0}^{\infty} d\tau_2\, w(\tau_2)\varphi_{xx}(\tau+\tau_1-\tau_2) \tag{H-54}$$

$$\varphi_{xy}(\tau) = \int_{0}^{\infty} w(\tau_1)\varphi_{xx}(\tau-\tau_1)\, d\tau_1 \tag{H-55}$$

Analytic operations on linear invariant systems are facilitated by the use of integral transforms which transform the convolution input-output relation of Eq. (H-51) into the algebraic operation of multiplication. Since members of stationary random ensembles must necessarily be visualized as existing for all negative and positive time, the two-sided Fourier transform is the appropriate transformation to employ in this instance. The Fourier transforms of the correlation functions defined above then appear quite naturally in analysis. The Fourier transform of the autocorrelation function

$$\Phi_{xx}(\omega) = \int_{-\infty}^{\infty} \varphi_{xx}(\tau)\exp(-j\omega\tau)\, d\tau \tag{H-56}$$

is called the *power spectral density function*, or power density spectrum of the random process $\{x(t)\}$. The term "power" is here used in a generalized sense, indicating the expected squared value of the members of the ensemble. $\Phi_{xx}(\omega)$ is indeed the spectral distribution of power density for $\{x(t)\}$ in that integration of $\Phi_{xx}(\omega)$ over frequencies in the band from $\omega_1$ to $\omega_2$ yields the mean-squared value of the process which consists only of those harmonic components of $\{x(t)\}$ that lie between $\omega_1$ and $\omega_2$. In particular, the mean-squared value of $\{x(t)\}$ itself is given by integration of the power density spectrum for the random process over the full range of $\omega$. This last result is seen as a specialization of the inverse transform relation corresponding to Eq. (H-56).

$$\varphi_{xx}(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \Phi_{xx}(\omega) \exp(j\omega\tau)\, d\omega \tag{H-57}$$

$$\overline{x^2} = \varphi_{xx}(0) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \Phi_{xx}(\omega)\, d\omega \tag{H-58}$$

The input-output relation for power spectral density functions is derived by calculating the Fourier transform of the autocorrelation function of the output of a linear invariant system as expressed by Eq. (H-54).

$$\begin{aligned}\Phi_{yy}(\omega) &= W(j\omega)W(-j\omega)\Phi_{xx}(\omega)\\ &= |W(j\omega)|^2\, \Phi_{xx}(\omega)\end{aligned} \tag{H-59}$$

where $W(j\omega)$ is the steady-state *sinusoidal response function* for the system, which is also the Fourier transform of the system weighting function.

$$W(j\omega) = \int_{-\infty}^{\infty} w(t) \exp(-j\omega t)\, dt \tag{H-60}$$

A particularly simple form for the power density spectrum is a constant, $\Phi_{xx}(\omega) = \Phi_0$. This implies that power density is distributed uniformly over all frequency components in the full infinite range. By analogy with the corresponding situation in the case of white light, such a random process, usually a noise, is called *white noise*. The autocorrelation function for white noise is a delta function.

$$\begin{aligned}\varphi_{nn}(\tau) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \Phi_0 \exp(j\omega\tau)\, d\omega\\ &= \Phi_0\, \delta(\tau)\end{aligned} \tag{H-61}$$

The mean-squared value of white noise, $\varphi_{nn}(0)$, is infinite, and so the process is not physically realizable. However, it does serve as a very useful approximation to situations in which the noise is wideband compared with the bandwidth of the system, and the concept is useful in analytic operations.

The Fourier transform of the cross-correlation function is called the *cross power spectral density function.*

$$\Phi_{xy}(\omega) = \int_{-\infty}^{\infty} \varphi_{xy}(\tau) \exp(-j\omega\tau)\, d\tau \tag{H-62}$$

If $x(t)$ is the input to, and $y(t)$ the output from, a linear invariant system, so that $\varphi_{xy}(\tau)$ is given by Eq. (H-55), then the input-output cross power density spectrum is

$$\Phi_{xy}(\omega) = W(j\omega)\Phi_{xx}(\omega) \tag{H-63}$$

The calculation of the distribution of amplitudes of the random processes which appear at various points in linear systems is in general a most complicated problem. The only case in which a simple result is known is that of a gaussian process applied to a linear system; in this case it can be proved that the processes appearing at all points in the system are also gaussian. The credibility of this property is perhaps indicated by noting that the input-output relation for a linear system [Eq. (H-46)] is the limit of a sum of the form

$$y(t) = \lim_{\substack{\Delta\tau \to 0 \\ N \to \infty \\ \tau_N = t}} \sum_{i=-\infty}^{N} x(\tau_i) w(t,\tau_i)\, \Delta\tau_i \tag{H-64}$$

We have already noted that the sum of normally distributed random variables is normally distributed, and since any constant times a normally distributed variable is also normal, we may conclude that any linear combination of normally distributed random variables is normally distributed. Equation (H-64) expresses the output of a linear system as the limit of a linear combination of past values of the input. Thus, if the input at all past times has a normal distribution, the output at any time must also be normal. This property also holds for the higher-ordered joint distributions, with the result that if $x(t)$ is a gaussian process, so is $y(t)$.

Of even greater consequence is the empirically observed fact that *non-gaussian inputs tend to become more nearly gaussian as a result of linear filtering*. If the input were nongaussian white noise, one could refer to Eq. (H-64) and invoke the central limit theorem to argue that, as the filtering bandwidth is decreased, the number of terms in the sum for $y(t)$ which make a significant contribution increases, and thus the distribution of $y(t)$ should approach the normal, regardless of the distribution of the $x(\tau_i)$. In fact, this tendency is observed for nonwhite inputs as well; so the gaussian random process has the singular position of being that process toward which many others tend as a result of linear filtering. The most evident exception to this rule is a random ensemble in which every member contains a periodic component of the same period. Low-pass linear filtering tends to reduce these periodic signals to their fundamental sinusoidal components, and a sinusoid does not display the normal distribution of amplitudes. But if every member of a random ensemble has a periodic component of the same period, the process contains nonzero power at the fundamental frequency of these periodic components, and perhaps at some of the harmonics of this frequency as well. Nonzero power at any discrete frequency implies infinite power density at that frequency. Thus it might be said that if a random process has a finite power density spectrum, it may be expected to approach a gaussian process as a result of low-pass linear filtering. Unfortunately, it does not seem possible to phrase this general statement in quantitative terms.

## INSTANTANEOUS NONLINEARITIES

The second-order statistics of the output of a static single-valued nonlinearity depend on the form of the nonlinearity

$$y = y(x) \tag{H-65}$$

and the second-order joint probability density function for the input random process. For example, the autocorrelation function for the output process is

$$\varphi_{yy}(\tau) = \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2\, y(x_1)y(x_2)f_2(x_1, t_1; x_2, t_1 + \tau) \tag{H-66}$$

which is independent of $t_1$ if $\{x(t)\}$ is stationary. If interest is centered on closed-loop systems which contain a nonlinear device, it is rarely possible, as a practical matter, to determine the second-order probability density function for the input to the nonlinearity, unless one can argue, on the basis of linear filtering, that the process should be approximately gaussian. For a stationary gaussian process, $f_2(x_1, t_1; x_2, t_1 + \tau)$ is given by Eq. (H-27) with

$$\sigma_1^2 = \overline{x(t_1)x(t_1)} = \varphi_{xx}(0) = \sigma^2 \tag{H-67}$$

since we are considering zero-mean variables. Also,

$$\sigma_2^2 = \overline{x(t_1 + \tau)x(t_1 + \tau)} = \sigma^2 \tag{H-68}$$

$$\rho = \frac{\overline{x(t_1)x(t_1 + \tau)}}{\sigma_1\sigma_2} = \frac{1}{\sigma^2}\varphi_{xx}(\tau) = \rho_{xx}(\tau) \tag{H-69}$$

Equation (H-27) becomes

$$f_2(x_1, t_1; x_2, t_1 + \tau) = \frac{1}{2\pi\sigma^2\sqrt{1 - \rho^2}} \exp\left[-\frac{x_1^2 - 2\rho x_1 x_2 + x_2^2}{2\sigma^2(1 - \rho^2)}\right] \tag{H-70}$$

Thus, for a gaussian input, Eq. (H-66) is written

$$\begin{aligned}\varphi_{yy}(\tau) &= \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2\, y(x_1)y(x_2)\frac{1}{2\pi\sigma^2\sqrt{1 - \rho^2}} \\ &\qquad \times \exp\left[-\frac{x_1^2 - 2\rho x_1 x_2 + x_2^2}{2\sigma^2(1 - \rho^2)}\right] \\ &= \int_{-\infty}^{\infty} du_1 \int_{-\infty}^{\infty} du_2\, y(\sigma u_1)y(\sigma u_2)\frac{1}{2\pi\sqrt{1 - \rho^2}} \\ &\qquad \times \exp\left[-\frac{u_1^2 - 2\rho u_1 u_2 + u_2^2}{2(1 - \rho^2)}\right]\end{aligned} \tag{H-71}$$

The evaluation of this double integral can be systematized through the expansion of the gaussian joint probability density function into a double series of Hermite polynomials. These functions are used in varying forms; perhaps the most convenient for the present purpose is

$$H_k(u) = (-1)^k \exp\left(\frac{u^2}{2}\right) \frac{d^k}{du^k}\left[\exp\left(-\frac{u^2}{2}\right)\right] \tag{H-72}$$

The first few of these functions are

$$H_0(u) = 1 \tag{H-73a}$$

$$H_1(u) = u \tag{H-73b}$$

$$H_2(u) = u^2 - 1 \tag{H-73c}$$

$$H_3(u) = u^3 - 3u \tag{H-73d}$$

and in general,

$$H_{k+1}(u) = uH_k(u) - kH_{k-1}(u) \tag{H-74}$$

These functions form a complete set and are orthogonal over the doubly infinite interval with respect to the weighting function $\exp(-u^2/2)$. The orthogonality conditions are

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} H_k(u)H_j(u)\exp\left(-\frac{u^2}{2}\right) du = \begin{cases} k! & j = k \\ 0 & j \neq k \end{cases} \tag{H-75}$$

Expansion of the gaussian probability density function in terms of these Hermite polynomials gives (Ref. 1)

$$\frac{1}{\sqrt{1-\rho^2}} \exp\left[-\frac{u_1^2 - 2\rho u_1 u_2 + u_2^2}{2(1-\rho^2)}\right] = \exp\left(-\frac{u_1^2 + u_2^2}{2}\right) \sum_{k=0}^{\infty} H_k(u_1)H_k(u_2)\frac{\rho^k}{k!} \tag{H-76}$$

With this expansion, Eq. (H-71) becomes

$$\varphi_{yy}(\tau) = \frac{1}{2\pi}\sum_{k=0}^{\infty}\frac{\rho^k}{k!}\int_{-\infty}^{\infty} du_1 \int_{-\infty}^{\infty} du_2\, y(\sigma u_1)y(\sigma u_2) \times \exp\left(-\frac{u_1^2 + u_2^2}{2}\right) H_k(u_1)H_k(u_2)$$

$$= \sum_{k=0}^{\infty} a_k^2 \rho_{xx}^k(\tau) \tag{H-77}$$

where

$$a_k = \frac{1}{\sqrt{2\pi k!}}\int_{-\infty}^{\infty} y(\sigma u)\exp\left(-\frac{u^2}{2}\right) H_k(u)\, du \tag{H-78}$$

This expresses the output autocorrelation function as a power series in the normalized input autocorrelation function. Note that $H_k(u)$ is odd for $k$ odd and is even for $k$ even; thus, in the common case of an odd nonlinearity, $a_k = 0$ for $k$ even.

It is clear that the output of the nonlinearity contains power at higher frequencies than the input. The normalized autocorrelation function $\rho_{xx}(\tau)$ has a magnitude less than or equal to unity everywhere; so the powers of this function appearing in Eq. (H-77) fall off to zero faster than the function itself. For example, if the input process has the familiar exponential autocorrelation function, $\rho_{xx}^3(\tau)$ will decay three times faster, as indicated in Fig. H-4. But this contribution to the output autocorrelation function has a power spectral density function which is three times wider than that of the input process. The higher-ordered terms in the expansion of the output autocorrelation function make still broader band contributions to the output power spectral density function. This is in a way analogous to the response of a nonlinear device to a sinusoidal input, which consists of a fundamental frequency component plus higher-frequency harmonics.

Other statistics of the output of the nonlinearity can also be expressed in terms of these $a_k$. The mean of the output is

$$\begin{aligned}\overline{y(t)} &= \int_{-\infty}^{\infty} y(x)f(x,t)\,dx \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} y(\sigma u)\exp\left(-\frac{u^2}{2}\right)du \\ &= a_0 \end{aligned} \tag{H-79}$$

in the case of an unbiased stationary gaussian input. Also, in this case, the input-output cross-correlation function is

$$\begin{aligned}\varphi_{xy}(\tau) &= \frac{1}{2\pi}\sum_{k=0}^{\infty}\frac{\rho^k}{k!}\int_{-\infty}^{\infty}du_1\int_{-\infty}^{\infty}du_2\,\sigma u_1 y(\sigma u_2)\exp\left(-\frac{u_1^2+u_2^2}{2}\right)H_k(u_1)H_k(u_2) \\ &= \sum_{k=0}^{\infty} a_k\rho^k\left[\frac{1}{\sqrt{2\pi}k!}\int_{-\infty}^{\infty}\sigma u_1\exp\left(-\frac{u_1^2}{2}\right)H_k(u_1)\,du_1\right] \\ &= \sigma a_1\rho_{xx}(\tau) \\ &= \frac{a_1}{\sigma}\varphi_{xx}(\tau)\end{aligned} \tag{H-80}$$

using Eq. (H-73*b*) and the orthogonality conditions of Eq. (H-75). The input-output cross-correlation function for a static single-valued nonlinearity with an unbiased gaussian input is thus found to be just a constant times the input autocorrelation function, as it is for a static linear gain. In the linear case, the constant is the gain, whereas in the nonlinear case, the constant depends both on the nonlinearity and on the rms value of the input.

## REFERENCES

1. Cramer, H.: "Mathematical Methods of Statistics," Princeton University Press, Princeton, N.J., 1946, p. 133.
2. Davenport, W. B., Jr., and W. L. Root: "An Introduction to Random Signals and Noise," McGraw-Hill Book Company, New York, 1958.
3. Laning, J. H., Jr., and R. H. Battin: "Random Processes in Automatic Control," McGraw-Hill Book Company, New York, 1956.
4. Lee, Y. W.: "Statistical Theory of Communication," John Wiley & Sons, Inc., New York, 1960.

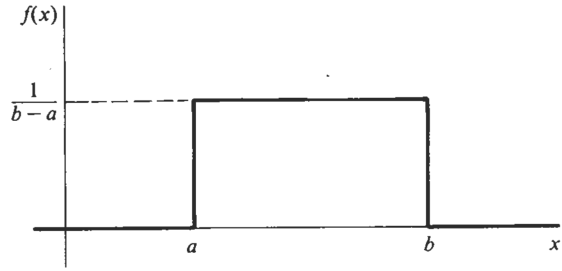



*Figure H-1* *The uniform probability density function.*

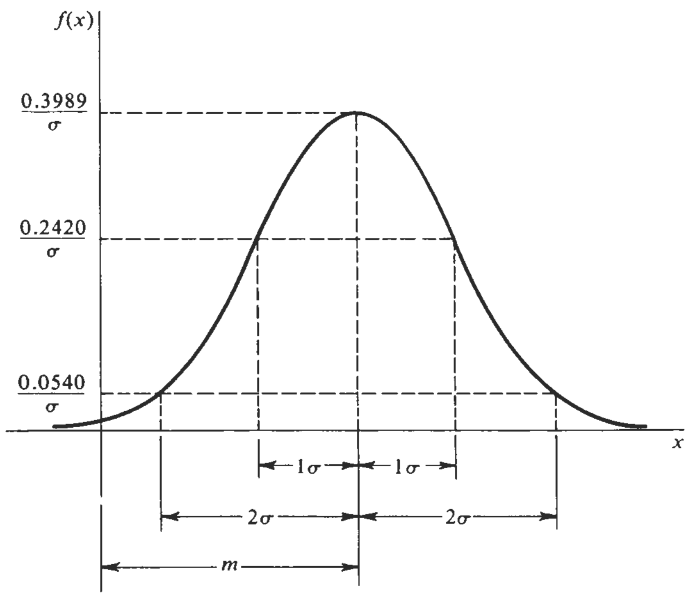



*Figure H-2* *The normal probability density function.*

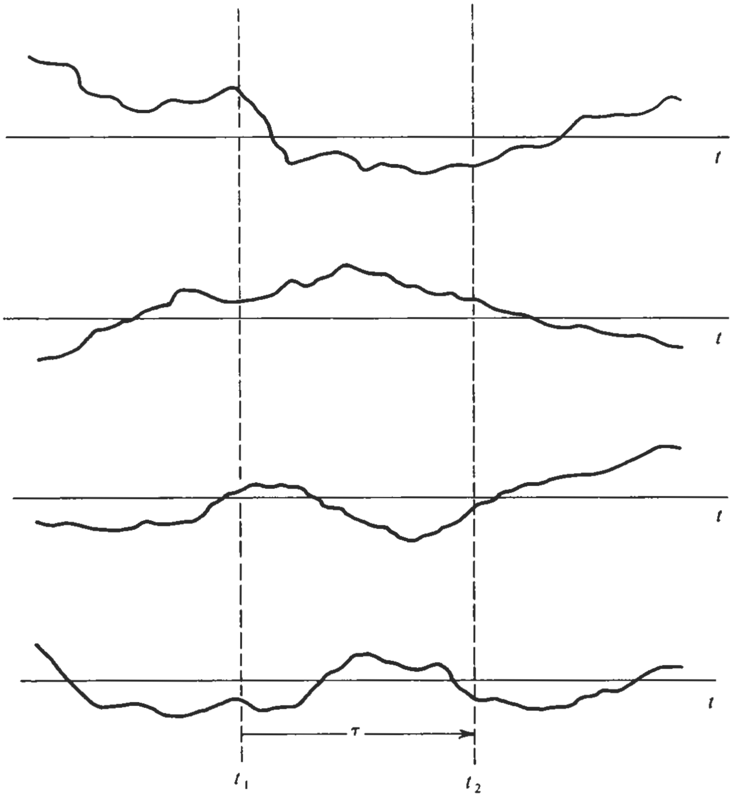
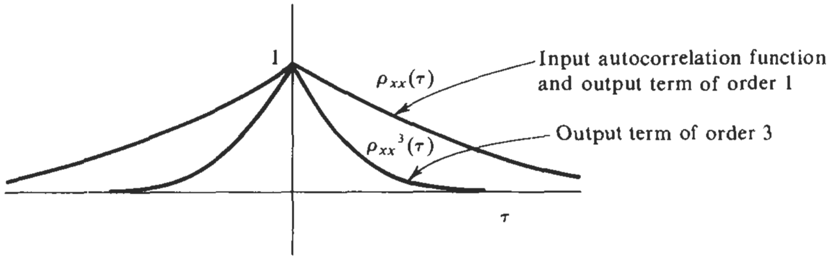
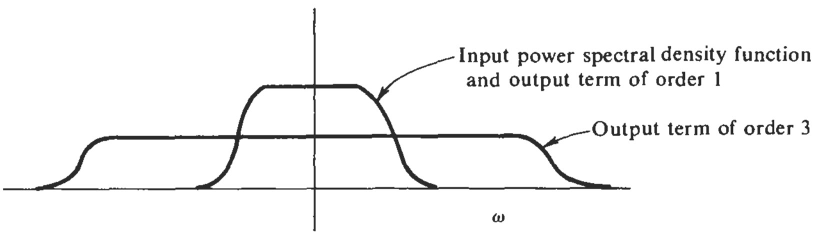



*Figure H-3* *Members of the ensemble* $\{x(t)\}$.




(*a*) Autocorrelation functions




(*b*) Power spectral density functions

***Figure H-4*** *Characteristics of the nonlinearity input and output random processes.*