HST.582J / 6.555J / 16.456J Biomedical Signal and Image Processing
Spring 2007

Chapter 11 - RANDOM SIGNALS: BASIC PROPERTIES

Introduction

In the preceding chapters, we have assumed that signals can be completely specified as either discrete or continuous functions of time. Such complete mathematical specifications are rarely, if ever, available for signals recorded from living systems. Far more commonly, signals are characterized by a set of properties or features. For example, we might know that the EKG consists of more-or-less regular peaks not exceeding 100 mv in amplitude, or that speech has most of its energy below 10 kHz. Because such characterizations are incomplete, there will in general exist a large class, or *ensemble* of signals that share these properties.

When an ensemble of signals characterized by certain properties is processed by a dynamic system such as a linear filter, response signals will form another ensemble characterized by a new set of properties. Our task here and in Chapter 12 will be to determine properties of the ensemble of response signals from knowledge of properties of the input ensemble and system characteristics. Alternatively, we will design systems such that the output signals have certain desired properties, given a set of input properties. While these tasks cannot be solved in general, when the known input properties are long-term *time averages,* and the system has special properties (such being linear or memoryless), it will be possible to specify certain time averages of the response signals.

Signals described in terms of averages are called *random signals, random processes,* or *stochastic processes.* The term "random" is used because the waveforms of such signals are typically irregular and complicated. This term does not necessarily imply that such signals are unpredictable. More often than not, it means that, for a specific purpose, only an appropriate set of averages needs to be known for a class of signals. For example, it might be possible to completely specify a speech signal if we knew the motions of the articulators for all times, and such detailed information would be of great value to speech scientists. On the other hand, for an engineer who is designing a telecommunication system, knowledge of the long-term average spectrum of speech might suffice. To give another example, the electrocardiogram might be predictable from the electrical potentials of every cardiac muscle fibers, but such knowledge would be too cumbersome for deciding whether a patient needs a pacemaker. Thus, it is perfectly appropriate to model the same signal as being random for one particular purpose, and as being deterministic (i.e. completely specified mathematically) for another purpose. In short, the applicability of random signal models is a matter of attitude, not an inherent property of signals.

The study of random signals is part of *probability theory.* Probability is a branch of mathematics, and therefore starts from specific assumptions (axioms) to obtain results (theorems) by deduction. In this chapter and in Chapter 12 we will assume that certain average properties of a class of signals are *known*, apply a *known* transformation to these signals, and deduce appropriate averages for the transformed signals. Specifically, this chapter introduces basic definitions and

properties of random signals, while Chapter 12 treats the processing of random signals by linear systems. In Chapter 13, we will also consider how the infinite-time averages that are assumed to be known in probability theory can be estimated from finite data. The problems of estimating infinite averages from finite data, and of making reliable decisions on the basis of incomplete or noisy data belong to the subject matter of *statistics.*

11.1 Time averages

11.1.1 Definition

The *time-average* or *mean* of a discrete random signal $x[n]$ is defined as:

$$< x[n] > \stackrel{\triangle}{=} \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} x[n] \tag{11.1}$$

This definition is meaningful only if the limiting operation in (11.1) is well defined. Therefore, when we use time-averages, we restrict our attention to the class of signals for which such a limit exists. This class will include for example periodic signals: For such signals, the time average is the DC Fourier coefficient X_0. On the other hand, time averages are not defined for signals that grow monotonically with time (such as exponential signals), or finite-duration signals. In general, the notion of time average is most useful for signals that are *stationary,* i.e. signals whose characteristics do not change over long time intervals. This limitation is not as restrictive as it may appear because, in many instances, time-varying signals can be considered to be approximately stationary for certain time intervals. This is for example the case for speech signals because the motions of the articulators are slow relative to the time constants of the vocal-tract resonances.

The time-average of a continuous-time signal $x(t)$ is

$$< x(t) > \stackrel{\triangle}{=} \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t)\ dt \tag{11.2}$$

Because properties of time averages for continuous-time signals are similar to those for discrete-time signals, we will only state results for the discrete-time case.

11.1.2 Properties of time averages

Two basic properties of time averages are used to derive time averages for combinations and transformations of signals:

1. *Stationarity:* Time averages do not change if signals are delayed by any number n_0 of samples:
$$< x[n-n_0] > \ = \ < x[n] > \ \stackrel{\triangle}{=} \mu_x \tag{11.3}$$

The short-hand notation μ_x makes stationarity of the mean apparent. Stationarity is easy to prove for the practically-important case of bounded signals:

$$< x[n - n_0] > \; = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=n_0-N}^{n_0+N} x[n] \tag{11.4}$$

Assuming that $n_0 > \;\; 0$, the sum in (11.4) can be decomposed into three terms:

$$\sum_{n=n_0-N}^{n_0+N} x[n] = \sum_{n=-N}^{N} x[n] + \sum_{n=N+1}^{n_0+N} x[n] - \sum_{n=-N}^{n_0-N-1} x[n]$$

Dividing each term by $2N + 1$ and taking the limit, the first term on the right tends to $x[n]$, while the last two terms tend to zero because the sum over a finite number n_0 of bounded terms is bounded.

2. *Linearity:* The time average of a weighted sum of signals is the weighted sum of the averages:
$$< ax[n] + by[n] > \; = a < x[n] > \; + b < y[n] > \tag{11.5}$$
for a and b arbitrary constants.

11.1.3 Averages of functions of a signal

The notion of time-average can be generalized to functions of a signal (Fig. 11.1A):

$$< g(x[n]) > \; \stackrel{\triangle}{=} \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} g(x[n]) \tag{11.6}$$

For example, the *mean power* of a signal is the time average of its square (Fig. 11.1B):

$$P_x \stackrel{\triangle}{=} \; < x[n]^2 > \; = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} x[n]^2 \tag{11.7}$$

The mean AC power is the average of the square of the signal once the DC component (i.e. the mean) has been subtracted out (Fig. 11.1C):

$$\sigma_x^2 \stackrel{\triangle}{=} \; < (x[n] - \mu_x)^2 > \tag{11.8}$$

The average AC power is also called the *variance,* and its square root σ_x is the *standard deviation.* The total power is the sum of the AC power and the DC power:

$$P_x \; = \; \sigma_x^2 \; + \; \mu_x^2 \tag{11.9}$$

This can be verified by applying the linearity property to the definition of the AC power:

$$\sigma_x^2 = < (x[n]-\mu_x)^2 > = < x[n]^2 - 2\mu_x x[n] + \mu_x^2 > = < x[n]^2 > \; - 2\mu_x < x[n] > \; + \; < \mu_x^2 > = P_x - \mu_x^2$$

In this derivation, we have assumed that $< \mu_x > \; = \mu_x$ because the mean of a constant signal is that constant.

11.1.4 Time averages and probabilities

Time averages are closely related to the notion of probability. Intuitively, the *probability* of an event is the limit of the frequency of occurrence of that event when the number of trials becomes large. To see the relation between probabilities and time averages, consider the average of $u(x[n])$, where $u(x)$ is the unit step function (Fig. 11.2):

$$< u(x[n]) > \; = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} u(x[n])$$

The function $u(x[n])$ is always equal to either 1 when $x[n] \;\geq\; 0$ or 0 otherwise. Therefore the sum in the above equation is the number of samples over the interval $[-N,\; N]$ for which $x[n] \geq 0$:

$$u(x[n]) = \lim_{N\to\infty} \frac{\text{number of samples in } [-N, N] \text{ for which } x[n] \;\geq\; 0}{\text{total number of samples in } [-N,\; N]}$$

This formula means that $< u(x[n]) >$ is the limit for large N of the frequency of occurrence of non-negative samples, i.e. the probability that $x[n]$ is non-negative, $P\,(x[n] \geq 0)$. More generally, given an arbitrary "threshold" X_0, the probability that $x[n]$ is greater than or equal to X_0 is given by the time average:

$$P\,(x[n] \;\geq\; X_0) \;=\; < u(x[n] - X_0) > \tag{11.10}$$

This computation is shown in Fig. 11.1D. The probability that $x[n]$ is between the two thresholds X_1 and X_2 is:

$$P\,(X_1 \;\leq\; x[n] \;\leq\; X_2) \;=\; < \Pi_{X_1,X_2}(x[n]) >$$

where

$$\Pi_{X_1,X_2}(X) \triangleq \begin{cases} 1 & \text{if } X_1 \leq X \leq X_2 \\ 0 & \text{otherwise} \end{cases}$$

Thus, for our purposes, probabilities are special time averages.

11.1.5 Chebyshev's inequality

Variances and probabilities are related by an important inequality due to Chebyshev. This inequality provides an upper bound on the probability that a signal $x[n]$ deviates from its mean by more than an arbitrary amount ϵ:

$$P\,(|x[n] \;-\; \mu_x| \;\geq\; \epsilon) \;\leq\; \frac{\sigma_x^2}{\epsilon^2} \tag{11.11}$$

To prove this inequality, we consider the signal:

$$y[n] \;\triangleq (x[n] \;-\; \mu_x)^2$$

From the definition of the variance (11.8), we see that $y[n] = \; < \sigma_x^2 >$. For any threshold ϵ^2, $y[n]$ is always greater than the rectangular signal that is equal to the threshold ϵ^2 when $y[n]$ is above threshold, and to zero otherwise (Fig. 11.3):

$$\epsilon^2 \; u(y[n] \;-\; \epsilon^2) \;\leq\; y[n]$$

Replacing $y[n]$ by its definition in terms of $x[n]$, and taking time averages, we obtain

$$\epsilon^2 \; P\left((x[n] \; - \; \mu_x)^2 \; \geq \; \epsilon^2\right) \; \leq \; \sigma_x^2$$

Dividing by ϵ^2, and noting that $P((x[n] \; - \; \mu_x)^2 \; \leq \; \epsilon^2) \; = \; P(|x[n] \; - \; \mu_x| \; \leq \; \epsilon)$, we obtain (11.11), completing the proof of Chebyshev's inequality.

It is remarkable that Chebyshev's inequality holds for arbitrary random signals. However, for specific classes of signals, tighter bounds can often be obtained. For example, according to Chebyshev's inequality, the probability that a signal deviates from its mean by more than 2 standard deviations is less than 25%. For Gaussian signals (see Sec. 11.4.3), this probability is actually less than 5%. Rather than being a practical bound, Chebyshev's inequality is theoretically important, because it justifies the common mathematical practice of *least squares estimation.* In many applications, a desired signal $y[n]$ must be approximated by a signal $\hat{y}[n]$. Least-squares estimation consists in minimizing the variance of the error signal $e[n] \; \stackrel{\triangle}{=} \; y[n] \; - \; \hat{y}[n]$. Minimizing the variance makes sense because, according to Chebyshev's inequality, the error signal is unlikely to be large if its variance is small. Examples of least-squares estimation are presented in Chapters 12 and 13.

11.2 Autocorrelation functions

Averages of functions of random signals that were considered in the previous section (such as means, variances, and simple probabilities) depend only on the present value of the input signal, and are therefore called *memoryless.* For digital filters, on the other hand, outputs generally depend on past values of the input as well as on its present value (the only exception is a pure gain). Thus, in order to process random signals by digital filters, we need to define time averages that characterize how rapidly signals vary with time. Specifically, if a signal reaches a certain value X at time n, it is likely to remain in the vicinity of X for times shortly following n. In other words, signal samples separated by short intervals are not, in general, independent from one another. Therefore we need to know time averages characterizing these *dependencies (or correlations)* between samples taken *at different times.* The most important of these time averages is the autocorrelation function.

11.2.1 Filtering a random signal

To demonstrate how autocorrelation functions arise, consider the simple, first-order FIR filter whose output $y[n]$ is defined as a function of the input $x[n]$ by the difference equation:

$$y[n] \; = \; \frac{1}{2} \; (x[n] \; + \; x[n-1]) \tag{11.12}$$

Knowing the mean of the input signal suffices to compute the mean of the output signal:

$$\mu_y = \; < \frac{1}{2}(x[n] \; + \; x[n-1]) > \; = \frac{1}{2}(< x[n] > \; + \; \; < x[n-1] >) = \mu_x$$

However, the output power cannot be computed from the input power alone:

$$P_y \triangleq \; < \frac{1}{4}(x[n] \; + \; x[n-1])^2 > \; = \frac{1}{4}\left[< x[n]^2 > \; + \; < x[n-1]^2 > \; + 2 < x[n] \; x[n-1] >\right]$$

Using the time-invariance property of averages, this becomes:

$$P_y = \frac{1}{2}(P_x + \quad < x[n] \; x[n-1] >) \tag{11.13}$$

Thus, we must know not only the input power P_x, but also the time average $\quad < x[n]x[n-1] >$ in order to compute P_y. More generally, if the input signal is to be processed by an N^{th}-order FIR filter:

$$y[n] \; = \; \sum_{k=0}^{N} \; b_k \; x[n-k]$$

then time averages of the form $\quad < x[n]x[n-k] >$ must be known for all $0 \leq k \leq N$ in order to compute the output power.

11.2.2 Definition of autocorrelation and aurocovariance functions

The above arguments lead us to define the *autocorrelation function* of a discrete random signal $x[n]$ as the time average

$$R_x[k] \; \triangleq \; < x[n] \; x[n-k] > \tag{11.14}$$

By the time-invariance property of time averages, the autocorrelation function does not depend on absolute time n, only on the time separation, or *lag* k. Intuitively, the autocorrelation function measures the resemblance between successive samples of a signal. If the signal varies slowly, samples at time n and time $n-1$ will be nearly always of the same sign, and the mean of their product $R_x[1]$ will be large. Conversely, if a zero-mean signal varies rapidly, even samples separated by short intervals will be equally likely to be of the same sign or of opposite signs, so that the autocorrelation function will approach zero. Fig. 11.1E shows a signal flow diagram for computing the autocorrelation function for one particular lag k.

A time average closely related to the autocorrelation function is the *autocovariance function* $C_x[k]$, which is the autocorrelation function of the signal minus its mean:

$$C_x[k] \triangleq \; < (x[n] - \mu_x)(x[n-k] - \mu_x) > \; = R_x[k] - \mu_x^2 \tag{11.15}$$

Being an autocorrelation function, the autocovariance function has the same properties as autocorrelation functions.

11.2.3 Properties of autocorrelation and autocovariance functions

1. Autocorrelation functions are even functions of lag:

$$R_x[-k] = \; < x[n]x[n+k] > \; = \; < x[n-k]x[n] > \; = R_x[k] \tag{11.16}$$

2. The autocorrelation function evaluated at the origin is the mean power, and the autocovariance function is the variance:

$$R_x[0] = \ < x[n]^2 > \ = P_x \tag{11.17a}$$

$$C_x[0] = \ < (x[n] - \mu_x)^2 > \ = \sigma_x^2 \tag{11.17b}$$

3. Autocorrelation functions are always maximum at the origin:

$$|R_x[k]| \ \leq \ R_x[0] \qquad\qquad -\infty \ < k \ < \infty \tag{11.18}$$

This property is a special case of the *Cauchy-Schwarz inequality,* which states that, for arbitrary random signals $x[n]$ and $y[n]$, one has:

$$< x[n] \ y[n] >^2 \ \leq \ < x[n]^2 > < y[n]^2 > \tag{11.19}$$

To show this, consider the following time average:

$$< (ax[n] + y[n])^2 > \ = \ a^2 < x[n]^2 > \ + 2a < x[n] \ y[n] > \ + \ < y[n]^2 > \tag{11.20}$$

where a is a parameter. This quadratic function of a must always be positive because it is the time average of a positive function. Therefore, its discriminant D must be nonpositive:

$$D \stackrel{\triangle}{=} \ < x[n] \ y[n] >^2 - \ < x[n]^2 > \ < y[n]^2 > \ \leq \ \ 0 \tag{11.21}$$

Rearranging terms yields (11.19). Eq. (11.18) is obtained by letting $y[n] \ = \ x[n-k]$ in (11.19).

4. For large lags, the autocovariance function of a signal having no periodic component approaches zero:

$$\lim_{|k| \to \infty} C_x[k] = 0 \tag{11.22a}$$

In other words, samples of a random signal separated by large lags are *uncorrelated.* This fits with the intuitive notion that random signals must have irregular, unpredictable waveforms. From (11.18) and (11.15), it follows that the limit of the autocorrelation function for large lags is the square of the mean:

$$\lim_{|k| \to \infty} R_x[k] = \mu_x^2 \tag{11.22b}$$

5. A random signal $w[n]$ is said to be *white* if its autocovariance function is an impulse at lag 0:

$$C_w[k] = \sigma_w^2 \delta[k] = \begin{cases} \sigma_w^2 & \text{if } k = 0 \\ 0 & \text{otherwise} \end{cases} \tag{11.23a}$$

From (11.15), if a white signal is in addition zero-mean, its autocorrelation function is an impulse at the origin. In general, the autocorrelation function of a white signal is the sum of an impulse and a constant:

$$R_w[k] \ = \ \sigma_w^2 \ \delta[k] \ + \ \mu_w^2 \tag{11.23b}$$

The term "white" comes from the fact that such signals have energy at all frequencies, just as white light contains all visible wavelengths. White noise plays a fundamental role in random-signal theory because random signals with arbitrary frequency spectra can be generated by passing white noise through appropriate linear filters. Many physical and biological signals can be modeled as white noise. Examples include the thermal noise across a resistor, or the acoustic signal produced by uttering the sound "FF". The signal produced by an ideal random noise generator is white noise.

11.2.4 Autocorrelation examples

1. The autocorrelation function of a sine wave is a cosine at the same frequency. Specifically, let $s[n]$ be a sinusoidal signal with frequency f:

$$s[n] = A\sin(2\pi fn + \phi) \tag{11.24}$$

Its autocorrelation function is

$$R_s[k] = \ < A^2 \ \sin\ (2\pi fn \ + \ \phi) \ \sin\ (2\pi f[n+k] \ + \ \phi) >$$

Making use of a trigonometric identity and the linearity of time averages, this becomes

$$R_s[k] = \frac{A^2}{2}\left(< \cos\ 2\pi fk > \ - \ < \cos(2\pi f[2n+k] + 2\phi) >\right)$$

The first term depends only on lag k, not on time n, and its mean is therefore constant. The second term is sinusoidal function of n, which has zero mean. Therefore,

$$R_s[k] = \frac{A^2}{2} \ \cos\ 2\pi fk \tag{11.25}$$

Note that the autocorrelation function does not depend on phase ϕ, and that $R_x[0]$ is the mean power $A^2/2$, as expected. The autocorrelation function of a sine wave is shown in Fig. 11.4A.

2. The autocorrrelation function of a periodic signal with period N is also periodic with period N:

$$\text{if } x[n+N] = x[n], \ \text{ then } R_x[k+N] = \ < x[n]x[n+k+N] > \ = \ < x[n]x[n+k] > \ = R_x[k] \tag{11.26}$$

Fig. 11.4B show the waveforms and autocorrelation functions of two periodic signals, a sine wave and a 3-component signal. In Sec. 11.3.4, we apply this property to the detection of periodic signals in noise.

3. To illustrate the use of autocorrelation functions, we continue the example of a random signal processed by the simple FIR filter (11.12). From (11.13) and (11.14), the mean power of the output signal is

$$P_y \ = \ \frac{1}{2}\left(R_x[0] \ + \ R_x[1]\right)$$

Two extreme cases can be considered:

1. If the input is white and zero-mean, then $R_x[1] \ = \ 0$, so that the output power is half the input power. This shows that a smoothing filter such as (11.12) is useful in reducing interference by zero-mean, white noise.

2. If consecutive samples of the input are strongly correlated (i.e., if $R_x[1] \ \approx \ R_x[0]$), then the output power is nearly equal to the input power, and the filter will not be effective in reducing noise.

Note however, that, in the case of white noise, passing the signal through the filter a second time results in a smaller reduction in noise power than the factor of 2 achieved in the first pass. This is because the filter introduces correlation between successive samples, that is, $R_y[1] \ \neq 0$, as shown in Fig. 11.4C.

4. For a slightly more elaborate example, consider the signal $y[n]$ obtained by passing zero-mean, white noise with variance σ_w^2 through a first-order, recursive digital filter:

$$y[n] \;=\; a\; y[n-1] \;+\; w[n] \tag{11.27}$$

The autocorrelation function $R_y[k]$ can be evaluated by inspection in this simple case. More general techniques will be given in Chapter 12. The basic idea is to first evaluate $R_y[0]$, then obtain $R_y[k]$ by induction on k. To evaluate $R_y[0]$, we start from the difference equation (11.27):

$$R_y[0] = \; < y^2[n] > \; = \; < (ay[n-1] + w[n])^2 >$$

Making use of linearity, this becomes

$$R_y[0] = a^2 \quad < y^2[n-1] > \quad + \quad < w^2[n] > \; + 2a < w[n]y[n-1] >$$

Further making use of stationarity, we get

$$R_y[0] = a^2 R_y[0] + \sigma_w^2 + 2a < w[n]y[n-1] >$$

The last term in the sum is zero because $y[n-1]$ depends only on the past values, $w[n-1], w[n-2], w[n-3], \ldots$, but not on the present value $w[n]$. Since, $w[n]w[n-i] = 0$ for $i \neq 0$ (zero-mean white noise), $w[n]y[n-1]$ is also zero. Therefore, rearranging terms, we obtain:

$$R_y[0] = \frac{\sigma_w^2}{1-a^2} \tag{11.28}$$

$R_y[k]$ can now be derived from $R_y[0]$ by induction on k. Again, we start from the difference equation, which is multiplied by $y[n-k]$, then time-averaged:

$$R_y[k] = \; < y[n]y[n-k] > \; = \; < ay[n-1]y[n-k] + w[n]y[n-k] >$$

Making use of linearity and stationarity, we obtain

$$R_y[k] = a < y[n-1]y[n-k] > \quad + \quad < w[n]y[n-k] > = aR_y[k-1] + \; < w[n] \; y[n-k] >$$

If $k > 0$, the quantity $< w[n]y[n-k] >$ is zero for the same reason as $< w[n]y[n-1] >$ above. Therefore, $R_y[k] = aR_y[k-1]$ for $k > 0$ and, by induction on k,

$$R_y[k] = a^{|k|} R_y[0] = a^{|k|} \; \frac{\sigma_w^2}{1-a^2} \tag{11.29}$$

The autocorrelation function $R_y[k]$ is shown in Fig. 11.4D for several values of $a > 0$. The larger a, the more correlated (the less white) $y[n]$ becomes.

11.3 Crosscorrelation functions

11.3.1 Filtering a sum of random signals

In many applications, one is not dealing with a single random signal, but with combinations of signals. For example, the signal $x[n]$ recorded from the chest of a human subject can be modeled

as the sum of the electrocardiogram $s[n]$ and a noise (disturbance) $d[n]$ reflecting the activity of other muscles as well as noise generated by the amplifier:

$$x[n] = s[n] + d[n] \tag{11.30}$$

We process this signal through a digital filter in the hope of improving the signal-to-noise ratio. In order to express the signal-to-noise ratio in the filter output, we will need to know the autocorrelation function of $x[n]$:

$$R_x[k] \;=\; < x[n]x[n+k] > \;=\; < (s[n] \;+\; d[n])\;(s[n+k] \;+\; d[n+k]) > \tag{11.31}$$

Making use of (11.14), this becomes:

$$R_x[k] = R_s[k] + R_d[k] + \;< s[n]\; d[n+k] > \quad + \quad < d[n]\; s[n+k] > \tag{11.32}$$

This relation can be concisely written by defining the *crosscorrelation function* of two random signals $x[n]$ and $y[n]$:

$$R_{xy}[k] \stackrel{\triangle}{=} \;< x[n]\; y[n+k] > \;=\; < x[n-k]y[n] > \tag{11.33}$$

Just as the autocorrelation function is a measure of the dependence between successive samples of a single signal, the crosscorrelation function $R_{xy}[k]$ measures the resemblance between the signal $x[n]$ and a delayed version of $y[n]$. Often, a large correlation between two signals indicates that they have a common source. For example, the input and the output of a filter, or the outputs of two filters that have the same input are in general correlated. Making use of the definition of the crosscorrelation function, Equation (11.32) becomes:

$$R_x[k] = R_s[k] + R_d[k] + R_{sd}[k] + R_{ds}[k] \tag{11.34}$$

A close relative of the crosscorrelation function is the *crosscovariance function,* which is the crosscorrelation function of the centered signals $x[n] - \mu_x$ and $y[n] - \mu_y$:

$$C_{xy}[k] \stackrel{\triangle}{=} \;< (x[n] \;-\; \mu_x)(y[n+k] \;-\; \mu_y) > \;= R_{xy}[k] - \mu_x\mu_y \tag{11.35}$$

11.3.2 Properties of crosscorrelation and crosscovariance functions

1. Unlike autocorrelation functions, crosscorrelation functions are not even. In fact,

$$R_{xy}[-k] = \;< x[n]y[n-k] > \;=\; < x[n+k]y[n] > \;= R_{yx}[k] \tag{11.36}$$

2. From Schwarz' inequality (11.19), it is clear that crosscorrelation functions obey the following inequality:

$$|R_{xy}[k]|^2 \leq R_x[0]R_y[0] \tag{11.37}$$

The *correlation coefficient* $\rho_{xy}[k]$ is the crosscovariance function normalized by the product of the standard deviations:

$$\rho_{xy}[k] \stackrel{\triangle}{=} \frac{C_{xy}[k]}{\sigma_x\sigma_y} \tag{11.38}$$

From (11.37), the correlation coefficient is bounded between -1 and 1.

3. Two signals are said to be *uncorrelated* if their crosscovariance function is zero for all k. Often, this indicates that the signals are generated by independent processes. For example, the electrocardiogram is uncorrelated with 60-Hz noise, or with the EMG potentials produced by other muscles. On the other hand, the heart rate signal is correlated with breathing rate because, through the autonomic nervous system, breathing influences heart rate. From (11.35), if two signals are uncorrelated, and at least one of them is zero-mean, then their crosscorrelation function is zero for all lags.

11.3.3 Crosscorrelation examples

1. If $x[n]$ and $y[n]$ are two sine waves with different frequencies, they are uncorrelated. On the other hand, the crosscorrelation of two sine waves having the same frequency is a sine wave at the same frequency. Specifically, let

$$x[n] \;=\; A_x \;\sin\;(2\pi f_x n \;+\; \phi_x) \tag{11.39a}$$

and

$$y[n] \;=\; A_y \;\sin\;(2\pi f_y n \;+\; \phi_y) \tag{11.39b}$$

The crosscorrelation function is

$$R_{xy}[k] = \;< x[n]\; y[n+k] >\; = A_x A_y \quad < \sin(2\pi f_x n + \phi_x) \sin(2\pi f_y [n+k] + \phi_y) >$$

Applying a trigonometric identity, this expression can be decomposed into a difference of two sine waves with frequencies $f_x \;-\; f_y$ and $f_x \;+\; f_y$, respectively. If $f_x \;\neq\; f_y$, both of these terms are sinusoidal functions of n, so that their means are both zero, and $R_{xy}[k]$ is also zero. If, on the other hand, $f_x \;=\; f_y \;=\; f$, the first term is a function of k, not n, so that

$$R_{xy}[k] \;=\; \frac{A_x\; A_y}{2}\;\cos\;(2\pi f k \;+\; \phi_y \;-\; \phi_x) \tag{11.40}$$

Note that, unlike the autocorrelation function, the crosscorrelation function depends on the relative phase $\phi_y \;-\; \phi_x$. We show below that this result has applications to the detection of sinusoidal signals in noise.

2. Let $x[n]$ be a periodic signal expressed as a discrete Fourier series:

$$x[n] \;=\; \sum_{i=0}^{\infty}\; A_i \;\sin(2\pi i n/N \;+\; \phi_i) \tag{11.41}$$

From the previous example, all the frequency components are uncorrelated with each other. Therefore, the autocorrelation function $R_x[k]$ is the sum of the autocorrelation functions of all the sinusoidal components. From (11.25), the autocorrelation of a sine wave is a cosine. Thererfore

$$R_x[k] \;=\; \sum_{i=0}^{\infty}\; \frac{A_i^2}{2}\;\cos\;2\pi i k/N \tag{11.42}$$

Specializing this relation for $k \;=\; 0$ gives Parseval's theorem for discrete Fourier series.

3. If $y[n]$ is obtained by delaying $x[n]$ by n_0 samples (i.e. $y[n] = x[n-n_0]$), then the crosscorrelation function of the two signals is maximum for $k = n_0$:

$$R_{xy}[k] = < x[n]y[n+k] > = < x[n]x[n+k-n_0] > = R_x[k-n_0] \tag{11.43}$$

From (11.18), $R_x[k-n_0]$ is maximum for $k = n_0$. This result has applications in radar and sonar engineering, where the delay between the emission and the echo is the measure of the distance to the target. Bats excel at determining the distance of a target (e.g. a fly) by crosscorrelating their own vocalization with the received echo. In order to determine the location of sound sources in space, the brain performs a kind of crosscorrelation operation on the neural outputs of the two ears, thereby estimating the interaural time difference.

4. We now return to the problem of filtering the electrocardiogram signal $x[n] = s[n] + d[n]$. If the signal $s[n]$ and the noise $d[n]$ are uncorrelated, then (11.34) becomes:

$$R_x[k] = R_s[k] + R_d[k] \tag{11.44}$$

Suppose now that this signal is processed by the first-order FIR filter defined by (11.12). The mean power in the output signal is:

$$P_y = \frac{1}{2}\left(R_x[0] + R_x[1]\right) = \frac{1}{2}\left(R_s[0] + R_s[1]\right) + \frac{1}{2}\left(R_d[0] + R_d[1]\right) \tag{11.45}$$

This is the sum of a term due to the signal and a term due to the noise. If we further assume that $s[n]$ varies slowly, then $R_s[1] \approx R_s[0]$. On the other hand, in many cases the noise will be almost white, so that $R_d[1] \approx 0$. If these two assumptions are verified, the output power becomes:

$$P_y \approx P_s + \frac{1}{2} P_d \tag{11.46}$$

Thus, before filtering, the signal-to-noise ratio is $\frac{P_s}{P_d}$, while after filtering, it becomes 2 $\frac{P_s}{P_d}$. In other words, for this particular (and extreme) set of assumptions, filtering results in a 3-dB improvement in signal-to-noise ratio. In contrast, if the signal and the noise were strongly correlated, it would be difficult to improve the signal-to-noise ratio by filtering because any filter would have similar effects on the signal and the noise. Chapter 12 introduces general techniques for designing filters that do the best job of separating signal from noise.

5. Consider two uncorrelated signals $u[n]$ and $v[n]$. We form two new signals $x[n]$ and $y[n]$ from the sum and difference of $u[n]$ and $v[n]$, respectively:

$$x[n] = u[n] + v[n], \text{ and } y[n] = u[n] - v[n]$$

Then, the crosscorrelation function of the two new signals is:

$$R_{xy}[k] = < x[n]y[n+k] > = < (u[n]+v[n])(u[n+k]-v[n+k]) >$$

Making use of linearity, this becomes

$$R_{xy}[k] = < u[n]u[n+k] > + < v[n]u[n+k] > - < u[n]v[n+k] > - < v[n]\, v[n+k] >$$

$$R_{xy}[k] = R_u[k] + R_{vu}[k] - R_{uv}[k] - R_v[k]$$

Since $R_{uv}[k] = 0$, this simplifies to:

$$R_{xy}[k] = R_u[k] - R_v[k]$$

Thus, $x[n]$ and $y[n]$ are correlated because they both depend on $u[n]$ and $v[n]$.

11.3.4 Detection of a sine wave in noise

Let $s[n] = A \cos(2\pi f n + \phi)$ be a sinusoidal signal buried in additive, zero-mean noise $d[n]$. Further assume that the signal and the noise are uncorrelated. We will show that the signal can be recovered exactly by forming the crosscorrelation function between the noisy signal and the unit cosine wave $c[n] \stackrel{\triangle}{=} \cos 2\pi f n$ with the same frequency as $s[n]$. The crosscorrelation function is:

$$R_{cx}[k] = < c[n](s[n+k] + d[n+k]) > = R_{cs}[k] + R_{cd}[k] \tag{11.47}$$

The second term is zero because $c[n]$ and $d[n]$ are uncorrelated. (Otherwise, $s[n]$ and $d[n]$ would also be correlated, which is against our assumption.) Therefore

$$R_{cx}[k] = R_{cs}[k] = A/2 \cos(2\pi f k + \phi) \tag{11.48}$$

which is equal to $s[k]$ within a factor of 2. Although, in principle, the signal can be recovered exactly, in practice, it is only possible to average over a finite time, so that some noise will always remain even after crosscorrelating. Figure 11.5A shows the waveform of a sinusoidal signal buried in noise, while Fig. 11.5B shows an estimate of the crosscorrelation $R_{cx}[k]$ based on a large number of signal samples. Clearly, the sine wave is much more visible in the crosscorrelation function than in the original signal.

A major difficulty with this method is that, in many cases, the period of the signal is not known in advance. In such cases, a second method based on the autocorrelation function can be applied for detecting periodicities in a noisy signal. Specifically, suppose that the signal $x[n]$ is the sum of a periodic component $s[n]$ and a zero-mean disturbance $d[n]$ which is uncorrelated with $s[n]$. The autocorrelation function is:

$$R_x[k] = R_s[k] + R_d[k] + R_{sd}[k] + R_{ds}[k] \tag{11.50}$$

The last two terms are zero because the signal and the noise are uncorrelated. The autocorrelation function of the disturbance signal $R_d[k]$ tends to fall off with increasing lag so that, if $|k|$ is greater than a certain value k_0, $R_d[k] \approx 0$. In that range of lags, $R_x[k] \approx R_s[k]$, which is a periodic function of lag. Therefore, it will be easier to detect periodicities from the autocorrelation function $R_x[k]$ than from the original signal $x[n]$. Once periods are identified, the first method based on crosscorrelation can be used to recover each periodic component of the signal. Figure 11.5D shows an estimate of the autocorrelation function of the noisy sinusoidal signal shown on top. Note that information about the phase of the sine wave is lost. This is not true for the crosscorrelation method of Fig. 11.5B.

11.4 Probability density functions

11.4.1 Definition

While autocorrelation functions play a key role in describing the responses of linear filters to random signals, these functions are less useful for nonlinear systems. For example, knowing the autocorrelation function of the input $x[n]$ does not, in general, allow us to compute the power at the output of the simple nonlinear system $y[n] = x[n]^2$. In the special case of *memoryless*

systems, however, there is a family of time averages which plays a role similar to that of the autocorrelation function for linear systems, in that it can be used to compute the means and mean powers of the output signals: This is the (first-order) *probability density function (p.d.f).* We have seen in Sec. 11.1 that the probability that the signal $x[n]$ is between X and $X + \Delta X$ is the time average of a rectangular function of width ΔX:

$$P(X \leq x[n] < X + \Delta X) = < \Pi_{\Delta X}(x[n] - X) > \tag{11.51a}$$

where

$$\Pi_{\Delta X}(x) \triangleq \begin{cases} 1 & \text{if } 0 \leq x < \Delta X \\ 0 & \text{otherwise} \end{cases} \tag{11.51b}$$

The probability density function $f_x(X)$ is obtained by dividing this probability by the interval width ΔX, then taking the limit when ΔX approaches 0:

$$f_x(X) \triangleq \lim_{\Delta X \to 0} \frac{1}{\Delta X} P(X \leq x[n] < X + \Delta X) = \lim_{\Delta X \to 0} < \frac{1}{\Delta X} \Pi_{\Delta X}(x[n] - X) > \tag{11.52}$$

The rectangular function $\frac{1}{\Delta X}\Pi_{\Delta X}(x)$ that is being averaged has area 1 for all values of ΔX. Therefore, in the limit when ΔX tends to zero, this function tends to a unit impulse $\delta(X)$. This means that the probability density function can formally be written as the time-average of a δ function:

$$f_x(X) = < \delta(x[n] - X) > \tag{11.53}$$

The probability density function $f_x(X)$ is the probability that the signal amplitude $x[n]$ lies in a very narrow interval centered at $x[n] = X$, divided by the width of this interval. The lower-case x in $f_x(X)$ refers to the random signal $x[n]$ that is being averaged, while the upper case X refers to the center of the interval for which the probability is computed. Thus, there is no relation between the two X's, and, for example, $f_y(X)\ \Delta X$ would refer to the probability that the signal $y[n]$ lies in an interval of width ΔX centered at X. Figure 11.6 illustrates in a concrete way how the p.d.f. can be considered as the limit of a time average.

11.4.2 Properties of probability density functions

1. Because the p.d.f. is the time-average of a non-negative function of $x[n]$, it must also be non-negative:

$$f_x(X) \geq 0 \quad \text{for all } X \tag{11.54}$$

2. From the definition of the p.d.f. (11.52), it is clear that the probability that $x[n]$ lies in the interval $[X_1,\ X_2]$ is obtained by integrating the p.d.f. over this interval:

$$P(X_1 \leq x[n] \leq X_2) = \int_{X_1}^{X_2} f_x(X)\ dX \tag{11.55}$$

3. Specializing (11.55) to the case when either edge of the interval is infinite gives:

$$P(x[n] \leq X) = \int_{-\infty}^{X} f_x(V)\ dV \tag{11.56a}$$

$$P(x[n] \geq X) = \int_{X}^{\infty} f_x(V)\ dV \tag{11.56b}$$

4. Further specializing (11.56b) to the interval $[-\infty,\ \infty]$ gives

$$\int_{-\infty}^{\infty} f_x(X)\ dX\ =\ 1 \tag{11.57}$$

because the probability that $x[n]$ lies between $-\infty$ and ∞ is always 1. This relation justifies the term probability *density* function.

11.4.3 Examples of probability density functions

1. The *uniform* p.d.f. over the interval $[a, b]$ (Fig. 11.7a):

$$f_x(X)\ \stackrel{\triangle}{=} \begin{cases} \frac{1}{b\ -\ a} & \text{if } a \le X \le b \\ 0 & \text{otherwise} \end{cases} \tag{11.58}$$

Most random number generators produce signals with a uniform p.d.f. between 0 and 1.

2. The *binomial* p.d.f. is a weighted sum of two impulses at $X\ =\ 0$ and $X\ =\ 1$:

$$f_x(X)\ \stackrel{\triangle}{=} (1\ -\ p)\ \delta(X)\ +\ p\ \delta(X\ -\ 1) \tag{11.59}$$

Binomial random signals can be generated by repeated flipping a coin, and, on the nth drawing, setting $x[n]$ to either 0 or 1 when "tail" or "head" is drawn, respectively. Then, the parameter p is the probability of "head", and $(1\ -\ p)$ the probability of "tail". It can be verified that the mean of a binomial signal is p and the variance $p\ (1\ -\ p)$. Binomial random signals can be used to model the discharge patterns of auditory neurons in response to acoustic stimuli. In this case, p represents the probability that a spike discharge occurs over a short time interval.

3. The *exponential* p.d.f. (Fig. 11.7b):

$$f_x(X)\ \stackrel{\triangle}{=} \lambda\ e^{-\lambda X}\ u(X) \tag{11.60}$$

Both the mean and the variance of an exponentially-distributed random signal are equal to the parameter λ. The exponential p.d.f. is an approximate characterization of the interspike interval distribution for auditory-nerve fibers.

4. By far the most important of all probability density functions is the *Gaussian* p.d.f (Fig. 11.7c):

$$f_x(X)\ =\ \frac{1}{\sqrt{2\pi}\sigma}\ e^{-\ \frac{(X\ -\ \mu)^2}{2\ \sigma^2}} \tag{11.61}$$

It can be verified that the parameters μ and σ are respectively the mean and standard deviation of the signal. Gaussian random signals are often good models for physical or biological signals because, according to the *central limit theorem* if many different signals with arbitrary p.d.f.'s are added together, then the p.d.f. of their sum approaches a Gaussian p.d.f. This situation is applicable, for example, to the evoked potential, which is a sum of potentials produced by many millions of neurons.

5. The *chi-squared* p.d.f. (Fig. 11.7d):

$$f_y(Y)\ \stackrel{\triangle}{=}\ \frac{1}{\sqrt{2\pi Y}\sigma}\ e^{-Y/2\sigma}\ u(Y) \tag{11.62}$$

If $x[n]$ is a zero-mean Gaussian signal with variance σ^2, then $y[n]\ =\ x[n]^2$ has a chi-squared p.d.f. with parameter σ.

11.4.4 Deriving time averages from probability density functions

While we have defined probability density functions as time-averages of functions of a signal, it is also possible to derive the mean from the p.d.f:

$$< x[n] > \; = \int_{-\infty}^{\infty} X f_x(X) dX \tag{11.63}$$

In this context, the term *expected value,* which is often defined by (11.63), is just another name for the mean of a random signal.

A formal proof of (11.63) can easily be given using δ functions. The first step is to write $x[n]$ as a superposition of δ functions, much as we did to derive the convolution formula for linear filters:

$$x[n] \; = \; \int_{-\infty}^{\infty} \; X \; \delta(x[n] \; - \; X) \; dX \tag{11.64}$$

Taking the time average, and noting that, because integrals are linear operations, they commute with time averages, we obtain

$$< x[n] > \; = \int_{-\infty}^{\infty} X \quad < \delta(x[n] - X) > \quad dX \tag{11.65}$$

Remembering from (11.53) that $\quad < \delta(x[n] \; - \; X) > \; = f_x(X)$ completes the proof of (11.63).

We will now sketch a more intuitive proof for (11.63). For this purpose, we return to the definition of the time average:

$$< x[n] > \; = \lim_{N \to \infty} \frac{1}{2N+1} \sum_{n=-N}^{N} x[n] \tag{11.1}$$

The result of summing over $2N + 1$ terms does not depend on the order in which the terms are summed, and, in particular, we can sum over increasing values of the amplitude $x[n]$ rather than over increasing values of time n. Specifically, we divide the range of signal amplitudes into small adjacent intervals of width ΔX centered at $X_i \; = \; i \; \Delta X$

$$\sum_{n=-N}^{N} x[n] \; = \; \sum_{i=-\infty}^{\infty} \; \left[\sum_{X_i - \Delta X/2 \quad < x[n] \; \leq \; X_i + \Delta X/2} x[n] \right] \tag{11.66}$$

Assuming that ΔX is small enough that all values of $x[n]$ within the interval $[X_i - \Delta X/2, \; X_i + \Delta X/2]$ can be approximated by the center of the interval X_i, this sum can be approximated by:

$$\sum_{n=-N}^{N} x[n] \; \approx \; \sum_{i=-\infty}^{\infty} \; X_i \; (\text{ Number of samples in } [-N, N] \text{ for which } X_i - \Delta X/2 \quad < x[n] \; \leq \; X_i + \Delta X/2 \;) \tag{11.67}$$

Dividing this expression by $2N + 1$, and taking the limit for large N, we obtain:

$$< x[n] > \; \approx \; \sum_{i=-\infty}^{\infty} \; X_i \; P\left(X_i - \Delta X/2 \quad < x[n] \; \leq \; X_i + \Delta X/2\right) \tag{11.68}$$

When ΔX becomes very small, the probability that $x[n]$ lies in the interval of width ΔX centered at X_i approaches $f_x(X_i)\ \Delta X$:

$$< x[n] > \ \approx \sum_{i=-\infty}^{\infty} X_i\ f_x(X_i)\ \Delta X \tag{11.69}$$

In the limit, the sum over i approaches an integral over X_i, and we obtain (11.63).

11.5 Memoryless transformations of random signals

11.5.1 P.d.f for a memoryless transformation of a random signal

A major application of probability density functions is that, if the p.d.f. of a signal $x[n]$ is known, then we can compute the p.d.f. for the signal obtained by processing $x[n]$ through an arbitrary memoryless function $g(X)$. Specifically, let $y[n] = g(x[n])$, and assume for simplicity that $g(X)$ is a monotonic function of X. The probability that $x[n]$ lies in the small interval $[X,\ X + dX]$ is equal to the probability that $y[n]$ lies in the interval $[g(X),\ g(X + dX)]$. Defining $Y = g(X)$, this gives:

$$f_x(X)\ dX \ = \ f_y(Y)\ |dY| \ = \ f_y(Y)\ |g(X + dX) \ - \ g(X)| \tag{11.70}$$

Introducing the derivative $g'(X)$, this becomes:

$$f_x(X)\ dX \ = \ f_y(Y)\ |g'(X)|\ dX \tag{11.71}$$

Rearranging terms gives the final result:

$$f_y(Y) \ = \ \frac{f_x(X)}{|g'(X)|} \tag{11.72}$$

In the general case when $g(X)$ is not a monotonic function of X, there is no longer a one-to-one correspondence between X and Y. The p.d.f. $f_y(Y)$ is then equal to a sum of terms as in (11.72) for all the points X_i such that $g(X_i) = Y$.

11.5.2 Examples of memoryless transformations

1. Linear transformation. If $y[n] = a\ x[n] + b$, (11.72) gives:

$$f_y(Y) \ = \ \frac{1}{|a|}\ f_x\left(\frac{Y \ - \ b}{a}\right) \tag{11.73}$$

2. Logarithmic transformation. Assume that $x[n]$ has a uniform p.d.f. between 0 and 1, and that

$$y[n] \ = \ -\log(x[n])/\lambda. \tag{11.74}$$

Two cases need to be considered: If $Y < 0$, there is no X between 0 and 1 such that $Y = -\log(X)$, so that $f_y(Y) = 0$. If $Y \geq 0$, there is always a unique X between 0

and 1 such that $Y = -\log(X)/\lambda$. One then has $X = e^{-\lambda Y}$, and $|g'(X)| = 1/\lambda X$. Furthermore, because $x[n]$ is uniform between 0 and 1, $f_x(X) = 1$. Thus, from (11.44),

$$f_y(Y) = f_x(X)/|g'(X)| = \lambda X = \lambda e^{-\lambda Y} \tag{11.75}$$

Combining the two cases when $Y < 0$ and $Y \geq 0$, we conclude that $y[n]$ has an exponential p.d.f with parameter λ:

$$f_y(Y) = \lambda e^{-\lambda Y} u(Y) \tag{11.76}$$

This example illustrates how a uniform p.d.f can be changed into an exponential p.d.f. by an appropriate memoryless transformation. Such techniques can be used for generating signals with any desired p.d.f. using a random number generator that produces a signal with a uniform p.d.f.

11.5.3 Mean of a function of a random signal

Taken together, (11.63) and (11.72) make it possible to compute the mean of an arbitrary memoryless transformation of a random signal. Specifically, let $y[n] = g(x[n])$. From (11.72), we have:

$$< g(x[n]) > = < y[n] > = \int_{-\infty}^{\infty} Y f_y(Y) dY \tag{11.77}$$

Using (11.63), we can express $f_y(Y)\ dY$ as a function of $f_x(X)$, where $X = g^{-1}(Y)$, so that $|dX| = dY/|g'(X)|$:

$$f_y(Y)\ dY = f_x(X)/|g'(X)|\ dY = f_x(X)\ |dX| \tag{11.78}$$

Replacing $f_y(Y)\ dY$ by its value gives the desired result:

$$< g(x[n]) > = \int_{-\infty}^{\infty} g(X)\ f_x(X)\ dX \tag{11.79}$$

11.5.4 Example: Mean power of quantization noise

As an example of the application of (11.79), we will compute the mean power of the quantization error signal introduced in Chapter 1. It was argued that the error signal $q[n]$ is equally likely to take any values between $-Q/2$ and $Q/2$, where Q is the quantization step. In other words, its p.d.f. is uniform:

$$f_q(X) = \begin{cases} 1/Q & \text{if } -Q/2 < X \leq Q/2 \\ 0 & \text{otherwise} \end{cases} \tag{11.80}$$

Applying (11.79) to the function $g(X) = X^2$ in order to obtain the mean power:

$$P_q \triangleq < q[n]^2 > = \int_{-\infty}^{\infty} X^2\ f_q(X)\ dX = \frac{1}{Q} \int_{-Q/2}^{Q/2} X^2\ dX = \frac{Q^2}{12} \tag{11.81}$$

11.6 Summary

A signal is said to be random if it is characterized in terms of average properties rather than complete mathematical specifications. For our purposes, it suffices to consider time averages defined by:

$$< x[n] > \stackrel{\triangle}{=} \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} x[n]$$

Time averages possess the two basic properties of linearity and stationarity.

There are two important classes of systems for which it is possible to compute time averages of the output signal in terms of appropriate averages of the input: linear, time-invariant systems, and memoryless systems.

For linear systems, the autocorrelation function $R_x[k] = < x[n]x[n-k] >$ plays a key role in predicting responses to random signals. The autocorrelation function is an even function of lag k that is always maximum at the origin, where its value equals the mean power P_x. It is useful for identifying unknown periodicities in random signals. A random signal is said to be white if its autocovariance function is an impulse at the origin.

When two random signals are jointly present, we need to know not only the autocorrelation function of each signal, but also their crosscorrelation function $R_{xy}[k] = < x[n]y[n+k] >$. The crosscorrelation function is related to the autocorrelation functions by the Cauchy-Schwarz inequality:

$$|R_{xy}[k]|^2 \leq R_x[0]\ R_y[0]$$

Crosscorrelation functions are useful for time-delay estimation and for detecting periodic signals with known period in additive noise.

For memoryless systems, the appropriate set of time averages is the probability density function $f_x(X)$:

$$f_x(X) \stackrel{\triangle}{=} \lim_{\Delta X\to 0} \frac{1}{\Delta X} \quad < \Pi_{\Delta X}(x[n] - X) >$$

The p.d.f. can be used to compute time averages of outputs of arbitrary memoryless systems $y[n] = g(x[n])$ by means of the formula:

$$< g(x[n]) > = \int_{-\infty}^{\infty} g(X)\ f_x(X)\ dX$$

Further, if $g(X)$ is monotonic, the p.d.f. for the output signal is

$$f_y(Y) = \frac{f_x(X)}{|g'(X)|}, \qquad \text{where } X \stackrel{\triangle}{=} g^{-1}(Y)$$

11.A Ensemble averages

A limitation of the preceding results is that, because time averages eliminate all information on the time variations of the statistical characteristics of the signals, the time-average model is

only useful for *stationary* random signals. There are however signals for which variations are too rapid for stationarity to hold, even over limited times. For such non-stationary signals, we can use a more abstract probabilistic model based on *ensemble averages* rather than time averages.

To introduce the notion of ensemble average, consider a set of particles of gas in a rigid box. Each particle is constantly undergoing random Brownian motion. Conceptually, the trajectory of a particle can be followed over time, and its average position computed: this is the familiar time average. Alternatively, the average position of all the particles in the box can be determined at a specific time: This is called an *ensemble average.* Thus, the position of gas particles can be considered as a random function of two variables: time, and an ensemble variable identifying each particle. To be more specific, for each time t, we can compute the fraction of particles that lie between the positions X and $X + \Delta X$ along the vertical axis. In the limit when ΔX tends to zero, this defines a time-dependent *probability density function* $f_x(X;\ t)$. The ensemble average position $E\left(x(t)\right)$ along the vertical axis, or *expected value* of the position can then be computed by means of the formula

$$E\left(x(t)\right) \triangleq \int_{-\infty}^{\infty} X\ f_x(X;\ t)\ dX \tag{11.A.1}$$

This expression is formally similar to (11.63), which expresses the time-average of a random signal in terms of its p.d.f., but the two formulas have very different interpretations. In (11.52), the p.d.f. was *defined* as a time average, so that (11.63) expresses the consistency between two time averages. In contrast, (11.A.1) is a *definition* of the expected value from the ensemble-average p.d.f. $f_x(X;\ t)$ which is *a priori given* for each time t. Thus, the p.d.f. in (11.52) gives no information about how the position of an individual particle might change over time, while that in (11.A.1) does. This distinction is made clear by the fact that the ensemble-average p.d.f. $f_x(X;\ t)$ depends explicitly on time t in addition to the position variable X, while the time-average p.d.f. in (11.52) does not depend on time.

The notion of ensemble average can be extended to signals for which the definition of an ensemble is less obvious than for particles in a box. Suppose, for example that we record the EEG signal from the scalp of a patient. If we obtained another EEG record under identical experimental conditions, the signal would be different, but would share certain statistical regularities for every recording. We are thus led to consider that each EEG record is an element of the ensemble of signals that could have been recorded. In this case, the ensemble is not directly observable, it is an abstraction, a model characterizing our uncertainty about the signal. Similarly, we could consider each epoch of an evoked potential signal to be an element from an ensemble of signals, even if the epochs are in fact different segments from the same recording. A major advantage of introducing this abstract signal ensemble is that it allows us to model the time-dependence of the statistical characteristics of the signal. For example, for the evoked potential, we can now express the fact that the mean and variance may not be the same for every sample in the epoch. Despite the increased complexity of the ensemble-average model over the time-average model, formal results are often very similar for both models because they both use probability density functions.

11.A.1 Discrete Wiener process

To illustrate the usefulness of the ensemble average model, consider a zero-mean, stationary, white, Gaussian random signal $x[n]$ with variance σ^2. Its p.d.f. is:

$$f_x(X;\ n)\ =\ \frac{1}{\sqrt{2\pi}\ \sigma}\ e^{-\ \frac{X^2}{2\ \sigma^2}} \tag{11.A.2}$$

For this signal, the p.d.f. does not depend on time, so that the time-average model and the ensemble-average model give identical results. We now define a new signal $y[n]$ by forming the sum of all samples of $x[n]$ up to time n:

$$y[n]\ \stackrel{\triangle}{=} \begin{cases} 0 & \text{if } n \leq 0 \\ \sum_{i=1}^{n} x[i] & \text{if } n > 0 \end{cases} \tag{11.A.3}$$

We will compute the expected value of $y[n]$ as well as its autocorrelation function $R_y[n,m]\ \stackrel{\triangle}{=}\ E(y[n]\ y[m])$. In this case, the autocorrelation is a function of two variables because stationarity does not hold.

$$E(y[n])\ =\ \sum_{i=1}^{n} E(x[i])\ =\ 0 \tag{11.A.4}$$

$$E(y[n]\ y[m])\ =\ E\left(\left[\ \sum_{i=1}^{n}\ x[i]\ \right]\left[\ \sum_{j=1}^{m}\ x[j]\ \right]\right)\ =\ \sum_{i=1}^{n}\sum_{j=1}^{m} E(x[i]\ x[j]) \tag{11.A.5}$$

Because $x[n]$ is white and zero-mean, $E(x[i]\ x[j])$ is zero unless $i\ =\ j$, in which case it is σ^2. Among the $n\ \times\ m$ products in the expression for $E(y[n]\ y[m])$, the number of terms for which $i\ =\ j$ is the minimum of n and m. Therefore,

$$R_y[n,\ m]\ =\ E(y[n]\ y[m])\ =\ \sigma^2\ min(n,\ m) \tag{11.A.6}$$

In particular, the variance of $y[n]$ is:

$$\sigma_y[n]\ =\ E(y[n]^2)\ =\ n\ \sigma^2 \tag{11.A.7}$$

Thus, the variance of $y[n]$ increases linearly with time. Because $y[n]$ is a sum of n Gaussian signals, it is also Gaussian, and its probability density function is

$$f_y(Y;\ n)\ =\ \frac{1}{\sqrt{2\pi n}\ \sigma}\ e^{-\ \frac{Y^2}{2n\sigma^2}} \tag{11.A.8}$$

The random signal $y[n]$ is called a discrete-time *Wiener process.* It is a good model for the Brownian motion of a particle. Time-averages are not applicable to Wiener processes because these signals tend to wander farther and farther away from their mean value, as the expression for the variance shows. Figure 11.8 shows the waveform of a discrete-time Wiener process.

Although the ensemble average model is more powerful than the time average model, it is not always easy to make practical use of this increased power because, usually, a single record of a random signal is available, so that only time averages can be estimated from the data. In fact, many useful predictions of the ensemble average model apply only to stationary, ergodic signals, for which the two models are equivalent.

11.7 Further reading

Siebert: Chapter 19.
Oppenheim and Schafer: Chapter 2, Section 10, and Appendix A
Papoulis and Pillai: Chapters 7, 9

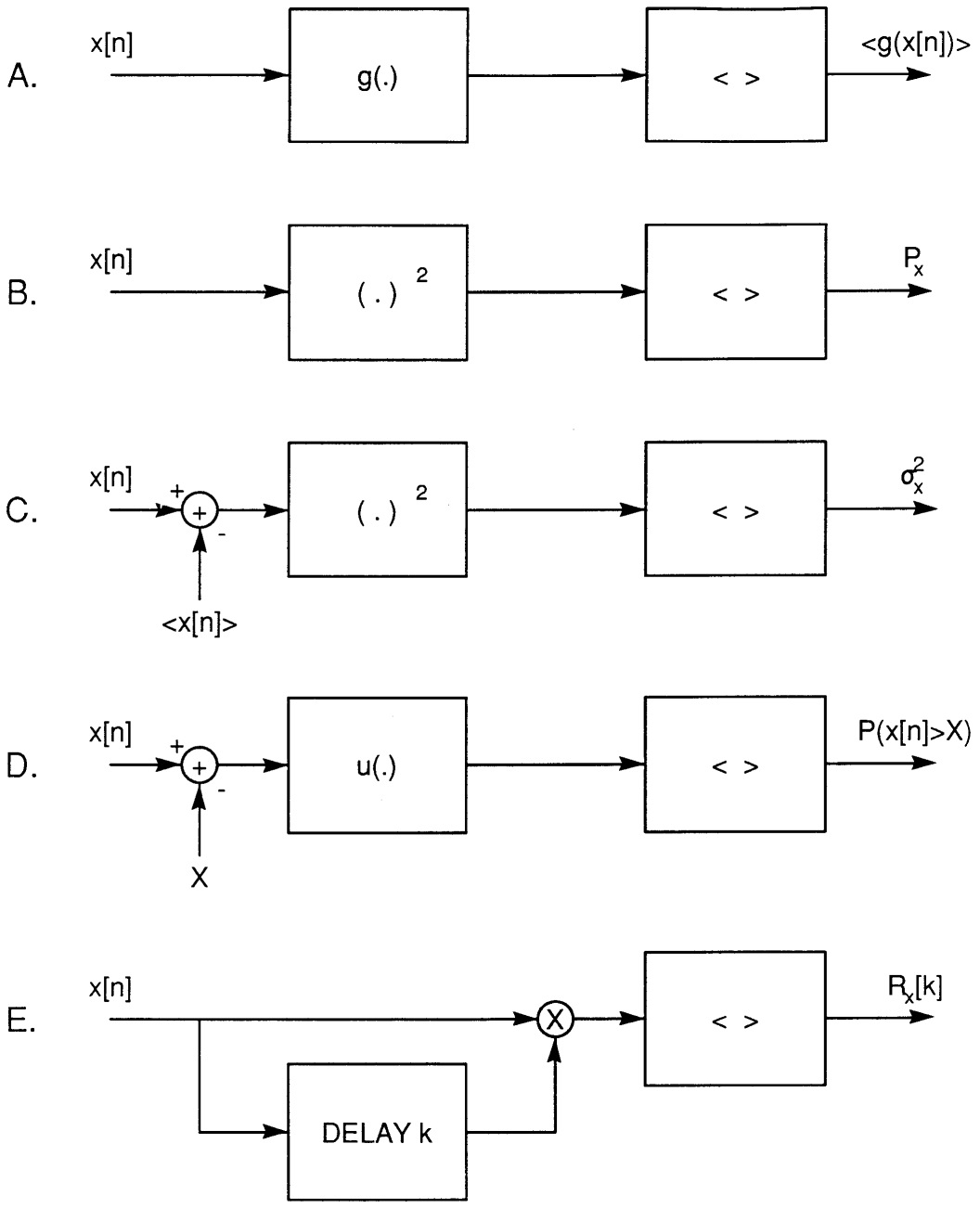

Figure 11.1: Computation of time averages of random signals. A. Average of a memoryless function $g(x[n])$. B. Mean Power. C. Variance. D. $P\left(x[n] \geq X\right)$. E. Autocorrelation function.

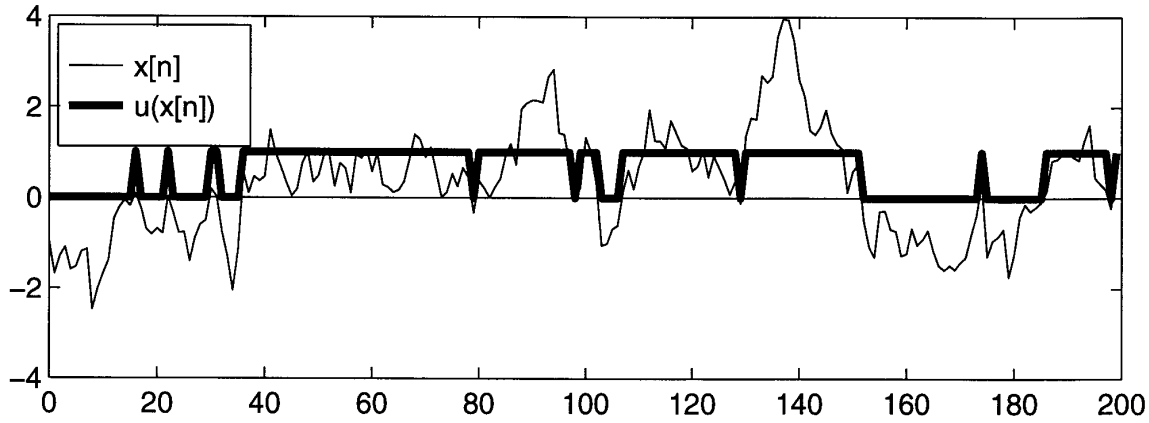

Figure 11.2: Use of step function $u(x[n])$ in computing $P\left(x[n] \geq 0\right)$.

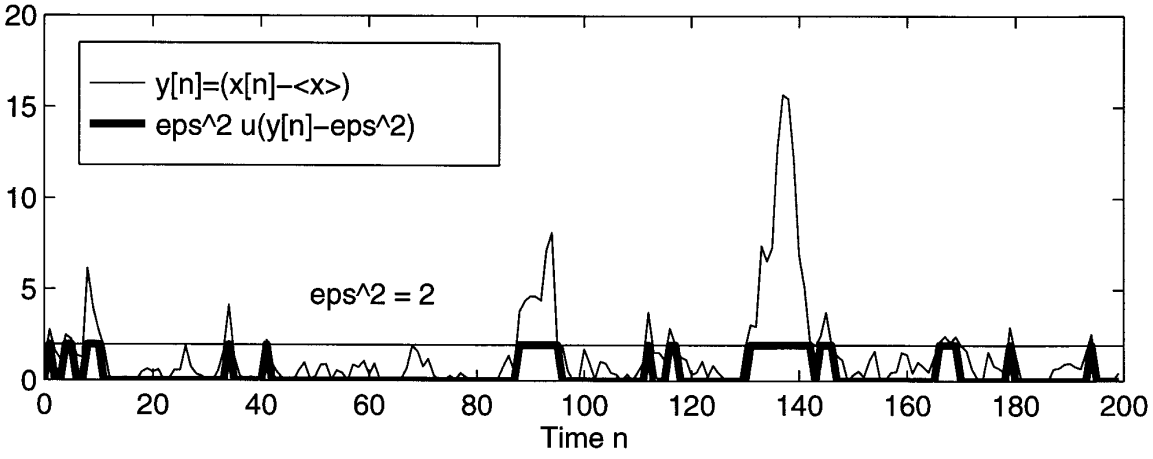

Figure 11.3: Signals used in proof of Chebyshev's inequality

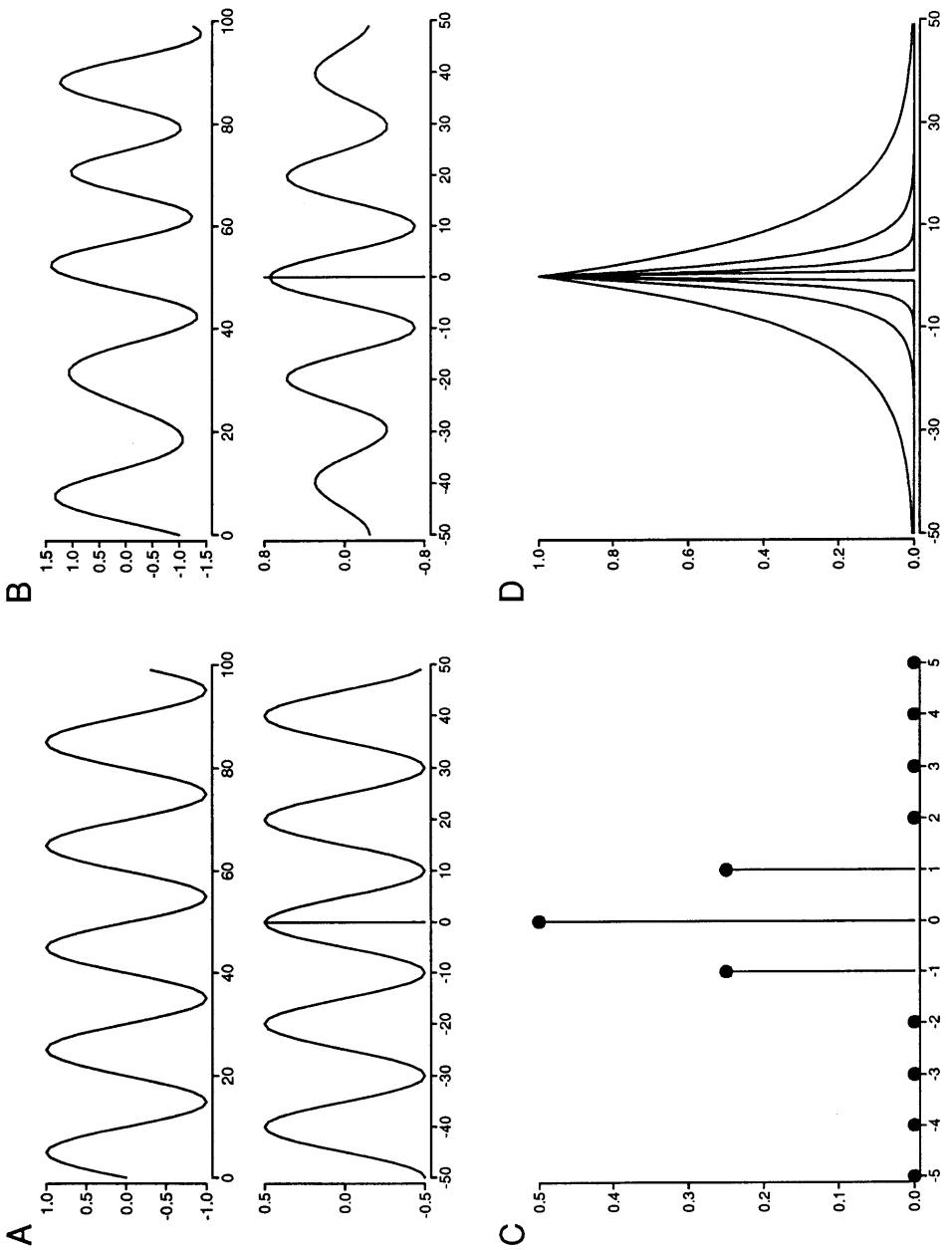

Figure 11.4: A. Waveform and autocorrelation function of the sine wave $\sin(0.1\pi fn)$. B. Same as in A. for the three component signal $\sin(0.1\pi fn) + 0.5\sin(0.08\pi fn) + 0.5\sin(0.12\pi fn)$. C. Autocorrelation function of the FIR filter $y[n] = 0.5(x[n] + x[n-1])$ for white noise input. D. Same as in C. for first-order recursive filter $y[n] = ay[n-1] + x[n]$. The autocorrelation function is shown for $a = 0, 0.5, 0.75, 0.9$.

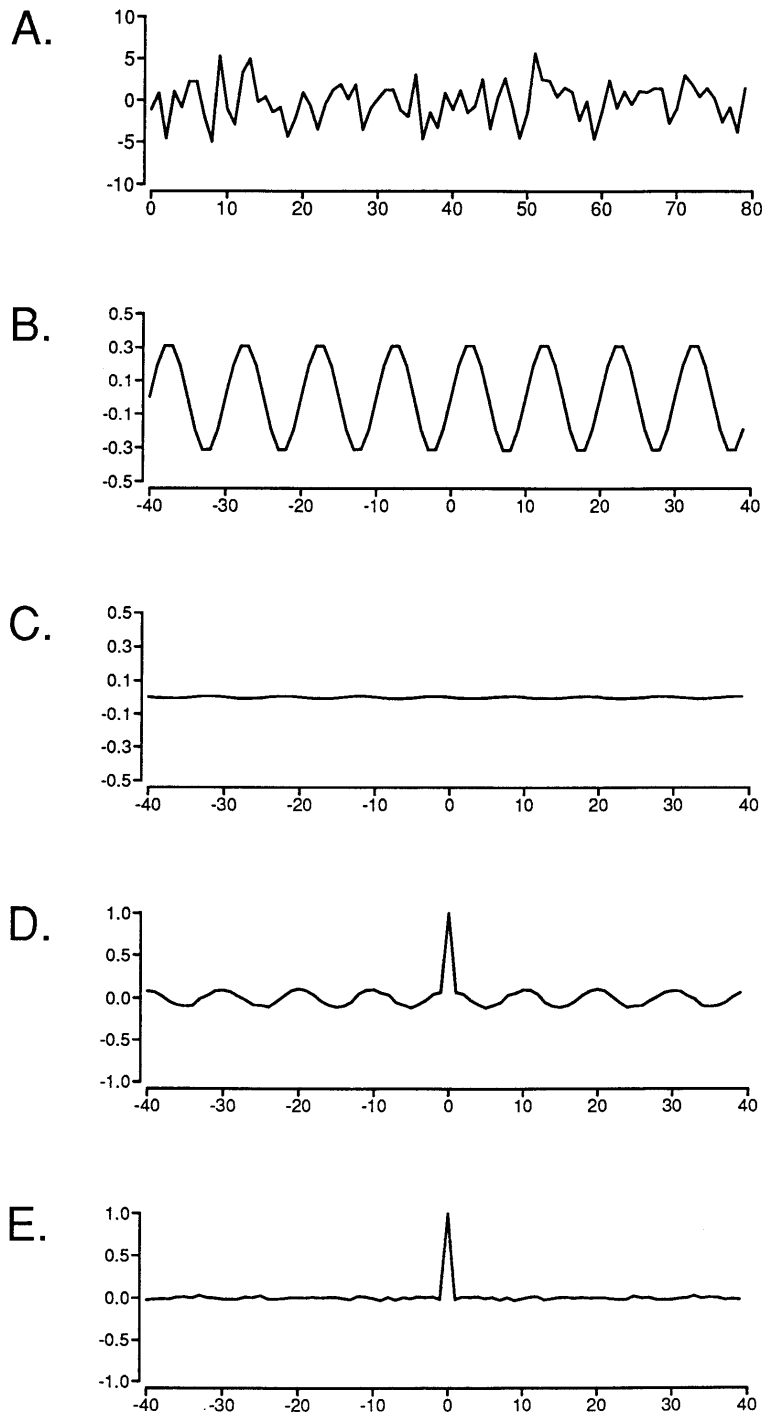

Figure 11.5: A. Waveform of a sine wave buried in additive, white noise. B. Estimated cross-correlation function between the signal in A. and the unit cosine. The estimate is based on 4000 samples. C. Cross-correlation between white noise and unit cosine. D. Estimated autocorrelation function for the signal in A. E. Autocorrelation function of the white noise.

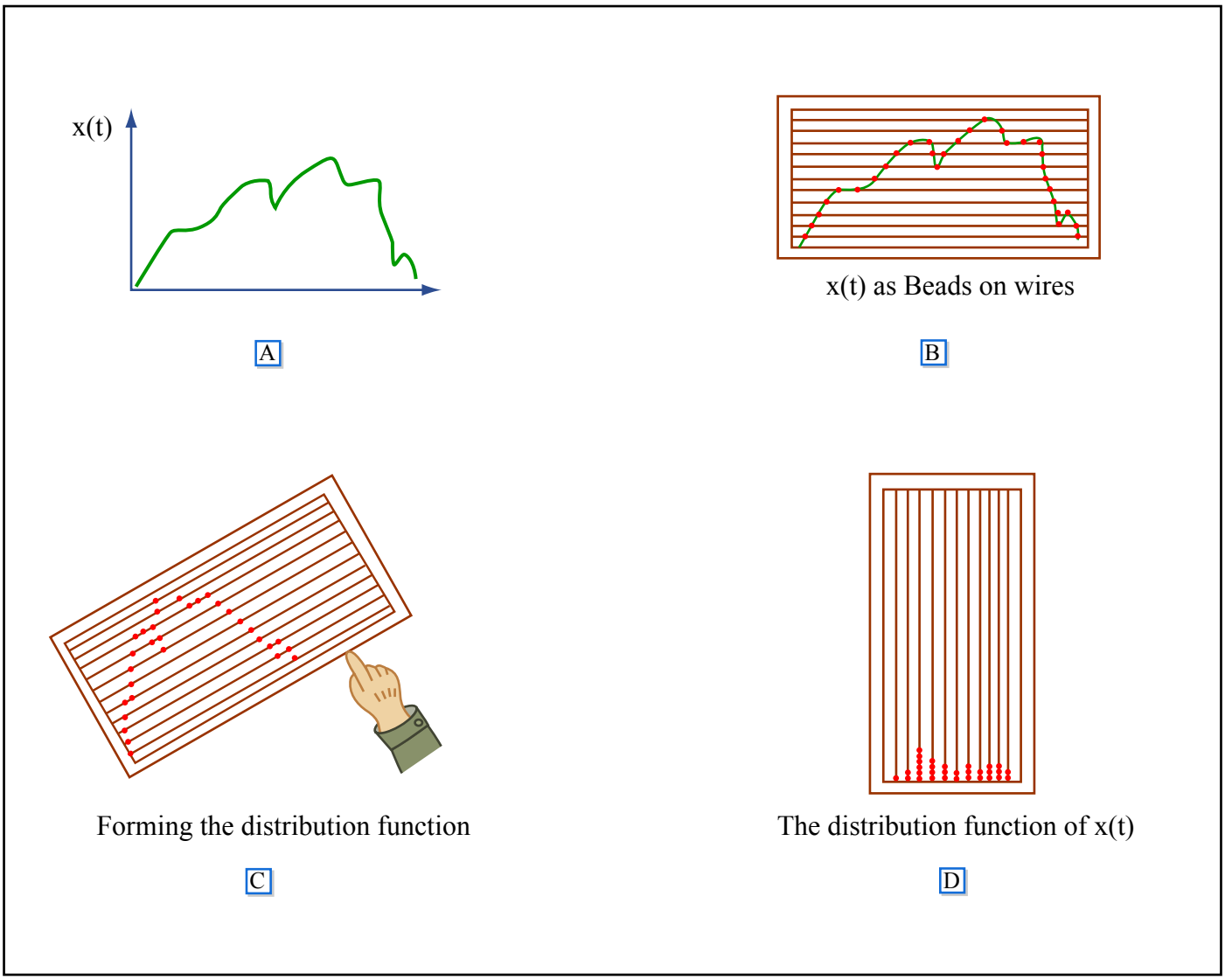

Figure by MIT OpenCourseWare.

Figure 11.6: Measurement of the probability density function of a random signal

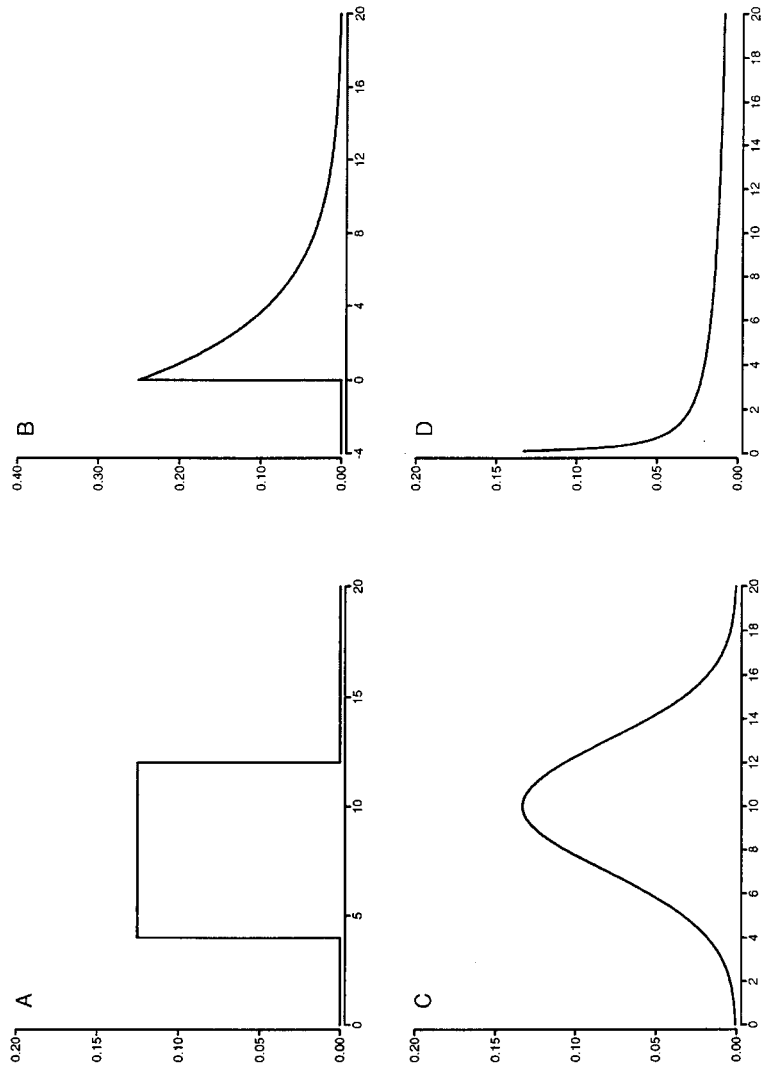

Figure 11.7: Common probability density functions. A. Uniform p.d.f. for $a = 4$ and $b = 12$. B. Exponential p.d.f. for $\lambda = 1/40$. C. Gaussian p.d.f. for $\mu = 10$ and $\sigma = 3$. D. Chi-squared p.d.f. for $\sigma = 3$.

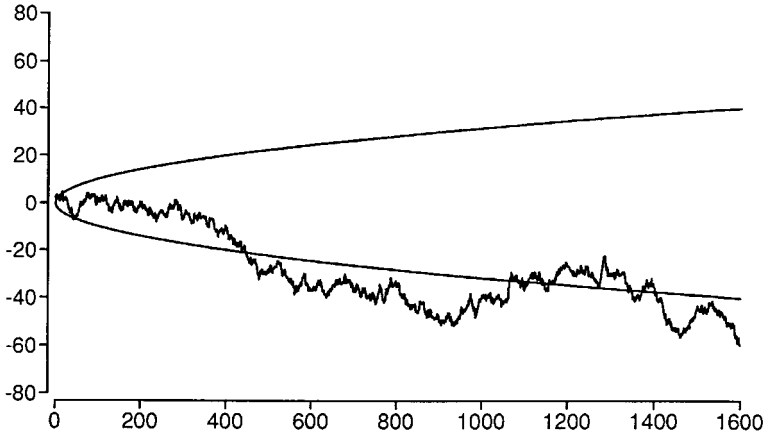

Figure 11.8: Waveform of discrete-time Wiener process with $\sigma = 1$. Smooth lines show +/-1 standard deviation