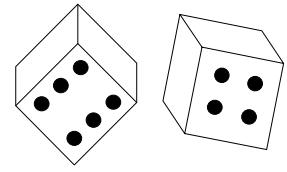
Advanced Digital Signal Processing and Noise Reduction, Second Edition.
Saeed V. Vaseghi

ISBNs: 0-471-62692-9 (Hardback): 0-470-84162-1 (Electronic)



3

The small probability of collision of the Earth and a comet can become very great in adding over a long sequence of centuries. It is easy to picture the effects of this impact on the Earth. The axis and the motion of rotation have changed, the seas abandoning their old position...

Pierre-Simon Laplace

PROBABILITY MODELS

Probability models form the foundation of information theory. Information itself is quantified in terms of the logarithm of probability. Probability models are used to characterise and predict the occurrence of random events in such diverse areas of applications as predicting the number of telephone calls on a trunk line in a specified period of the day, road traffic modelling, weather forecasting, financial data modelling, predicting the effect of drugs given data from medical trials, etc. In signal processing, probability models are used to describe the variations of random signals in applications such as pattern recognition, signal coding and signal estimation. This chapter begins with a study of the basic concepts of random signals and stochastic processes and the models that are used for the characterisation of random processes. Stochastic processes are classes of signals whose fluctuations in time are partially or completely random, such as speech, music, image, time-varying channels, noise and video. Stochastic signals are completely described in terms of a probability model, but can also be characterised with relatively simple statistics, such as the mean, the correlation and the power spectrum. We study the concept of ergodic stationary processes in which time averages obtained from a single realisation of a process can be used instead of ensemble averages. We consider some useful and widely used classes of random signals, and study the effect of filtering or transformation of a signal on its probability distribution.

3.1 Random Signals and Stochastic Processes

Signals, in terms of one of their most fundamental characteristics, can be classified into two broad categories: *deterministic* signals and *random* signals. Random functions of time are often referred to as *stochastic* signals. In each class, a signal may be continuous or discrete in time, and may have continuous-valued or discrete-valued amplitudes.

A deterministic signal can be defined as one that traverses a predetermined trajectory in time and space. The exact fluctuations of a deterministic signal can be completely described in terms of a function of time, and the exact value of the signal at any time is predictable from the functional description and the past history of the signal. For example, a sine wave $x(t)$ can be modelled, and accurately predicted either by a second-order linear predictive model or by the more familiar equation $x(t)=A\sin(2\pi ft+\phi)$.

Random signals have unpredictable fluctuations; hence it is not possible to formulate an equation that can predict the *exact* future value of a random signal from its past history. Most signals such as speech and noise are at least in part random. The concept of randomness is closely associated with the concepts of information and noise. Indeed, much of the work on the processing of random signals is concerned with the extraction of information from noisy observations. If a signal is to have a capacity to convey information, it must have a degree of randomness: a predictable signal conveys no information. Therefore the random part of a signal is either the information content of the signal, or noise, or a mixture of both information and noise. Although a random signal is not completely predictable, it often exhibits a set of well-defined statistical characteristic values such as the maximum, the minimum, the mean, the median, the variance and the power spectrum. A random process is described in terms of its statistics, and most completely in terms of a probability model from which all its statistics can be calculated.

Example 3.1 Figure 3.1(a) shows a block diagram model of a deterministic discrete-time signal. The model generates an output signal $x(m)$ from the P past samples as

$$x(m)=h_1\big(x(m-1),x(m-2),...,x(m-P)\big) \tag{3.1}$$

where the function h_1 may be a linear or a non-linear model. A functional description of the model h_1 and the P initial sample values are all that is required to predict the future values of the signal $x(m)$. For example for a sinusoidal signal generator (or oscillator) Equation (3.1) becomes

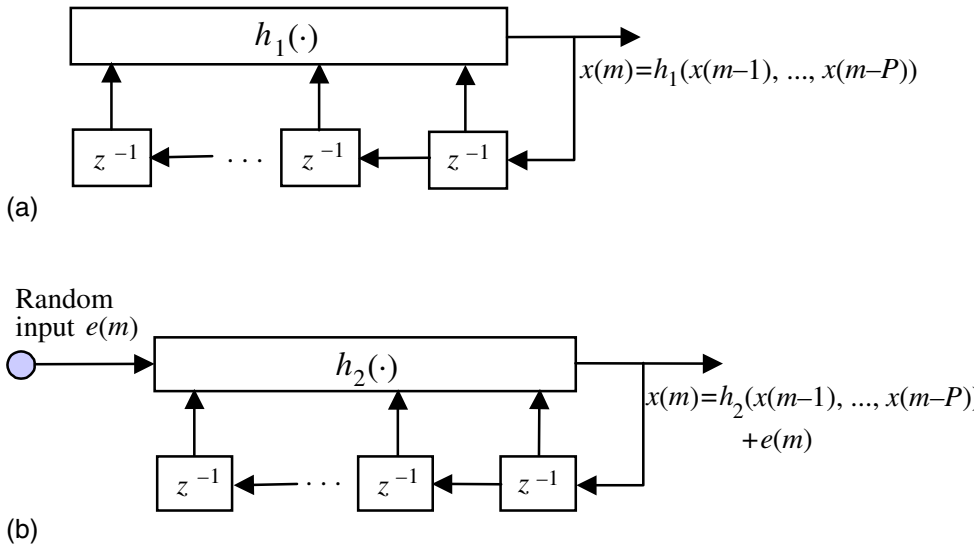


Figure 3.1 Illustration of deterministic and stochastic signal models: (a) a deterministic signal model, (b) a stochastic signal model.

$$x(m)= a\,x(m-1)- x(m-2) \tag{3.2}$$

where the choice of the parameter a=2cos(2πF_0 /F_s) determines the oscillation frequency F_0 of the sinusoid, at a sampling frequency of F_s. Figure 3.1(b) is a model for a stochastic random process given by

$$x(m)=h_2\big(x(m-1),x(m-2),...,x(m-P)\big)+e(m) \tag{3.3}$$

where the random input $e(m)$ models the unpredictable part of the signal $x(m)$, and the function h_2 models the part of the signal that is correlated with the past samples. For example, a narrowband, second-order autoregressive process can be modelled as

$$x(m)= a_1\,x(m-1)+ a_2\,x(m-2)+e(m) \tag{3.4}$$

where the choice of the parameters a_1 and a_2 will determine the centre frequency and the bandwidth of the process.

3.1.1 Stochastic Processes

The term "stochastic process" is broadly used to describe a random process that generates sequential signals such as speech or noise. In signal processing terminology, a stochastic process is a probability model of a class of random signals, e.g. Gaussian process, Markov process, Poisson process, etc. The classic example of a stochastic process is the so-called Brownian motion of particles in a fluid. Particles in the space of a fluid move randomly due to bombardment by fluid molecules. The random motion of each particle is a single realisation of a stochastic process. The motion of all particles in the fluid forms the collection or the space of different realisations of the process.

In this chapter, we are mainly concerned with discrete-time random processes that may occur naturally or may be obtained by sampling a continuous-time band-limited random process. The term "discrete-time stochastic process" refers to a class of discrete-time random signals, $X(m)$, characterised by a probabilistic model. Each realisation of a discrete stochastic process $X(m)$ may be indexed in time and space as $x(m,s)$, where m is the discrete time index, and s is an integer variable that designates a space index to each realisation of the process.

3.1.2 The Space or Ensemble of a Random Process

The collection of all realisations of a random process is known as the ensemble, or the space, of the process. For an illustration, consider a random noise process over a telecommunication network as shown in Figure 3.2. The noise on each telephone line fluctuates randomly with time, and may be denoted as $n(m,s)$, where m is the discrete time index and s denotes the line index. The collection of noise on different lines form the ensemble (or the space) of the noise process denoted by $N(m)=\{n(m,s)\}$, where $n(m,s)$ denotes a realisation of the noise process $N(m)$ on the line s. The "true" statistics of a random process are obtained from the averages taken over the ensemble of many different realisations of the process. However, in many practical cases, only one realisation of a process is available. In Section 3.4, we consider the so-called ergodic processes in which time-averaged statistics, from a single realisation of a process, may be used instead of the ensemble-averaged statistics.

Notation The following notation is used in this chapter: $X(m)$ denotes a random process, the signal $x(m,s)$ is a particular realisation of the process $X(m)$, the random signal $x(m)$ is any realisation of $X(m)$, and the collection

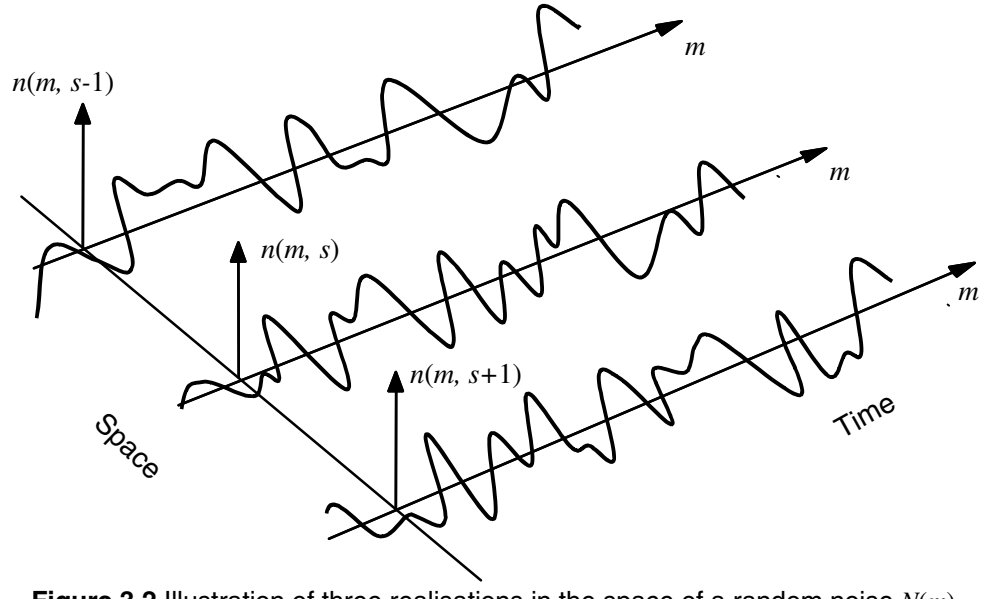


Figure 3.2 Illustration of three realisations in the space of a random noise $N(m)$.

of all realisations of $X(m)$, denoted by $\{x(m,s)\}$, form the ensemble or the space of the random process $X(m)$.

3.2 Probabilistic Models

Probability models provide the most complete mathematical description of a random process. For a fixed time instant m, the collection of sample realisations of a random process $\{x(m,s)\}$ is a random variable that takes on various values across the space s of the process. The main difference between a random variable and a random process is that the latter generates a time series. Therefore, the probability models used for random variables may also be applied to random processes. We start this section with the definitions of the probability functions for a random variable.

The space of a random variable is the collection of all the values, or outcomes, that the variable can assume. The space of a random variable can be partitioned, according to some criteria, into a number of subspaces. A subspace is a collection of signal values with a common attribute, such as a cluster of closely spaced samples, or the collection of samples with their amplitude within a given band of values. Each subspace is called an event, and the probability of an event A, $P(A)$, is the ratio of the number of

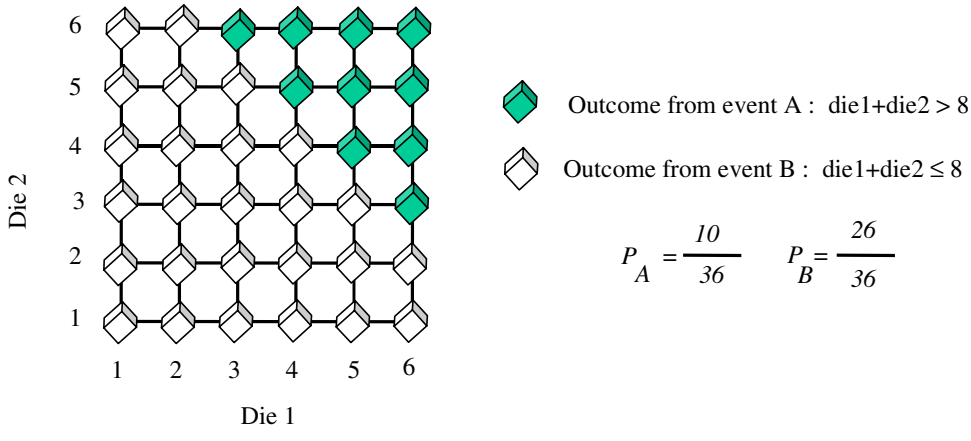


Figure 3.3 A two-dimensional representation of the outcomes of two dice, and the subspaces associated with the events corresponding to the sum of the dice being greater than 8 or, less than or equal to 8.

observed outcomes from the space of A, N_A, divided by the total number of observations:

$$P(A) = \frac{N_A}{\sum_{\text{All events } i} N_i} \tag{3.5}$$

From Equation (3.5), it is evident that the sum of the probabilities of all likely events in an experiment is unity.

Example 3.2 The space of two discrete numbers obtained as outcomes of throwing a pair of dice is shown in Figure 3.3. This space can be partitioned in different ways; for example, the two subspaces shown in Figure 3.3 are associated with the pair of numbers that add up to less than or equal to 8, and to greater than 8. In this example, assuming the dice are not loaded, all numbers are equally likely, and the probability of each event is proportional to the total number of outcomes in the space of the event.

3.2.1 Probability Mass Function (pmf)

For a discrete random variable X that can only assume discrete values from a finite set of N numbers $\{x_1, x_2, ..., x_N\}$, each outcome x_i may be considered as an event and assigned a probability of occurrence. The probability that a

discrete-valued random variable X takes on a value of x_i, $P(X= x_i)$, is called the *probability mass function* (*pmf*). For two such random variables X and Y, the probability of an outcome in which X takes on a value of x_i and Y takes on a value of y_j, $P(X=x_i, Y=y_j)$, is called the joint probability mass function. The joint pmf can be described in terms of the conditional and the marginal probability mass functions as

$$\begin{aligned} P_{X,Y}(x_i, y_j) &= P_{Y|X}(y_j | x_i) P_X(x_i) \\ &= P_{X|Y}(x_i | y_j) P_Y(y_j) \end{aligned} \tag{3.6}$$

where $P_{Y|X}(y_j | x_i)$ is the probability of the random variable Y taking on a value of y_j conditioned on X having taken a value of x_i, and the so-called marginal pmf of X is obtained as

$$\begin{aligned} P_X(x_i) &= \sum_{j=1}^{M} P_{X,Y}(x_i, y_j) \\ &= \sum_{j=1}^{M} P_{X|Y}(x_i | y_j) P_Y(y_j) \end{aligned} \tag{3.7}$$

where M is the number of values, or outcomes, in the space of the discrete random variable Y. From Equations (3.6) and (3.7), we have *Bayes' rule* for the conditional probability mass function, given by

$$\begin{aligned} P_{X|Y}(x_i | y_j) &= \frac{1}{P_Y(y_j)} P_{Y|X}(y_j | x_i) P_X(x_i) \\ &= \frac{P_{Y|X}(y_j | x_i) P_X(x_i)}{\sum_{i=1}^{M} P_{Y|X}(y_j | x_i) P_X(x_i)} \end{aligned} \tag{3.8}$$

3.2.2 Probability Density Function (pdf)

Now consider a continuous-valued random variable. A continuous-valued variable can assume an infinite number of values, and hence, the probability that it takes on a given value vanishes to zero. For a continuous-valued

random variable X the cumulative distribution function (cdf) is defined as the probability that the outcome is less than x as:

$$F_X(x) = Prob(X \leq x) \tag{3.9}$$

where $Prob(\cdot)$ denotes probability. The probability that a random variable X takes on a value within a band of Δ centred on x can be expressed as

$$\begin{aligned}\frac{1}{\Delta} Prob(x-\Delta/2 \leq X \leq x+\Delta/2) &= \frac{1}{\Delta}[Prob(X \leq x+\Delta/2) - Prob(X \leq x-\Delta/2)] \\ &= \frac{1}{\Delta}[F_X(x+\Delta/2) - F_X(x-\Delta/2)]\end{aligned} \tag{3.10}$$

As Δ tends to zero we obtain the *probability density function* (*pdf*) as

$$\begin{aligned}f_X(x) &= \lim_{\Delta \to 0} \frac{1}{\Delta}[F_X(x+\Delta/2) - F_X(x-\Delta/2)] \\ &= \frac{\partial F_X(x)}{\partial x}\end{aligned} \tag{3.11}$$

Since $F_X(x)$ increases with x, the pdf of x, which is the rate of change of $F_X(x)$ with x, is a non-negative-valued function; i.e. $f_X(x) \geq 0$. The integral of the pdf of a random variable X in the range $\pm\infty$ is unity:

$$\int_{-\infty}^{\infty} f_X(x)dx = 1 \tag{3.12}$$

The conditional and marginal probability functions and the Bayes rule, of Equations (3.6)–(3.8), also apply to probability density functions of continuous-valued variables.

Now, the probability models for random variables can also be applied to random processes. For a continuous-valued random process $X(m)$, the simplest probabilistic model is the univariate pdf $f_{X(m)}(x)$, which is the probability density function that a sample from the random process $X(m)$ takes on a value of x. A bivariate pdf $f_{X(m)X(m+n)}(x_1, x_2)$ describes the probability that the samples of the process at time instants m and $m+n$ take on the values x_1, and x_2 respectively. In general, an M-variate pdf

$f_{X(m_1)X(m_2)\cdots X(m_M)}(x_1, x_2, \ldots, x_M)$ describes the pdf of M samples of a random process taking specific values at specific time instants. For an M-variate pdf, we can write

$$\int_{-\infty}^{\infty} f_{X(m_1)\cdots X(m_M)}(x_1, \ldots, x_M)\, dx_M = f_{X(m_1)\cdots X(m_{M-1})}(x_1, \ldots, x_{M-1}) \tag{3.13}$$

and the sum of the pdfs of all possible realisations of a random process is unity, i.e.

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X(m_1)\cdots X(m_M)}(x_1, \ldots, x_M)\, dx_1 \ldots dx_M = 1 \tag{3.14}$$

The probability of a realisation of a random process at a specified time instant may be conditioned on the value of the process at some other time instant, and expressed in the form of a conditional probability density function as

$$f_{X(m)|X(n)}(x_m | x_n) = \frac{f_{X(n)|X(m)}(x_n | x_m) f_{X(m)}(x_m)}{f_{X(n)}(x_n)} \tag{3.15}$$

If the outcome of a random process at any time is independent of its outcomes at other time instants, then the random process is uncorrelated. For an uncorrelated process a multivariate pdf can be written in terms of the products of univariate pdfs as

$$f_{[X(m_1)\cdots X(m_M)|X(n_1)\cdots X(n_N)]}(x_{m_1}, \ldots, x_{m_M} | x_{n_1}, \ldots, x_{n_N}) = \prod_{i=1}^{M} f_{X(m_i)}(x_{m_i}) \tag{3.16}$$

Discrete-valued stochastic processes can only assume values from a finite set of allowable numbers [x_1, x_2, ..., x_n]. An example is the output of a binary message coder that generates a sequence of 1s and 0s. Discrete-time, discrete-valued, stochastic processes are characterised by multivariate probability mass functions (pmf) denoted as

$$P_{[x(m_1)\cdots x(m_M)]}(x(m_1) = x_i, \ldots, x(m_M) = x_k) \tag{3.17}$$

The probability that a discrete random process $X(m)$ takes on a value of x_m at time instant m can be conditioned on the process taking on a value x_n at some other time instant n, and expressed in the form of a conditional pmf as

$$P_{X(m)|X(n)}(x_m|x_n)=\frac{P_{X(n)|X(m)}(x_n|x_m)P_{X(m)}(x_m)}{P_{X(n)}(x_n)} \tag{3.18}$$

and for a statistically independent process we have

$$P_{[X(m_1)\cdots X(m_M)|X(n_1)\cdots X(n_N)]}(x_{m_1},\ldots,x_{m_M}|x_{n_1},\ldots,x_{n_N})=\prod_{i=1}^{M}P_{X(m_i)}(X(m_i)=x_{m_i}) \tag{3.19}$$

3.3 Stationary and Non-Stationary Random Processes

Although the amplitude of a signal $x(m)$ fluctuates with time m, the characteristics of the process that generates the signal may be time-invariant (stationary) or time-varying (non-stationary). An example of a non-stationary process is speech, whose loudness and spectral composition changes continuously as the speaker generates various sounds. A process is stationary if the parameters of the probability model of the process are time-invariant; otherwise it is non-stationary (Figure 3.4). The stationarity property implies that all the parameters, such as the mean, the variance, the power spectral composition and the higher-order moments of the process, are time-invariant. In practice, there are various degrees of stationarity: it may be that one set of the statistics of a process is stationary, whereas another set is time-varying. For example, a random process may have a time-invariant mean, but a time-varying power.

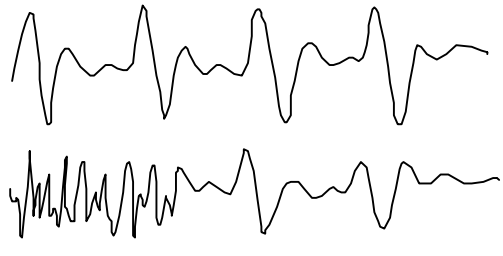


Figure 3.4 Examples of a quasistationary and a non-stationary speech segment.

Example 3.3 In this example, we consider the *time-averaged* values of the mean and the power of: (a) a stationary signal $A\sin\omega t$ and (b) a transient signal $Ae^{-\alpha t}$.
The mean and power of the sinusoid are

$$Mean(A\sin\omega t) = \frac{1}{T}\int_T A\sin\omega t\, dt = 0, \qquad \text{constant} \tag{3.20}$$

$$Power(A\sin\omega t) = \frac{1}{T}\int_T A^2\sin^2\omega t\, dt = \frac{A^2}{2}, \qquad \text{constant} \tag{3.21}$$

Where T is the period of the sine wave. The mean and the power of the transient signal are given by:

$$Mean(Ae^{-\alpha t}) = \frac{1}{T}\int_t^{t+T} Ae^{-\alpha\tau}\, d\tau = \frac{A}{\alpha T}(1-e^{-\alpha T})e^{-\alpha t}, \qquad \text{time-varying} \tag{3.22}$$

$$Power(Ae^{-\alpha t}) = \frac{1}{T}\int_t^{t+T} A^2 e^{-2\alpha\tau}\, d\tau = \frac{A^2}{2\alpha T}(1-e^{-2\alpha T})e^{-2\alpha t}, \quad \text{time-varying} \tag{3.23}$$

In Equations (3.22) and (3.23), the signal mean and power are exponentially decaying functions of the time variable t.

Example 3.4 Consider a non-stationary signal $y(m)$ generated by a binary-state random process described by the following equation:

$$y(m) = \bar{s}(m)x_0(m) + s(m)x_1(m) \tag{3.24}$$

where $s(m)$ is a binary-valued state indicator variable and $\bar{s}(m)$ denotes the binary complement of $s(m)$. From Equation (3.24), we have

$$y(m) = \begin{cases} x_0(m) & \text{if } s(m) = 0 \\ x_1(m) & \text{if } s(m) = 1 \end{cases} \tag{3.25}$$

Let μ_{x_0} and P_{x_0} denote the mean and the power of the signal $x_0(m)$, and μ_{x_1} and P_{x_1} the mean and the power of $x_1(m)$ respectively. The expectation of $y(m)$, given the state $s(m)$, is obtained as

$$\begin{aligned}\mathcal{E}[y(m)|s(m)] &= \bar{s}(m)\mathcal{E}[x_0(m)] + s(m)\mathcal{E}[x_1(m)] \\ &= \bar{s}(m)\mu_{x_0} + s(m)\mu_{x_1}\end{aligned} \tag{3.26}$$

In Equation (3.26), the mean of $y(m)$ is expressed as a function of the state of the process at time m. The power of $y(m)$ is given by

$$\begin{aligned}\mathcal{E}\left[y^2(m)|s(m)\right] &= \bar{s}(m)\mathcal{E}\left[x_0^2(m)\right] + s(m)\mathcal{E}\left[x_1^2(m)\right] \\ &= \bar{s}(m)P_{x_0} + s(m)P_{x_1}\end{aligned} \tag{3.27}$$

Although many signals are non-stationary, the concept of a stationary process has played an important role in the development of signal processing methods. Furthermore, even non-stationary signals such as speech can often be considered as approximately stationary for a short period of time. In signal processing theory, two classes of stationary processes are defined: (a) strict-sense stationary processes and (b) wide-sense stationary processes, which is a less strict form of stationarity, in that it only requires that the first-order and second-order statistics of the process should be time-invariant.

3.3.1 Strict-Sense Stationary Processes

A random process $X(m)$ is stationary in a strict sense if all its distributions and statistical parameters are time-invariant. Strict-sense stationarity implies that the n^{th} order distribution is translation-invariant for all $n=1, 2,3, \ldots$:

$$\begin{aligned}&Prob[x(m_1)\leq x_1, x(m_2)\leq x_2, \ldots, x(m_n)\leq x_n)] \\ &\quad = Prob[x(m_1+\tau)\leq x_1, x(m_2+\tau)\leq x_2, \ldots, x(m_n+\tau)\leq x_n)]\end{aligned} \tag{3.28}$$

From Equation (3.28) the statistics of a strict-sense stationary process including the mean, the correlation and the power spectrum, are time-invariant; therefore we have

$$\mathcal{E}[x(m)] = \mu_x \tag{3.29}$$

$$\mathcal{E}[x(m)x(m+k)] = r_{xx}(k) \tag{3.30}$$

and

$$\mathcal{E}[|X(f,m)|^2] = \mathcal{E}[|X(f)|^2] = P_{XX}(f) \tag{3.31}$$

where μ_x, $r_{xx}(m)$ and $P_{XX}(f)$ are the mean value, the autocorrelation and the power spectrum of the signal $x(m)$ respectively, and $X(f,m)$ denotes the frequency–time spectrum of $x(m)$.

3.3.2 Wide-Sense Stationary Processes

The strict-sense stationarity condition requires that all statistics of the process should be time-invariant. A less restrictive form of a stationary process is so-called wide-sense stationarity. A process is said to be wide-sense stationary if the mean and the autocorrelation functions of the process are time invariant:

$$\mathcal{E}[x(m)] = \mu_x \tag{3.32}$$

$$\mathcal{E}[x(m)x(m+k)] = r_{xx}(k) \tag{3.33}$$

From the definitions of strict-sense and wide-sense stationary processes, it is clear that a strict-sense stationary process is also wide-sense stationary, whereas the reverse is not necessarily true.

3.3.3 Non-Stationary Processes

A random process is non-stationary if its distributions or statistics vary with time. Most stochastic processes such as video signals, audio signals, financial data, meteorological data, biomedical signals, etc., are non-stationary, because they are generated by systems whose environments and parameters vary over time. For example, speech is a non-stationary process generated by a time-varying articulatory system. The loudness and the frequency composition of speech changes over time, and sometimes the change can be quite abrupt. Time-varying processes may be modelled by a combination of stationary random models as illustrated in Figure 3.5. In Figure 3.5(a) a non-stationary process is modelled as the output of a time-varying system whose parameters are controlled by a stationary process. In Figure 3.5(b) a time-varying process is modelled by a chain of time-invariant states, with each state having a different set of statistics or

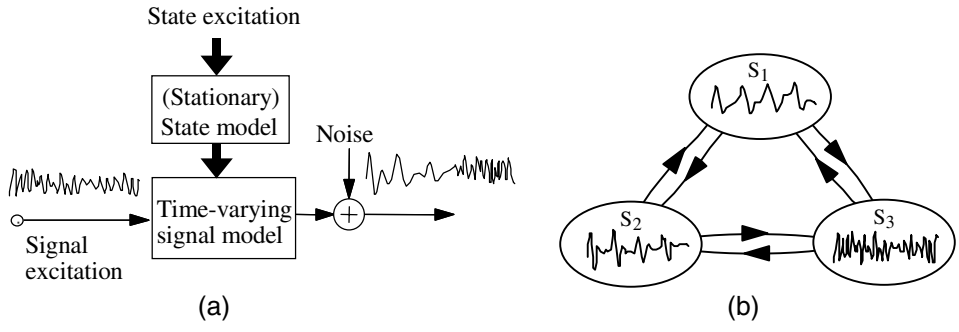


Figure 3.5 Two models for non-stationary processes: (a) a stationary process drives the parameters of a continuously time-varying model; (b) a finite-state model with each state having a different set of statistics.

probability distributions. Finite state statistical models for time-varying processes are discussed in detail in Chapter 5.

3.4 Expected Values of a Random Process

Expected values of a process play a central role in the modelling and processing of signals. Furthermore, the probability models of a random process are usually expressed as functions of the expected values. For example, a Gaussian pdf is defined as an exponential function of the mean and the covariance of the process, and a Poisson pdf is defined in terms of the mean of the process. In signal processing applications, we often have a suitable statistical model of the process, e.g. a Gaussian pdf, and to complete the model we need the values of the expected parameters. Furthermore in many signal processing algorithms, such as spectral subtraction for noise reduction described in Chapter 11, or linear prediction described in Chapter 8, what we essentially need is an estimate of the mean or the correlation function of the process. The expected value of a function, $h(X(m_1), X(m_2), ..., X(m_M))$, of a random process X is defined as

$$\mathcal{E}[h(X(m_1),\ldots,X(m_M))] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h(x_1,\ldots,x_M) f_{X(m_1)\cdots X(m_M)}(x_1,\ldots,x_M)\, dx_1 \ldots dx_M \tag{3.34}$$

The most important, and widely used, expected values are the mean value, the correlation, the covariance, and the power spectrum.

3.4.1 The Mean Value

The mean value of a process plays an important part in signal processing and parameter estimation from noisy observations. For example, in Chapter 3 it is shown that the optimal linear estimate of a signal from a noisy observation, is an interpolation between the mean value and the observed value of the noisy signal. The mean value of a random vector $[X(m_1), ..., X(m_M)]$ is its average value across the ensemble of the process defined as

$$\mathcal{E}[X(m_1),\ldots,X(m_M)] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (x_1,\ldots,x_M) f_{X(m_1)\cdots X(m_M)}(x_1,\ldots,x_M)\, dx_1 \cdots dx_M \tag{3.35}$$

3.4.2 Autocorrelation

The correlation function and its Fourier transform, the power spectral density, are used in modelling and identification of patterns and structures in a signal process. Correlators play a central role in signal processing and telecommunication systems, including predictive coders, equalisers, digital decoders, delay estimators, classifiers and signal restoration systems. The autocorrelation function of a random process $X(m)$, denoted by $r_{xx}(m_1,m_2)$, is defined as

$$\begin{aligned} r_{xx}(m_1,m_2) &= \mathcal{E}[x(m_1)x(m_2)] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x(m_1)x(m_2) f_{X(m_1),X(m_1)}\big(x(m_1),x(m_2)\big)\, dx(m_1)\, dx(m_2) \end{aligned} \tag{3.36}$$

The autocorrelation function $r_{xx}(m_1,m_2)$ is a measure of the similarity, or the mutual relation, of the outcomes of the process X at time instants m_1 and m_2. If the outcome of a random process at time m_1 bears no relation to that at time m_2 then $X(m_1)$ and $X(m_2)$ are said to be independent or uncorrelated and $r_{xx}(m_1,m_2)=0$. For a wide-sense stationary process, the autocorrelation function is time-invariant and depends on the time difference $m= m_1–m_2$:

$$r_{xx}(m_1+\tau, m_2+\tau) = r_{xx}(m_1,m_2) = r_{xx}(m_1-m_2) = r_{xx}(m) \tag{3.37}$$

The autocorrelation function of a real-valued wide-sense stationary process is a symmetric function with the following properties:

$$r_{xx}(-m) = r_{xx}(m) \tag{3.38}$$

$$r_{xx}(m) \leq r_{xx}(0) \tag{3.39}$$

Note that for a zero-mean signal, $r_{xx}(0)$ is the signal power.

Example 3.5 Autocorrelation of the output of a linear time-invariant (LTI) system. Let $x(m)$, $y(m)$ and $h(m)$ denote the input, the output and the impulse response of a LTI system respectively. The input–output relation is given by

$$y(m)=\sum_k h_k x(m-k) \tag{3.40}$$

The autocorrelation function of the output signal $y(m)$ can be related to the autocorrelation of the input signal $x(m)$ by

$$\begin{aligned} r_{yy}(k) &= \mathcal{E}[y(m)y(m+k)] \\ &= \sum_i \sum_j h_i h_j \mathcal{E}[x(m-i)x(m+k-j)] \\ &= \sum_i \sum_j h_i h_j r_{xx}(k+i-j) \end{aligned} \tag{3.41}$$

When the input $x(m)$ is an uncorrelated random signal with a unit variance, Equation (3.41) becomes

$$r_{yy}(k)=\sum_i h_i h_{k+i} \tag{3.42}$$

3.4.3 Autocovariance

The autocovariance function $c_{xx}(m_1,m_2)$ of a random process $X(m)$ is measure of the scatter, or the dispersion, of the random process about the mean value, and is defined as

$$\begin{aligned} c_{xx}(m_1,m_2) &= \mathcal{E}[(x(m_1)-\mu_x(m_1))(x(m_2)-\mu_x(m_2))] \\ &= r_{xx}(m_1,m_2) - \mu_x(m_1)\mu_x(m_2) \end{aligned} \tag{3.43}$$

where $\mu_x(m)$ is the mean of $X(m)$. Note that for a zero-mean process the autocorrelation and the autocovariance functions are identical. Note also that $c_{xx}(m_1,m_1)$ is the variance of the process. For a stationary process the autocovariance function of Equation (3.43) becomes

$$c_{xx}(m_1,m_2) = c_{xx}(m_1-m_2) = r_{xx}(m_1-m_2) - \mu_x^2 \tag{3.44}$$

3.4.4 Power Spectral Density

The power spectral density (PSD) function, also called the power spectrum, of a random process gives the spectrum of the distribution of the power among the individual frequency contents of the process. The power spectrum of a wide sense stationary process $X(m)$ is defined, by the Wiener–Khinchin theorem in Chapter 9, as the Fourier transform of the autocorrelation function:

$$\begin{aligned} P_{XX}(f) &= \mathcal{E}[X(f)X^*(f)] \\ &= \sum_{m=-\infty}^{\infty} r_{xx}(k)e^{-j2\pi fm} \end{aligned} \tag{3.45}$$

where $r_{xx}(m)$ and $P_{XX}(f)$ are the autocorrelation and power spectrum of $x(m)$ respectively, and f is the frequency variable. For a real-valued stationary process, the autocorrelation is symmetric, and the power spectrum may be written as

$$P_{XX}(f) = r_{xx}(0) + \sum_{m=1}^{\infty} 2r_{xx}(m)\cos(2\pi fm) \tag{3.46}$$

The power spectral density is a real-valued non-negative function, expressed in units of watts per hertz. From Equation (3.45), the autocorrelation sequence of a random process may be obtained as the inverse Fourier transform of the power spectrum as

$$r_{xx}(m) = \int_{-1/2}^{1/2} P_{XX}(f)\, e^{j2\pi fm}\, df \tag{3.47}$$

Note that the autocorrelation and the power spectrum represent the second order statistics of a process in the time and frequency domains respectively.

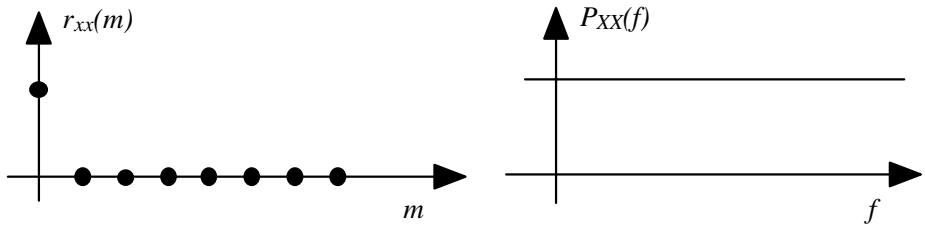


Figure 3.6 Autocorrelation and power spectrum of white noise.

Example 3.6 Power spectrum and autocorrelation of white noise (Figure3.6). A noise process with uncorrelated independent samples is called a white noise process. The autocorrelation of a stationary white noise $n(m)$ is defined as:

$$r_{nn}(k) = \mathcal{E}[n(m)n(m+k)] = \begin{cases} \text{Noisepower} & k = 0 \\ 0 & k \neq 0 \end{cases} \tag{3.48}$$

Equation (3.48) is a mathematical statement of the definition of an uncorrelated white noise process. The equivalent description in the frequency domain is derived by taking the Fourier transform of $r_{nn}(k)$:

$$P_{NN}(f) = \sum_{k=-\infty}^{\infty} r_{nn}(k)e^{-j2\pi fk} = r_{nn}(0) = \text{noise power} \tag{3.49}$$

The power spectrum of a stationary white noise process is spread equally across all time instances and across all frequency bins. White noise is one of the most difficult types of noise to remove, because it does not have a localised structure either in the time domain or in the frequency domain.

Example 3.7 Autocorrelation and power spectrum of impulsive noise. Impulsive noise is a random, binary-state ("on/off") sequence of impulses of random amplitudes and random time of occurrence. In Chapter 12, a random impulsive noise sequence $n_i(m)$ is modelled as an amplitude-modulated random binary sequence as

$$n_i(m) = n(m)b(m) \tag{3.50}$$

where $b(m)$ is a binary-state random sequence that indicates the presence or the absence of an impulse, and $n(m)$ is a random noise process. Assuming

that impulsive noise is an uncorrelated process, the autocorrelation of impulsive noise can be defined as a binary-state process as

$$r_{nn}(k,m) = \mathcal{E}[n_i(m)n_i(m+k)] = \sigma_n^2\,\delta(k)b(m) \tag{3.51}$$

where σ_n^2 is the noise variance. Note that in Equation (3.51), the autocorrelation is expressed as a binary-state function that depends on the on/off state of impulsive noise at time *m*. The power spectrum of an impulsive noise sequence is obtained by taking the Fourier transform of the autocorrelation function:

$$P_{NN}(f,m) = \sigma_n^2\, b(m) \tag{3.52}$$

3.4.5 Joint Statistical Averages of Two Random Processes

In many signal processing problems, for example in processing the outputs of an array of sensors, we deal with more than one random process. Joint statistics and joint distributions are used to describe the statistical inter-relationship between two or more random processes. For two discrete-time random processes $x(m)$ and $y(m)$, the joint pdf is denoted by

$$f_{X(m_1)\cdots X(m_M),Y(n_1)\cdots Y(n_N)}(x_1,\ldots,x_M,y_1,\ldots,y_N) \tag{3.53}$$

When two random processes, $X(m)$ and $Y(m)$ are uncorrelated, the joint pdf can be expressed as product of the pdfs of each process as

$$\begin{aligned} &f_{X(m_1)\cdots X(m_M),Y(n_1)\cdots Y(n_N)}(x_1,\ldots,x_M,y_1,\ldots,y_N) \\ &\quad = f_{X(m_1)\cdots X(m_M)}(x_1,\ldots,x_M)\, f_{Y(n_1)\cdots Y(n_N)}(y_1,\ldots,y_N) \end{aligned} \tag{3.54}$$

3.4.6 Cross-Correlation and Cross-Covariance

The cross-correlation of two random process $x(m)$ and $y(m)$ is defined as

$$\begin{aligned} r_{xy}(m_1,m_2) &= \mathcal{E}[x(m_1)y(m_2)] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x(m_1)y(m_2)f_{X(m_1)Y(m_2)}(x(m_1),y(m_2))\,dx(m_1)\,dy(m_2) \end{aligned} \tag{3.55}$$

For wide-sense stationary processes, the cross-correlation function $r_{xy}(m_1,m_2)$ depends only on the time difference $m=m_1-m_2$:

$$r_{xy}(m_1+\tau,m_2+\tau)=r_{xy}(m_1,m_2)=r_{xy}(m_1-m_2)=r_{xy}(m) \tag{3.56}$$

The cross-covariance function is defined as

$$\begin{aligned} c_{xy}(m_1,m_2) &= \mathcal{E}\left[\left(x(m_1)-\mu_x(m_1)\right)\left(y(m_2)-\mu_y(m_2)\right)\right] \\ &= r_{xy}(m_1,m_2)-\mu_x(m_1)\mu_y(m_2) \end{aligned} \tag{3.57}$$

Note that for zero-mean processes, the cross-correlation and the cross-covariance functions are identical. For a wide-sense stationary process the cross-covariance function of Equation (3.57) becomes

$$c_{xy}(m_1,m_2)=c_{xy}(m_1-m_2)=r_{xy}(m_1-m_2)-\mu_x\mu_y \tag{3.58}$$

Example 3.8 Time-delay estimation. Consider two signals $y_1(m)$ and $y_2(m)$, each composed of an information bearing signal $x(m)$ and an additive noise, given by

$$y_1(m)=x(m)+n_1(m) \tag{3.59}$$

$$y_2(m)=A\,x(m-D)+n_2(m) \tag{3.60}$$

where A is an amplitude factor and D is a time delay variable. The cross-correlation of the signals $y_1(m)$ and $y_2(m)$ yields

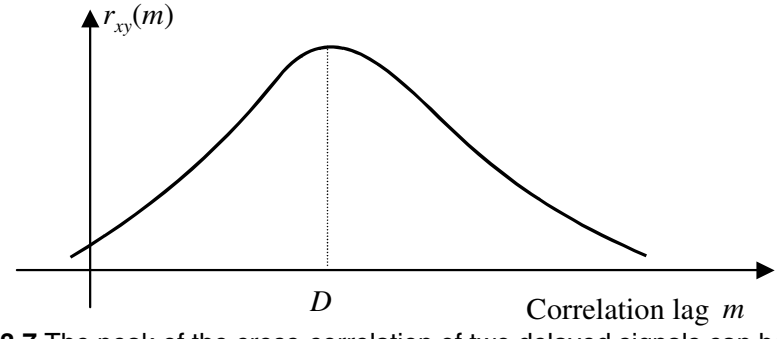


Figure 3.7 The peak of the cross-correlation of two delayed signals can be used to estimate the time delay D.

$$\begin{aligned} r_{y_1y_2}(k) &= \mathcal{E}[y_1(m)y_2(m+k)] \\ &= \mathcal{E}\{[x(m)+n_1(m)][Ax(m-D+k)+n_2(m+k)]\} \\ &= Ar_{xx}(k-D)+r_{xn_2}(k)+Ar_{xn_1}(k-D)+r_{n_1n_2}(k) \end{aligned} \tag{3.61}$$

Assuming that the signal and noise are uncorrelated, we have $r_{y_1y_2}(k) = Ar_{xx}(k-D)$. As shown in Figure 3.7, the cross-correlation function has its maximum at the lag D.

3.4.7 Cross-Power Spectral Density and Coherence

The cross-power spectral density of two random processes $X(m)$ and $Y(m)$ is defined as the Fourier transform of their cross-correlation function:

$$\begin{aligned} P_{XY}(f) &= \mathcal{E}[X(f)Y^*(f)] \\ &= \sum_{m=-\infty}^{\infty} r_{xy}(m)e^{-j2\pi fm} \end{aligned} \tag{3.62}$$

Like the cross-correlation the cross-power spectral density of two processes is a measure of the similarity, or coherence, of their power spectra. The coherence, or spectral coherence, of two random processes is a normalised form of the cross-power spectral density, defined as

$$C_{XY}(f) = \frac{P_{XY}(f)}{\sqrt{P_{XX}(f)P_{YY}(f)}} \tag{3.63}$$

The coherence function is used in applications such as time-delay estimation and signal-to-noise ratio measurements.

3.4.8 Ergodic Processes and Time-Averaged Statistics

In many signal processing problems, there is only a single realisation of a random process from which its statistical parameters, such as the mean, the correlation and the power spectrum can be estimated. In such cases, time-averaged statistics, obtained from averages along the time dimension of a single realisation of the process, are used instead of the "true" ensemble averages obtained across the space of different realisations of the process.

This section considers ergodic random processes for which time-averages can be used instead of ensemble averages. *A stationary stochastic process is said to be ergodic if it exhibits the same statistical characteristics along the time dimension of a single realisation as across the space* (*or ensemble*) *of different realisations of the process*. Over a very long time, a single realisation of an ergodic process takes on all the values, the characteristics and the configurations exhibited across the entire space of the process. For an ergodic process $\{x(m,s)\}$, we have

$$\underset{\text{along time } m}{\textit{statistical averages}}[x(m,s)] = \underset{\text{across space } s}{\textit{statistical averages}}[x(m,s)] \tag{3.64}$$

where the *statistical averages*[.] function refers to any statistical operation such as the mean, the variance, the power spectrum, etc.

3.4.9 Mean-Ergodic Processes

The time-averaged estimate of the mean of a signal $x(m)$ obtained from N samples is given by

$$\hat{\mu}_X = \frac{1}{N}\sum_{m=0}^{N-1} x(m) \tag{3.65}$$

A stationary process is said to be mean-ergodic if the time-averaged value of an infinitely long realisation of the process is the same as the ensemble-mean taken across the space of the process. Therefore, for a mean-ergodic process, we have

$$\lim_{N\to\infty} \mathcal{E}[\hat{\mu}_X] = \mu_X \tag{3.66}$$

$$\lim_{N\to\infty} \text{var}[\hat{\mu}_X] = 0 \tag{3.67}$$

where μ_x is the "true" ensemble average of the process. Condition (3.67) is also referred to as mean-ergodicity in the mean square error (or minimum variance of error) sense. The time-averaged estimate of the mean of a signal, obtained from a random realisation of the process, is itself a random variable, with is own mean, variance and probability density function. If the number of observation samples N is relatively large then, from the central limit theorem the probability density function of the estimate $\hat{\mu}_X$ is Gaussian. The expectation of $\hat{\mu}_X$ is given by

$$\mathcal{E}[\hat{\mu}_x]=\mathcal{E}\left[\frac{1}{N}\sum_{m=0}^{N-1}x(m)\right]=\frac{1}{N}\sum_{m=0}^{N-1}\mathcal{E}[x(m)]=\frac{1}{N}\sum_{m=0}^{N-1}\mu_x=\mu_x \tag{3.68}$$

From Equation (3.68), the time-averaged estimate of the mean is unbiased. The variance of $\hat{\mu}_X$ is given by

$$\begin{aligned}\text{Var}[\hat{\mu}_x]&=\mathcal{E}[\hat{\mu}_x^2]-\mathcal{E}^2[\hat{\mu}_x]\\&=\mathcal{E}[\hat{\mu}_x^2]-\mu_x^2\end{aligned} \tag{3.69}$$

Now the term $\mathcal{E}[\hat{\mu}_x^2]$ in Equation (3.69) may be expressed as

$$\begin{aligned}\mathcal{E}[\hat{\mu}_x^2]&=\mathcal{E}\left[\left(\frac{1}{N}\sum_{m=0}^{N-1}x(m)\right)\left(\frac{1}{N}\sum_{k=0}^{N-1}x(k)\right)\right]\\&=\frac{1}{N}\sum_{m=-(N-1)}^{N-1}\left(1-\frac{|m|}{N}\right)r_{xx}(m)\end{aligned} \tag{3.70}$$

Substitution of Equation (3.70) in Equation (3.69) yields

$$\begin{aligned}\text{Var}[\hat{\mu}_x^2]&=\frac{1}{N}\sum_{m=-(N-1)}^{N-1}\left(1-\frac{|m|}{N}\right)r_{xx}(m)-\mu_x^2\\&=\frac{1}{N}\sum_{m=-(N-1)}^{N-1}\left(1-\frac{|m|}{N}\right)c_{xx}(m)\end{aligned} \tag{3.71}$$

Therefore the condition for a process to be mean-ergodic, in the mean square error sense, is

$$\lim_{N\to\infty}\frac{1}{N}\sum_{m=-(N-1)}^{N-1}\left(1-\frac{|m|}{N}\right)c_{xx}(m)=0 \tag{3.72}$$

3.4.10 Correlation-Ergodic Processes

The time-averaged estimate of the autocorrelation of a random process, estimated from N samples of a realisation of the process, is given by

$$\hat{r}_{xx}(m) = \frac{1}{N}\sum_{k=0}^{N-1} x(k)x(k+m) \tag{3.73}$$

A process is correlation-ergodic, in the mean square error sense, if

$$\lim_{N\to\infty} \mathcal{E}[\hat{r}_{xx}(m)] = r_{xx}(m) \tag{3.74}$$

$$\lim_{N\to\infty} \mathrm{Var}[\hat{r}_{xx}(m)] = 0 \tag{3.75}$$

where $r_{xx}(m)$ is the ensemble-averaged autocorrelation. Taking the expectation of $\hat{r}_{xx}(m)$ shows that it is an unbiased estimate, since

$$\mathcal{E}[\hat{r}_{xx}(m)] = \mathcal{E}\left[\frac{1}{N}\sum_{k=0}^{N-1} x(k)x(k+m)\right] = \frac{1}{N}\sum_{k=0}^{N-1} \mathcal{E}[x(k)x(k+m)] = r_{xx}(m) \tag{3.76}$$

The variance of $\hat{r}_{xx}(m)$ is given by

$$\mathrm{Var}[\hat{r}_{xx}(m)] = \mathcal{E}[\hat{r}_{xx}^2(m)] - r_{xx}^2(m) \tag{3.77}$$

The term $\mathcal{E}[\hat{r}_{xx}^2(m)]$ in Equation (3.77) may be expressed as

$$\begin{aligned}\mathcal{E}[\hat{r}_{xx}^2(m)] &= \frac{1}{N^2}\sum_{k=0}^{N-1}\sum_{j=0}^{N-1} \mathcal{E}[x(k)x(k+m)x(j)x(j+m)] \\ &= \frac{1}{N^2}\sum_{k=0}^{N-1}\sum_{j=0}^{N-1} \mathcal{E}[z(k,m)z(j,m)] \\ &= \frac{1}{N}\sum_{k=-N+1}^{N-1}\left(1-\frac{|k|}{N}\right) r_{zz}(k,m)\end{aligned} \tag{3.78}$$

where $z(i,m)=x(i)x(i+m)$. Therefore the condition for correlation ergodicity in the mean square error sense is given by

$$\lim_{N\to\infty}\left[\frac{1}{N}\sum_{k=-N+1}^{N-1}\left(1-\frac{|k|}{N}\right) r_{zz}(k,m) - r_{xx}^2(m)\right] = 0 \tag{3.79}$$

3.5 Some Useful Classes of Random Processes

In this section, we consider some important classes of random processes extensively used in signal processing applications for the modelling of signals and noise.

3.5.1 Gaussian (Normal) Process

The Gaussian process, also called the normal process, is perhaps the most widely applied of all probability models. Some advantages of Gaussian probability models are the following:

(a) Gaussian pdfs can model the distribution of many processes including some important classes of signals and noise.
(b) Non-Gaussian processes can be approximated by a weighted combination (i.e. a mixture) of a number of Gaussian pdfs of appropriate means and variances.
(c) Optimal estimation methods based on Gaussian models often result in linear and mathematically tractable solutions.
(d) The sum of many independent random processes has a Gaussian distribution. This is known as the central limit theorem.

A scalar Gaussian random variable is described by the following probability density function:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left[-\frac{(x-\mu_x)^2}{2\sigma_x^2}\right] \tag{3.80}$$

where μ_x and σ_x^2 are the mean and the variance of the random variable x. The Gaussian process of Equation (3.80) is also denoted by $\mathcal{N}(x, \mu_x, \sigma_x^2)$. The maximum of a Gaussian pdf occurs at the mean μ_x, and is given by

$$f_X(\mu_x) = \frac{1}{\sqrt{2\pi}\,\sigma_x} \tag{3.81}$$

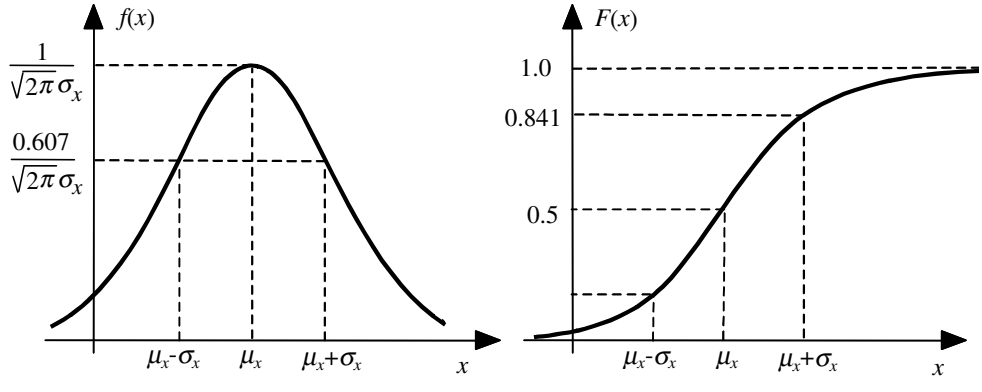


Figure 3.8 Gaussian probability density and cumulative density functions.

From Equation (3.80), the Gaussian pdf of x decreases exponentially with the increasing distance of x from the mean value μ_x. The distribution function $F(x)$ is given by

$$F_X(x) = \frac{1}{\sqrt{2\pi}\sigma_x} \int_{-\infty}^{x} \exp\left(-\frac{(\chi - \mu_x)^2}{2\sigma_x^2}\right) d\chi \tag{3.82}$$

Figure 3.8 shows the pdf and the cdf of a Gaussian model.

3.5.2 Multivariate Gaussian Process

Multivariate densities model vector-valued processes. Consider a P-variate Gaussian vector process $\{\boldsymbol{x}=[x(m_0), x(m_1), \ldots, x(m_{P-1})]^{\mathrm{T}}\}$ with mean vector $\boldsymbol{\mu}_x$, and covariance matrix $\boldsymbol{\Sigma}_{xx}$. The multivariate Gaussian pdf of $\boldsymbol{x}$ is given by

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{(2\pi)^{P/2}|\boldsymbol{\Sigma}_{\boldsymbol{xx}}|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_{\boldsymbol{x}})^{\mathrm{T}} \boldsymbol{\Sigma}_{\boldsymbol{xx}}^{-1} (\boldsymbol{x}-\boldsymbol{\mu}_{\boldsymbol{x}})\right] \tag{3.83}$$

where the mean vector $\boldsymbol{\mu}_{\boldsymbol{x}}$ is defined as

$$\boldsymbol{\mu}_{\boldsymbol{x}} = \begin{pmatrix} \mathcal{E}[x(m_0)] \\ \mathcal{E}[x(m_2)] \\ \vdots \\ \mathcal{E}[x(m_{P-1})] \end{pmatrix} \tag{3.84}$$

and the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{xx}}$ is given by

$$\boldsymbol{\Sigma}_{\boldsymbol{xx}} = \begin{pmatrix} c_{xx}(m_0, m_0) & c_{xx}(m_0, m_1) & \dots & c_{xx}(m_0, m_{P-1}) \\ c_{xx}(m_1, m_0) & c_{xx}(m_1, m_1) & \dots & c_{xx}(m_1, m_{P-1}) \\ \vdots & \vdots & \ddots & \vdots \\ c_{xx}(m_{P-1}, m_0) & c_{xx}(m_{P-1}, m_1) & \dots & c_{xx}(m_{P-1}, m_{P-1}) \end{pmatrix} \tag{3.85}$$

The Gaussian process of Equation (3.83) is also denoted by $\mathcal{N}(\boldsymbol{x}, \boldsymbol{\mu}_{\boldsymbol{x}}, \boldsymbol{\Sigma}_{\boldsymbol{xx}})$. If the elements of a vector process are uncorrelated then the covariance matrix is a diagonal matrix with zeros in the off-diagonal elements. In this case the multivariate pdf may be described as the product of the pdfs of the individual elements of the vector:

$$f_{\boldsymbol{X}}\left(\boldsymbol{x} = [x(m_0), \dots, x(m_{P-1})]^{\mathrm{T}}\right) = \prod_{i=0}^{P-1} \frac{1}{\sqrt{2\pi}\sigma_{xi}} \exp\left\{-\frac{[x(m_i) - \mu_{xi}]^2}{2\sigma_{xi}^2}\right\} \tag{3.86}$$

Example 3.9 Conditional multivariate Gaussian probability density function. Consider two vector realisations $\boldsymbol{x}(m)$ and $\boldsymbol{y}(m+k)$ from two vector-valued correlated stationary Gaussian processes $\mathcal{N}(\boldsymbol{x}, \boldsymbol{\mu}_{\boldsymbol{x}}, \boldsymbol{\Sigma}_{\boldsymbol{xx}})$ and $\mathcal{N}(\boldsymbol{y}, \boldsymbol{\mu}_{\boldsymbol{y}}, \boldsymbol{\Sigma}_{\boldsymbol{yy}})$. The joint probability density function of $\boldsymbol{x}(m)$ and $\boldsymbol{y}(m+k)$ is a multivariate Gaussian density $\mathcal{N}([\boldsymbol{x}(m), \boldsymbol{y}(m+k)], \boldsymbol{\mu}_{(\boldsymbol{x},\boldsymbol{y})}, \boldsymbol{\Sigma}_{(\boldsymbol{x},\boldsymbol{y})})$, with mean vector and covariance matrix given by

$$\boldsymbol{\mu}_{(\boldsymbol{x},\boldsymbol{y})} = \begin{bmatrix} \boldsymbol{\mu}_{\boldsymbol{x}} \\ \boldsymbol{\mu}_{\boldsymbol{y}} \end{bmatrix} \tag{3.87}$$

$$\boldsymbol{\Sigma}_{(\boldsymbol{x},\boldsymbol{y})} = \begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{xx}} & \boldsymbol{\Sigma}_{\boldsymbol{xy}} \\ \boldsymbol{\Sigma}_{\boldsymbol{yx}} & \boldsymbol{\Sigma}_{\boldsymbol{yy}} \end{bmatrix} \tag{3.88}$$

The conditional density of $\boldsymbol{x}(m)$ given $\boldsymbol{y}(m+k)$ is given from Bayes' rule as

$$f_{\boldsymbol{X}|\boldsymbol{Y}}\left(\boldsymbol{x}(m)\middle|\boldsymbol{y}(m+k)\right)=\frac{f_{\boldsymbol{X},\boldsymbol{Y}}\left(\boldsymbol{x}(m),\boldsymbol{y}(m+k)\right)}{f_{\boldsymbol{Y}}\left(\boldsymbol{y}(m+k)\right)} \tag{3.89}$$

It can be shown that the conditional density is also a multivariate Gaussian with its mean vector and covariance matrix given by

$$\begin{aligned}\boldsymbol{\mu}_{(\boldsymbol{x}|\boldsymbol{y})}&=\mathcal{E}\left[\boldsymbol{x}(m)\middle|\boldsymbol{y}(m+k)\right]\\&=\boldsymbol{\mu}_{\boldsymbol{x}}+\boldsymbol{\Sigma}_{\boldsymbol{xy}}\boldsymbol{\Sigma}_{\boldsymbol{yy}}^{-1}(\boldsymbol{y}-\boldsymbol{\mu}_{\boldsymbol{y}})\end{aligned} \tag{3.90}$$

$$\boldsymbol{\Sigma}_{(\boldsymbol{x}|\boldsymbol{y})}=\boldsymbol{\Sigma}_{\boldsymbol{xx}}-\boldsymbol{\Sigma}_{\boldsymbol{xy}}\boldsymbol{\Sigma}_{\boldsymbol{yy}}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{yx}} \tag{3.91}$$

3.5.3 Mixture Gaussian Process

Probability density functions of many processes, such as speech, are non-Gaussian. A non-Gaussian pdf may be approximated by a weighted sum (i.e. a mixture) of a number of Gaussian densities of appropriate mean vectors and covariance matrices. An M-mixture Gaussian density is defined as

$$f_{\boldsymbol{X}}(\boldsymbol{x})\ =\sum_{i=1}^{M}P_i\,\mathcal{N}_i(\boldsymbol{x},\boldsymbol{\mu}_{\boldsymbol{x}i},\boldsymbol{\Sigma}_{\boldsymbol{xx}i}) \tag{3.92}$$

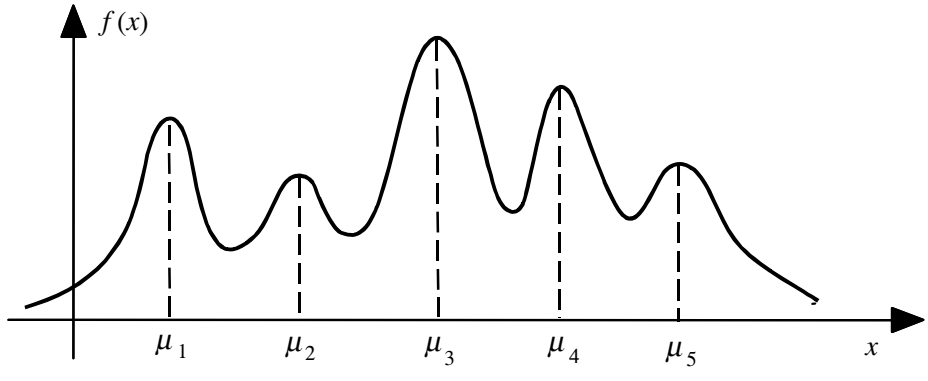


Figure 3.9 A mixture Gaussian pdf.

where $\mathcal{N}_i(\boldsymbol{x}, \boldsymbol{\mu}_{\boldsymbol{x}i}, \boldsymbol{\Sigma}_{\boldsymbol{xx}i})$ is a multivariate Gaussian density with mean vector $\boldsymbol{\mu}_{\boldsymbol{x}i}$ and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{xx}i}$, and P_i are the mixing coefficients. The parameter P_i is the prior probability of the i^{th} mixture component, and is given by

$$P_i = \frac{N_i}{\sum_{j=1}^{M} N_j} \tag{3.93}$$

where N_i is the number of observations associated with the mixture i. Figure 3.9 shows a non-Gaussian pdf modelled as a mixture of five Gaussian pdfs. Algorithms developed for Gaussian processes can be extended to mixture Gaussian densities.

3.5.4 A Binary-State Gaussian Process

Consider a random process $x(m)$ with two statistical states: such that in the state s_0 the process has a Gaussian pdf with mean $\mu_{x,0}$ and variance $\sigma^2_{x,0}$, and in the state s_1 the process is also Gaussian with mean $\mu_{x,1}$ and variance $\sigma^2_{x,1}$ (Figure 3.10). The state-dependent pdf of $x(m)$ can be expressed as

$$f_{X|S}(x(m)|s_i) = \frac{1}{\sqrt{2\pi}\sigma_{x,i}} \exp\left\{-\frac{1}{2\sigma^2_{x,i}}[x(m) - \mu_{x,i}]^2\right\}, \quad i=0, 1 \tag{3.94}$$

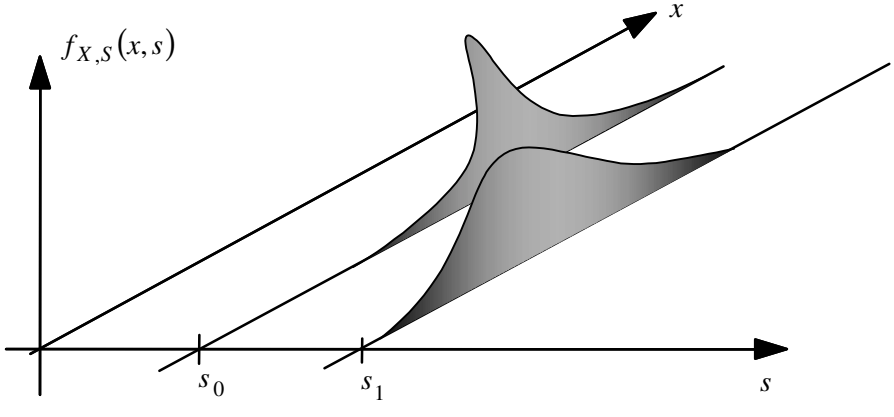


Figure 3.10 Illustration of a binary-state Gaussian process

The joint probability distribution of the binary-valued state s_i and the continuous-valued signal $x(m)$ can be expressed as

$$\begin{aligned} f_{X,S}(x(m), s_i) &= f_{X|S}(x(m)|s_i) P_S(s_i) \\ &= \frac{1}{\sqrt{2\pi}\sigma_{x,i}} \exp\left\{ -\frac{1}{2\sigma_{x,i}^2} [x(m) - \mu_{x,i}]^2 \right\} P_S(s_i) \end{aligned} \tag{3.95}$$

where $P_S(s_i)$ is the state probability. For a multistate process we have the following probabilistic relations between the joint and marginal probabilities:

$$\sum_S f_{X,S}(x(m), s_i) = f_X(x(m)) \tag{3.96}$$

$$\int_X f_{X,S}(x(m), s_i)\, dx = P_S(s_i) \tag{3.97}$$

and

$$\sum_S \int_X f_{X,S}(x(m), s_i)\, dx = 1 \tag{3.98}$$

Note that in a multistate model, the statistical parameters of the process *switch* between a number of different states, whereas in a single-state mixture pdf, a *weighted* combination of a number of pdfs models the process. In Chapter 5 on hidden Markov models we consider multistate models with a mixture pdf per state.

3.5.5 Poisson Process

The Poisson process is a continuous-time, integer-valued counting process, used for modelling the occurrence of a random event in various time intervals. An important area of application of the Poisson process is in queuing theory for the analysis and modelling of the distributions of demand on a service facility such as a telephone network, a shared computer system, a financial service, a petrol station, etc. Other applications of the Poisson distribution include the counting of the number of particles emitted in physics, the number of times that a component may fail in a system, and modelling of radar clutter, shot noise and impulsive noise. Consider an event-counting process $X(t)$, in which the probability of occurrence of the

event is governed by a rate function $\lambda(t)$, such that the probability that an event occurs in a small time interval Δt is

$$Prob(1 \text{ occurrence in the interval } (t, t+\Delta t)) = \lambda(t)\Delta t \tag{3.99}$$

Assuming that in the small interval Δt, no more than one occurrence of the event is possible, the probability of no occurrence of the event in a time interval of Δt is given by

$$Prob(0 \text{ occurrence in the interval } (t, t+\Delta t)) = 1 - \lambda(t)\Delta t \tag{3.100}$$

when the parameter $\lambda(t)$ is independent of time, $\lambda(t)=\lambda$, and the process is called a homogeneous Poisson process. Now, for a homogeneous Poisson process, consider the probability of k occurrences of an event in a time interval of $t+\Delta t$, denoted by $P(k, (0, t+\Delta t))$:

$$\begin{aligned} P(k,(0,t+\Delta t)) &= P(k,(0,t))P(0,(t,t+\Delta t)) + P(k-1,(0,t))P(1,(t,t+\Delta t)) \\ &= P(k,(0,t))(1-\lambda\Delta t) + P(k-1,(0,t))\lambda\Delta t \end{aligned} \tag{3.101}$$

Rearranging Equation (3.101), and letting Δt tend to zero, we obtain the following linear differential equation:

$$\frac{dP(k,t)}{dt} = -\lambda P(k,t) + \lambda P(k-1,t) \tag{3.102}$$

where $P(k,t)=P(k,(0, t))$. The solution of this differential equation is given by

$$P(k,t) = \lambda e^{-\lambda t} \int_0^t P(k-1,\tau) e^{\lambda\tau}\, d\tau \tag{3.103}$$

Equation (3.103) can be solved recursively: starting with $P(0,t)=e^{-\lambda t}$ and $P(1,t)=\lambda t\, e^{-\lambda t}$, we obtain the Poisson density

$$P(k,t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t} \tag{3.104}$$

From Equation (3.104), it is easy to show that for a homogenous Poisson process, the probability of k occurrences of an event in a time interval (t_1, t_2) is given by

$$P[k,(t_1,t_2)] = \frac{[\lambda(t_2 - t_1)]^k}{k!} e^{-\lambda(t_2-t_1)} \tag{3.105}$$

A Poisson counting process $X(t)$ is incremented by one every time the event occurs. From Equation (3.104), the mean and variance of a Poisson counting process $X(t)$ are

$$\mathcal{E}[X(t)] = \lambda t \tag{3.106}$$

$$r_{XX}(t_1,t_2) = \mathcal{E}[X(t_1)X(t_2)] = \lambda^2 t_1 t_2 + \lambda \min(t_1,t_2) \tag{3.107}$$

$$\text{Var}[X(t)] = \mathcal{E}\left[X^2(t)\right] - \mathcal{E}^2[X(t)] = \lambda t \tag{3.108}$$

Note that the variance of a Poisson process is equal to its mean value.

3.5.6 Shot Noise

Shot noise happens when there is randomness in a directional flow of particles: as in the flow of electrons from the cathode to the anode of a cathode ray tube, the flow of photons in a laser beam, the flow and recombination of electrons and holes in semiconductors, and the flow of photoelectrons emitted in photodiodes. Shot noise has the form of a random pulse sequence. The pulse sequence can be modelled as the response of a linear filter excited by a Poisson-distributed binary impulse input sequence (Figure 3.11). Consider a Poisson-distributed binary-valued impulse process $x(t)$. Divide the time axis into uniform short intervals of Δt such that only one occurrence of an impulse is possible within each time interval. Let $x(m\Delta t)$ be "1" if an impulse is present in the interval $m\Delta t$ to $(m+1)\Delta t$, and "0" otherwise. For $x(m\Delta t)$, we have

$$\mathcal{E}[x(m\Delta t)] = 1 \times P(x(m\Delta t) = 1) + 0 \times P(x(m\Delta t) = 0) = \lambda \Delta t \tag{3.109}$$

and

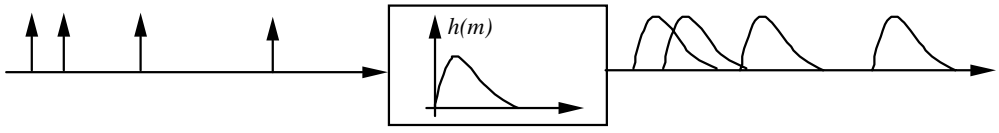


Figure 3.11 Shot noise is modelled as the output of a filter excited with a process.

$$\mathcal{E}[x(m\Delta t)x(n\Delta t)] = \begin{cases} 1\times P(x(m\Delta t)=1) = \lambda\Delta t, & m = n \\ 1\times P(x(m\Delta t)=1))\times P(x(n\Delta t)=1) = (\lambda\Delta t)^2, & m \neq n \end{cases} \tag{3.110}$$

A shot noise process $y(m)$ is defined as the output of a linear system with an impulse response $h(t)$, excited by a Poisson-distributed binary impulse input $x(t)$:

$$\begin{aligned} y(t) &= \int_{-\infty}^{\infty} x(\tau)h(t-\tau)d\tau \\ &= \sum_{k=-\infty}^{\infty} x(m\Delta t)h(t-m\Delta t) \end{aligned} \tag{3.111}$$

where the binary signal $x(m\Delta t)$ can assume a value of 0 or 1. In Equation (3.111) it is assumed that the impulses happen at the beginning of each interval. This assumption becomes more valid as Δt becomes smaller. The expectation of $y(t)$ is obtained as

$$\begin{aligned} \mathcal{E}[y(t)] &= \sum_{k=-\infty}^{\infty} \mathcal{E}[x(m\Delta t)]h(t-m\Delta t) \\ &= \sum_{k=-\infty}^{\infty} \lambda\Delta t\, h(t-m\Delta t) \end{aligned} \tag{3.112}$$

and

$$\begin{aligned} r_{yy}(t_1,t_2) &= \mathcal{E}[y(t_1)\,y(t_2)] \\ &= \sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} \mathcal{E}[x(m\Delta t)x(n\Delta t)]h(t_1-n\Delta t)h(t_2-m\Delta t) \end{aligned} \tag{3.113}$$

Using Equation (3.110), the autocorrelation of $y(t)$ can be obtained as

$$r_{yy}(t_1,t_2)=\sum_{m=-\infty}^{\infty}(\lambda\Delta t)h(t_1-m\Delta t)h(t_2-m\Delta t)+\sum_{m=-\infty}^{\infty}\sum_{\substack{n=-\infty\\ n\neq m}}^{\infty}(\lambda\Delta t)^2\,h(t_1-m\Delta t)h(t_2-n\Delta t) \tag{3.114}$$

3.5.7 Poisson–Gaussian Model for Clutters and Impulsive Noise

An impulsive noise process consists of a sequence of short-duration pulses of random amplitude and random time of occurrence whose shape and duration depends on the characteristics of the channel through which the impulse propagates. A Poisson process can be used to model the random time of occurrence of impulsive noise, and a Gaussian process can be used to model the random amplitude of the impulses. Finally, the finite duration character of real impulsive noise may be modelled by the impulse response of linear filter. The Poisson–Gaussian impulsive noise model is given by

$$x(m)=\sum_{k=-\infty}^{\infty}A_k h(m-\tau_k) \tag{3.115}$$

where $h(m)$ is the response of a linear filter that models the shape of impulsive noise, A_k is a zero-mean Gaussian process of variance σ^2 and τ_k is a Poisson process. The output of a filter excited by a Poisson-distributed sequence of Gaussian amplitude impulses can also be used to model clutters in radar. Clutters are due to reflection of radar pulses from a multitude of background surfaces and objects other than the radar target.

3.5.8 Markov Processes

A first-order discrete-time Markov process is defined as one in which the state of the process at time m depends only on its state at time m–1 and is independent of the process history before m–1. In probabilistic terms, a first-order Markov process can be defined as

$$\begin{aligned} f_X\left(x(m)=x_m \,|\, x(m-1)=x_{m-1},\ldots,x(m-N)=x_{m-N}\right) \\ = f_X\left(x(m)=x_m \,|\, x(m-1)=x_{m-1}\right) \end{aligned} \tag{3.116}$$

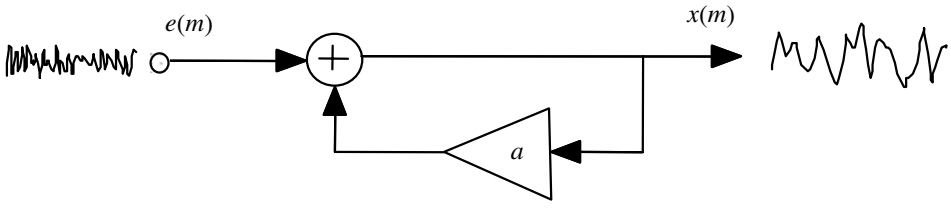


Figure 3.12 A first order autoregressive (Markov) process.

The marginal density of a Markov process at time m can be obtained by integrating the conditional density over all values of $x(m–1)$:

$$f_X(x(m)=x_m)=\int_{-\infty}^{\infty} f_X\left(x(m)=x_m|x(m-1)=x_{m-1}\right)f_X\left(x(m-1)=x_{m-1}\right)dx_{m-1} \tag{3.117}$$

A process in which the present state of the system depends on the past n states may be described in terms of n first-order Markov processes and is known as an n^{th} order Markov process. The term "Markov process" usually refers to a first order process.

Example 3.10 A simple example of a Markov process is a first-order autoregressive process (Figure 3.12) defined as

$$x(m)=ax(m-1)+e(m) \tag{3.118}$$

In Equation (3.118), $x(m)$ depends on the previous value $x(m–1)$ and the input $e(m)$. The conditional pdf of $x(m)$ given the previous sample value can be expressed as

$$\begin{aligned} f_X\left(x(m)|x(m-1),...,x(m-N)\right) &= f_X\left(x(m)|x(m-1)\right) \\ &= f_E\left(e(m)=x(m)-ax(m-1)\right) \end{aligned} \tag{3.119}$$

where $f_E(e(m))$ is the pdf of the input signal $e(m)$. Assuming that input $e(m)$ is a zero-mean Gaussian process with variance σ_e^2, we have

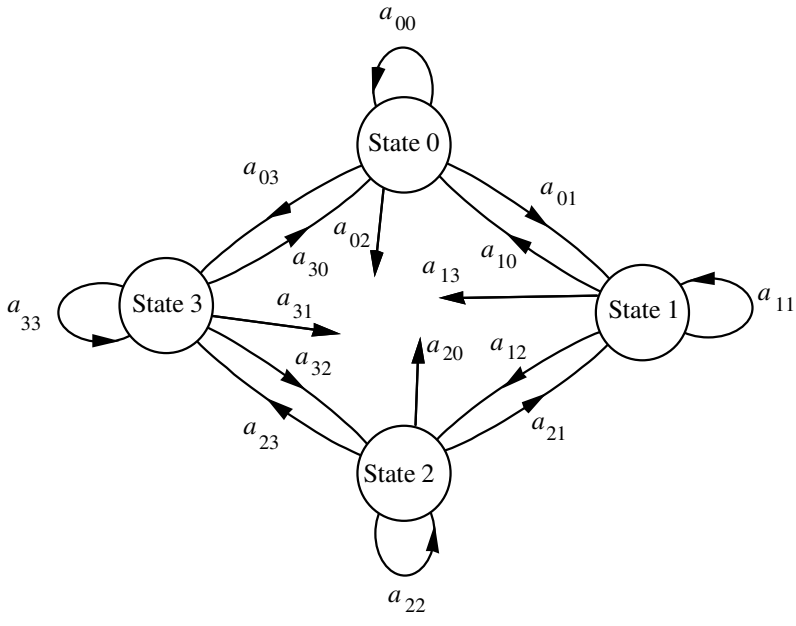


Figure 3.13 A Markov chain model of a four-state discrete-time Markov process.

$$\begin{aligned} f_X\left(x(m)|x(m-1)...,x(m-N)\right) &= f_X\left(x(m)|x(m-1)\right) \\ &= f_E\left(x(m)-ax(m-1)\right) \\ &= \frac{1}{\sqrt{2\pi}\sigma_e}\exp\left[-\frac{1}{2\sigma_e^2}\left(x(m)-ax(m-1)\right)^2\right] \end{aligned} \tag{3.120}$$

When the input to a Markov model is a Gaussian process the output is known as a Gauss–Markov process.

3.5.9 Markov Chain Processes

A discrete-time Markov process $x(m)$ with N allowable states may be modelled by a Markov chain of N states (Figure 3.13). Each state can be associated with one of the N values that $x(m)$ may assume. In a Markov chain, the Markovian property is modelled by a set of state transition probabilities defined as

$$a_{ij}(m-1,m)=Prob\big(x(m)=j|x(m-1)=i\big) \tag{3.121}$$

where $a_{ij}(m,m–1)$ is the probability that at time $m–1$ the process is in the state i and then at time m it moves to state j. In Equation (3.121), the transition probability is expressed in a general time-dependent form. The marginal probability that a Markov process is in the state j at time m, $P_j(m)$, can be expressed as

$$P_j(m) = \sum_{i=1}^{N} P_i(m-1)\,a_{ij}(m-1,m) \tag{3.122}$$

A Markov chain is defined by the following set of parameters:

number of states N
state probability vector

$$\boldsymbol{p}^{\mathrm{T}}(m)=[p_1(m),p_2(m),\ldots,p_N(m)]$$

and the state transition matrix

$$\boldsymbol{A}(m-1,m)=\begin{pmatrix} a_{11}(m-1,m) & a_{12}(m-1,m) & \ldots & a_{1N}(m-1,m) \\ a_{21}(m-1,m) & a_{22}(m-1,m) & \cdots & a_{2N}(m-1,m) \\ \vdots & \vdots & \ddots & \vdots \\ a_{N1}(m-1,m) & a_{N2}(m-1,m) & \ldots & a_{NN}(m-1,m) \end{pmatrix}$$

Homogenous and Inhomogeneous Markov Chains

A Markov chain with time-invariant state transition probabilities is known as a homogenous Markov chain. For a homogenous Markov process, the probability of a transition from state i to state j of the process is independent of the time of the transition m, as expressed in the following equation:

$$Prob\big(x(m)=j|x(m-1)=i\big)=a_{ij}(m-1,m)=a_{ij} \tag{3.123}$$

Inhomgeneous Markov chains have time-dependent transition probabilities. In most applications of Markov chains, homogenous models are used because they usually provide an adequate model of the signal process, and because homogenous Markov models are easier to train and use. Markov models are considered in Chapter 5.

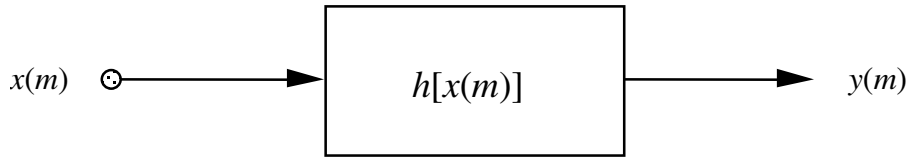


Figure 3.14 Transformation of a random process $x(m)$ to an output process $y(m)$.

3.6 Transformation of a Random Process

In this section we consider the effect of filtering or transformation of a random process on its probability density function. Figure 3.14 shows a generalised mapping operator $h(\cdot)$ that transforms a random input process X into an output process Y. The input and output signals $x(m)$ and $y(m)$ are realisations of the random processes X and Y respectively. If $x(m)$ and $y(m)$ are both discrete-valued such that $x(m) \in \{x_1,\ldots,x_N\}$ and $y(m) \in \{y_1,\ldots,y_M\}$ then we have

$$P_Y\left(y(m) = y_j\right) = \sum_{x_i \to y_j} P_X\left(x(m) = x_i\right) \tag{3.124}$$

where the summation is taken over all values of $x(m)$ that map to $y(m)=y_j$. Now consider the transformation of a discrete-time, *continuous-valued*, process. The probability that the output process Y has a value in the range $y(m)<Y<y(m)+\Delta y$ is

$$Prob[\,y(m)< Y < y(m)+\Delta y] = \int_{x(m)|\,y(m)<Y<y(m)+\Delta y} f_X\left(x(m)\right)dx(m) \tag{3.125}$$

where the integration is taken over all the values of $x(m)$ that yield an output in the range $y(m)$ to $y(m)+\Delta y$.

3.6.1 Monotonic Transformation of Random Processes

Now for a monotonic one-to-one transformation $y(m)=h[x(m)]$ (e.g. as in Figure 3.15) Equation (3.125) becomes

$$Prob\left(y(m)< Y < y(m)+\Delta y\right) = Prob\left(x(m)< X < x(m)+\Delta x\right) \tag{3.126}$$

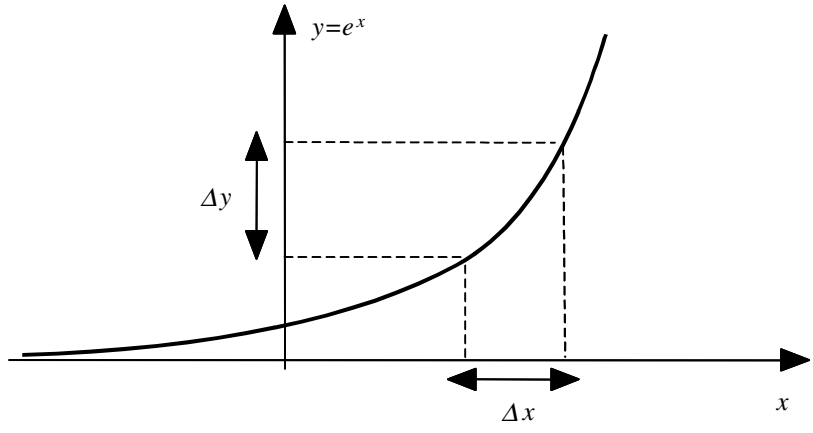


Figure 3.15 An example of a monotonic one-to-one mapping.

or, in terms of the cumulative distribution functions

$$F_Y(y(m)+\Delta y)-F_Y(y(m))=F_X(x(m)+\Delta x)-F_X(x(m)) \tag{3.127}$$

Multiplication of the left-hand side of Equation (3.127) by $\Delta y/\Delta y$ and the right-hand side by $\Delta x/\Delta x$ and re-arrangement of the terms yields

$$\frac{F_Y(y(m)+\Delta y)-F_Y(y(m))}{\Delta y}=\frac{\Delta x}{\Delta y}\,\frac{F_X(x(m)+\Delta x)-F_X(x(m))}{\Delta x} \tag{3.128}$$

Now as the intervals Δx and Δy tend to zero, Equation (3.128) becomes

$$f_Y(y(m))=\left|\frac{\partial x(m)}{\partial y(m)}\right| f_X(x(m)) \tag{3.129}$$

where $f_Y(y(m))$ is the probability density function. In Equation (3.129), substitution of $x(m)=h^{-1}(y(m))$ yields

$$f_Y(y(m))=\left|\frac{\partial\, h^{-1}(y(m))}{\partial\, y(m)}\right| f_X\left(h^{-1}(y(m))\right) \tag{3.130}$$

Equation (3.130) gives the pdf of the output signal in terms of the pdf of the input signal and the transformation.

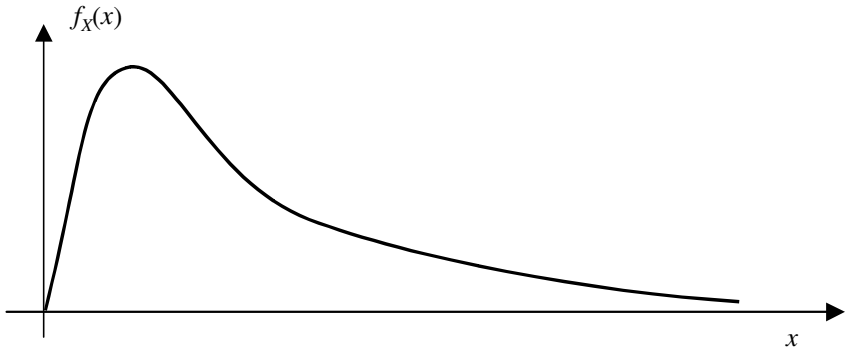


Figure 3.16 A log-normal distribution.

Example 3.11 Transformation of a Gaussian process to a log-normal process. Log-normal pdfs are used for modelling positive-valued processes such as power spectra. If a random variable $x(m)$ has a Gaussian pdf as in Equation (3.80) then the non-negative valued variable $y(m)=\exp(x(m))$ has a log-normal distribution (Figure 3.16) obtained using Equation (3.130) as

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\,\sigma_x\, y(m)} \exp\left\{-\frac{[\ln y(m) - \mu_x]^2}{2\sigma_x^2}\right\} \tag{3.131}$$

Conversely, if the input y to a logarithmic function has a log-normal distribution then the output $x=\ln y$ is Gaussian. The mapping functions for translating the mean and variance of a log-normal distribution to a normal distribution can be derived as

$$\mu_x = \ln \mu_y - \frac{1}{2}\ln\left(1+\sigma_y^2/\mu_y^2\right) \tag{3.132}$$

$$\sigma_x^2 = \ln\left(1+\sigma_y^2/\mu_y^2\right) \tag{3.133}$$

(μ_x, σ_x^2), and (μ_y, σ_y^2) are the mean and variance of x and y respectively. The inverse mapping relations for the translation of mean and variances of normal to log-normal variables are

$$\mu_y = \exp(\mu_x + \sigma_x^2/2) \tag{3.134}$$

$$\sigma_y^2 = \mu_x^2\,[\exp(\sigma_y^2) - 1] \tag{3.135}$$

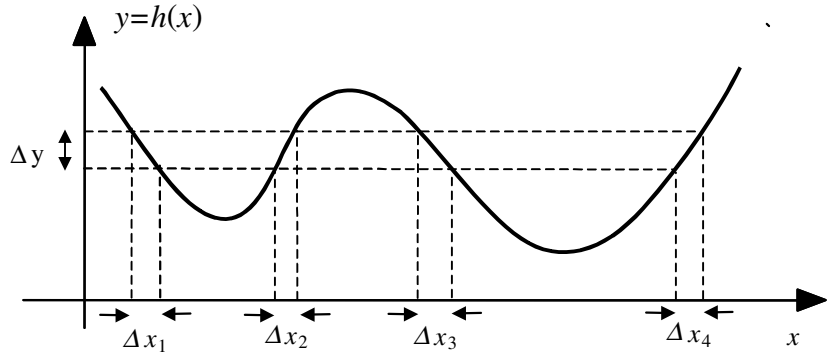


Figure 3.17 Illustration of a many to one transformation.

3.6.2 Many-to-One Mapping of Random Signals

Now consider the case when the transformation $h(\cdot)$ is a non-monotonic function such as that shown in Figure 3.17. Assuming that the equation $y(m)=h[x(m)]$ has K roots, there are K different values of $x(m)$ that map to the same $y(m)$. The probability that a realisation of the output process Y has a value in the range $y(m)$ to $y(m)+\Delta y$ is given by

$$Prob(y(m) < Y < y(m)+\Delta y) = \sum_{k=1}^{K} Prob(x_k(m) < X < x_k(m)+\Delta x_k) \quad (3.136)$$

where x_k is the k^{th} root of $y(m)=h(x(m))$. Similar to the development in Section 3.6.1, Equation (3.136) can be written as

$$\frac{F_Y(y(m)+\Delta y)-F_Y(y(m))}{\Delta y}\Delta y = \sum_{k=1}^{K} \frac{F_X(x_k(m)+\Delta x_k)-F_X(x_k(m))}{\Delta x_k}\Delta x_k \quad (3.137)$$

Equation (3.137) can be rearranged as

$$\frac{F_Y(y(m)+\Delta y)-F_Y(y(m))}{\Delta y} = \sum_{k=1}^{K} \frac{\Delta x_k}{\Delta y}\frac{F_X(x_k(m)+\Delta x_k)-F_X(x_k(m))}{\Delta x_k} \quad (3.138)$$

Now as the intervals Δx and Δy tend to zero Equation (3.138) becomes

$$f_Y(y(m)) = \sum_{k=1}^{K} \left| \frac{\partial x_k(m)}{\partial y(m)} \right| f_X(x_k(m)) = \sum_{k=1}^{K} \frac{1}{|h'(x_k(m))|} f_X(x_k(m)) \tag{3.139}$$

where $h'(x_k(m)) = \partial h(x_k(m)) / \partial x_k(m)$. Note that for a monotonic function, $K=1$ and Equation (3.139) becomes the same as Equation (3.130). Equation (3.139) can be expressed as

$$f_Y(y(m)) = \sum_{k=1}^{K} |J(x_k(m))|^{-1} f_X(x_k(m)) \tag{3.140}$$

where $J(x_k(m)) = h'(x_k(m))$ is called the Jacobian of the transformation. For a multi-variate transformation of a vector-valued process such as

$$\boldsymbol{y}(m) = \boldsymbol{H}(\boldsymbol{x}(m)) \tag{3.141}$$

the pdf of the output $\boldsymbol{y}(m)$ is given by

$$f_{\boldsymbol{Y}}(\boldsymbol{y}(m)) = \sum_{k=1}^{K} |\boldsymbol{J}(\boldsymbol{x}_k(m))|^{-1} f_{\boldsymbol{X}}(\boldsymbol{x}_k(m)) \tag{3.142}$$

where $|\boldsymbol{J}(\boldsymbol{x})|$, the Jacobian of the transformation $\boldsymbol{H}(\cdot)$, is the determinant of a matrix of derivatives:

$$|\boldsymbol{J}(\boldsymbol{x})| = \begin{vmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_P} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_P}{\partial x_1} & \frac{\partial y_P}{\partial x_2} & \cdots & \frac{\partial y_P}{\partial x_P} \end{vmatrix} \tag{3.143}$$

For a monotonic linear vector transformation such as

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} \tag{3.144}$$

the pdf of $\boldsymbol{y}$ becomes

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = |\boldsymbol{J}|^{-1} f_{\boldsymbol{X}}(\boldsymbol{H}^{-1}\boldsymbol{y}) \tag{3.145}$$

where $|\boldsymbol{J}|$ is the Jacobian of the transformation.

Example 3.12 The input–output relation of a $P \times P$ linear transformation matrix $\boldsymbol{H}$ is given by

$$\boldsymbol{y} = \boldsymbol{H}\,\boldsymbol{x} \tag{3.146}$$

The Jacobian of the linear transformation $\boldsymbol{H}$ is $|\boldsymbol{H}|$. Assume that the input $\boldsymbol{x}$ is a zero-mean Gaussian P-variate process with a covariance matrix of $\boldsymbol{\Sigma}_{\boldsymbol{xx}}$ and a probability density function given by:

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{(2\pi)^{P/2}|\boldsymbol{\Sigma}_{\boldsymbol{xx}}|^{1/2}} \exp\left[-\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\Sigma}_{\boldsymbol{xx}}^{-1}\boldsymbol{x}\right] \tag{3.147}$$

From Equations (3.145)–(3.147), the pdf of the output $\boldsymbol{y}$ is given by

$$\begin{aligned} f_{\boldsymbol{Y}}(\boldsymbol{y}) &= \frac{1}{(2\pi)^{P/2}|\boldsymbol{\Sigma}_{\boldsymbol{xx}}|^{1/2}} \exp\left(-\frac{1}{2}\boldsymbol{y}^{\mathrm{T}}\boldsymbol{H}^{-1^{\mathrm{T}}}\boldsymbol{\Sigma}_{\boldsymbol{xx}}^{-1}\boldsymbol{H}^{-1}\boldsymbol{y}\right)|\boldsymbol{H}|^{-1} \\ &= \frac{1}{(2\pi)^{P/2}|\boldsymbol{\Sigma}_{\boldsymbol{xx}}|^{1/2}|\boldsymbol{H}|} \exp\left(-\frac{1}{2}\boldsymbol{y}^{\mathrm{T}}\boldsymbol{\Sigma}_{\boldsymbol{yy}}^{-1}\boldsymbol{y}\right) \end{aligned} \tag{3.148}$$

where $\boldsymbol{\Sigma}_{\boldsymbol{yy}} = \boldsymbol{H}\boldsymbol{\Sigma}_{\boldsymbol{xx}}\boldsymbol{H}^{\mathrm{T}}$. Note that a linear transformation of a Gaussian process yields another Gaussian process.

3.7 Summary

The theory of statistical processes is central to the development of signal processing algorithms. We began this chapter with basic definitions of deterministic signals, random signals and random processes. A random process generates random signals, and the collection of all signals that can be generated by a random process is the space of the process. Probabilistic models and statistical measures, originally developed for random variables, were extended to model random signals. Although random signals are completely described in terms of probabilistic models, for many applications it may be sufficient to characterise a process in terms of a set of relatively simple statistics such as the mean, the autocorrelation function, the covariance and the power spectrum. Much of the theory and application of signal processing is concerned with the identification, extraction, and utilisation of structures and patterns in a signal process. The correlation and

its Fourier transform the power spectrum are particularly important because they can be used to identify the patterns in a stochastic process.

We considered the concepts of stationary, ergodic stationary and non-stationary processes. The concept of a stationary process is central to the theory of linear time-invariant systems, and furthermore even non-stationary processes can be modelled with a chain of stationary subprocesses as described in Chapter 5 on hidden Markov models. For signal processing applications, a number of useful pdfs, including the Gaussian, the mixture Gaussian, the Markov and the Poisson process, were considered. These pdf models are extensively employed in the remainder of this book. Signal processing normally involves the filtering or transformation of an input signal to an output signal. We derived general expressions for the pdf of the output of a system in terms of the pdf of the input. We also considered some applications of stochastic processes for modelling random noise such as white noise, clutters, shot noise and impulsive noise.

Bibliography

ANDERSON O.D. (1976) Time Series Analysis and Forecasting. The Box–Jenkins Approach. Butterworth, London.

AYRE A.J. (1972) Probability and Evidence. Columbia University Press, New York.

BARTLETT M.S. (1960) Stochastic Processes. Cambridge University Press.

BOX G.E.P and JENKINS G.M. (1976) Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco.

BREIPHOL A.M. (1970) Probabilistic System Analysis. Wiley, New York.

CARTER G. (1987) Coherence and Time Delay Estimation. Proc. IEEE, **75**, **2**, pp. 236–55.

CLARK A.B. and DISNEY R.L. (1985) Probability and Random Processes, 2nd Ed. Wiley, New York.

COOPER G.R. and MCGILLEM C.D. (1986) Probabilistic Methods of Signal and System Analysis Holt. Rinehart and Winston, New York.

DAVENPORT W.B. and ROOT W.L. (1958) Introduction to Random Signals and Noise. McGraw-Hill, New York.

DAVENPORT W.B. and WILBUR B. (1970) Probability and Random Processes: An Introduction for Applied Scientists and Engineers. McGraw-Hill, New York.

EINSTEIN A. (1956) Investigation on the Theory of Brownian Motion. Dover, New York.

GAUSS C.F. (1963) Theory of Motion of Heavenly Bodies. Dover, New York.

GARDENER W.A. (1986) Introduction to Random Processes: With Application to Signals and Systems. Macmillan, New York.

HELSTROM C.W. (1991) Probability and Stochastic Processes for Engineers. Macmillan, New York.

ISAACSON D. and MASDEN R. (1976) Markov Chains Theory and Applications. Wiley, New York.

JEFFREY H. (1961) Scientific Inference, 3rd Ed. Cambridge University Press.

JEFFREY H. (1973) Theory of Probability, 3rd Ed. Clarendon Press, Oxford.

KAY S.M. (1993) Fundamentals of Statistical Signal Processing. Estimation Theory. Prentice-Hall, Englewood Cliffs, NJ.

KOLMOGOROV A.N. (1956) Foundations of the Theory of Probability. Chelsea Publishing Company, New York.

KENDALL M. and STUART A. (1977) The Advanced Theory of Statistics. Macmillan.

LEON-GARCIA A. (1994) Probability and Random Processes for Electrical. Engineering. Addison Wesley, Reading, MA.

MARKOV A.A. (1913) An Example of Statistical Investigation in the text of *Eugen Onyegin* Illustrating Coupling of Tests in Chains. Proc. Acad. Sci. St Petersburg VI Ser., **7**, pp. 153–162.

MEYER P.L. (1970) Introductory Probability and Statistical Applications. Addison-Wesley, Reading, MA.

PEEBLES P.Z. (1987) Probability, Random Variables and Random Signal Principles. McGraw-Hill, New York.

PARZEN E. (1962) Stochastic Processes. Holden-Day, San Francisco.

PAPOPULIS A. (1984) Probability, Random Variables and Stochastic Processes. McGraw-Hill, New York.

PAPOPULIS A. (1977) Signal Analysis, McGraw-Hill, New York.

RAO C.R. (1973) Linear Statistical Inference and Its Applications. Wiley, New York.

ROZANOV Y.A. (1969) Probability Theory: A Concise Course, Dover, New York.

SHANMUGAN K.S. and BREIPOHL A.M. (1988) Random Signals: Detection, Estimation and Data Analysis. Wiley, New York.

THOMAS J.B. (1988) An introduction to Applied probability and Random Processes. Huntington, Krieger Publishing, New York.

WOZENCRAFT J.M. and JACOBS I.M. (1965) Principles of Communication Engineering. Wiley, New York