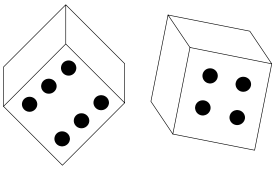

2

> 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 & INFORMATION MODELS

Probability models form the foundation of the information theory. Information itself is quantified in units of “bits” in terms of a logarithmic function of probability. Probability models are extensively used in communications and signal processing systems to characterise and predict the occurrence of random events in such diverse areas of applications as: speech/image recognition, audio/video coding, bioengineering, weather forecasting, financial data modelling, and predicting the demand on a service facility such as the number of telephone calls in a specified period of the day on a trunk line or in a cellular mobile phone network.

This chapter begins with a study of the basic concepts of random signals and the probability models used for the characterisation of random processes. The concepts of randomness, information and entropy are introduced and their close relationships explained. Random processes can be 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 stationary and non-stationary processes, and the concept of ergodic processes in which time averages obtained from a single realisation of a process can be used instead of ensemble averages. We consider some widely used classes of random processes, and study the effect of filtering or transformation of a signal on its probability distribution. Finally we consider several applications of probability models in communication signal processing.

2.1 Random Processes

Signals, in terms of one of their most fundamental characteristics, can be classified into two broad categories:

(a) *Deterministic* signals.
(b) *Random* 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 with 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 values. 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. *For a signal 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 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 2.1 Figure 2.1(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

$$x(m)=h_1\big(x(m-1),x(m-2),...,x(m-P)\big) \tag{2.1}$$

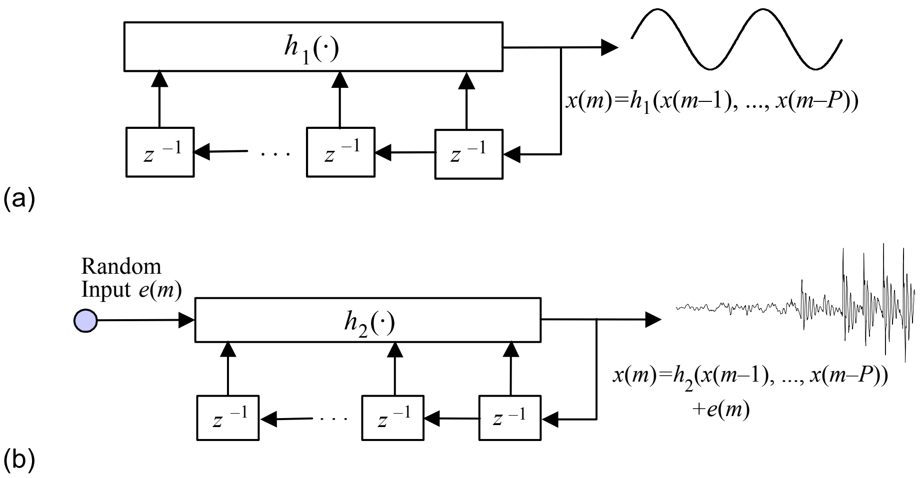

Figure 2.1 Illustration of deterministic and random signal models: (a) a deterministic signal model, (b) a random signal model.

where the function h_1 may be a linear or a non-linear model. A functional description of the model h_1 together with 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 (2.1) becomes

$$x(m)=a\,x(m-1)-x(m-2) \tag{2.2}$$

where the choice of the parameter $a=2\cos(2\pi F_0\ /F_s)$ determines the oscillation frequency F_0 of the sinusoid, at a sampling frequency of F_s. Figure 2.1(b) is a model for a random process given by

$$x(m)=h_2\big(x(m-1), x(m-2), ..., x(m-P)\big)+e(m) \tag{2.3}$$

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

$$x(m)=a_1\,x(m-1)+a_2\,x(m-2)+e(m) \tag{2.4}$$

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

2.1.2 The Space of a Random Process

The collection of all realisations of a random process is known as the space, or the ensemble, of the process. For an illustration, consider a random noise process over a communication network as shown in Figure 2.2. 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 space (or the ensemble) 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 such cases, 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 of all realisations of $X(m)$, denoted by $\{x(m,s)\}$, form the ensemble or the space of the random process $X(m)$.

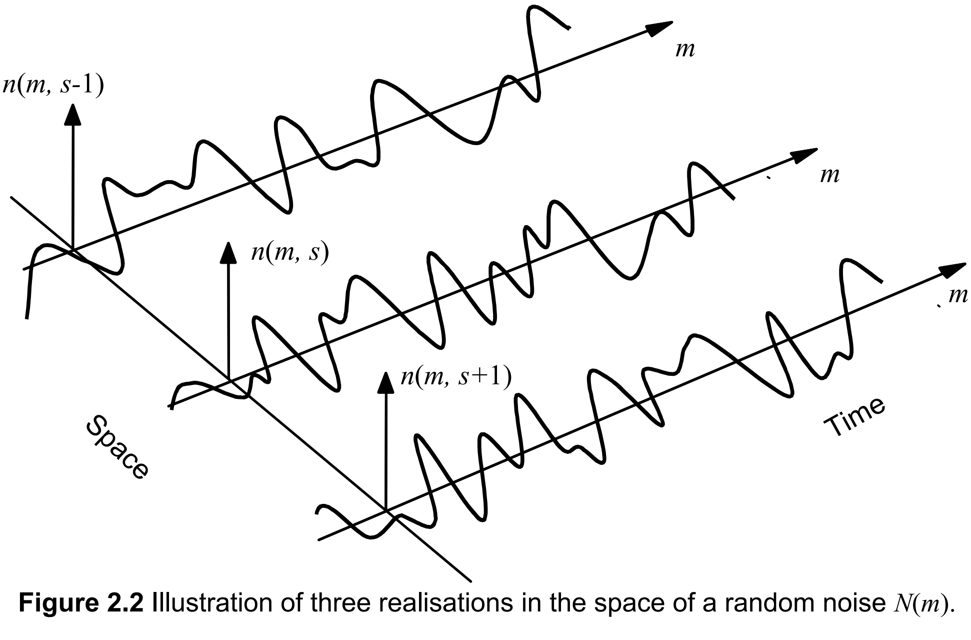

Figure 2.2 Illustration of three realisations in the space of a random noise $N(m)$.

2.2 Probabilistic Models

Probability models provide the most complete mathematical description of a random process. Consider a random process that generates a time-sequence of numbers $x(m)$. Let $\{x(m,s)\}$ denote a collection of different sequences generated by the same process where s is the index of the sequence space. 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 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.

2.2.1 Random Variables

Classical examples of random variables are the variables and outcomes in a gambling game, or in a chance process, such as throwing a coin, or a pair of dice. A random variable may be finite-valued or continuous-valued. For example a die has six faces, each face is an outcome and is assigned a numerical value from one to six, and a probability that is 1/6 for a fair die.
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

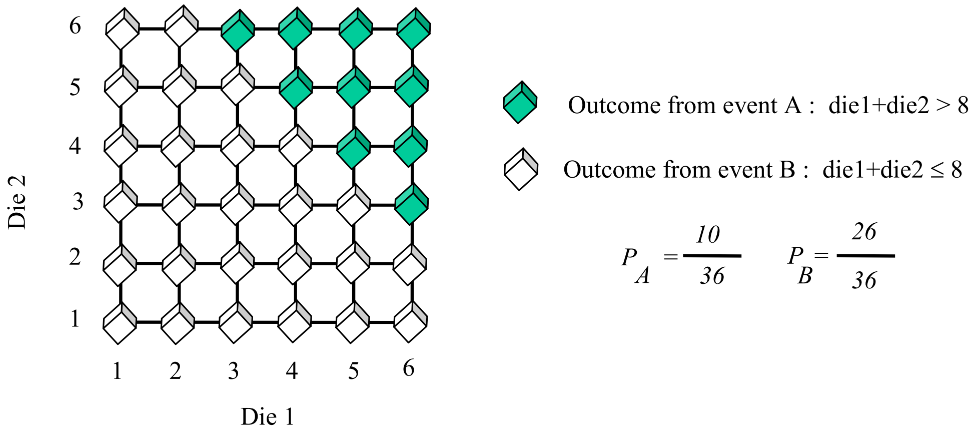

Figure 2.3 A two-dimensional representation of the outcomes of two dice, and the subspaces associated with the events corresponding to the sum of the dice being greater than 8 or, less than or equal to 8.

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

$$P(A) = \frac{N_A}{\sum\limits_{\text{All events } i} N_i} \tag{2.5}$$

From Equation (2.5), it is evident that the sum of the probabilities of all likely events in an experiment is unity.

Example 2.2 The space of two discrete numbers obtained as outcomes of throwing a pair of dice is shown in Figure 2.2. This space can be partitioned in different ways; for example, the two subspaces shown in Figure 2.3 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.

2.2.1 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

$$\begin{aligned} P_{X,Y}(x_i, y_j) &= P_{Y|X}(y_j | x_i) P_X(x_i) \\ &= P_{X|Y}(x_i | y_j) P_Y(y_j) \end{aligned} \tag{2.6}$$

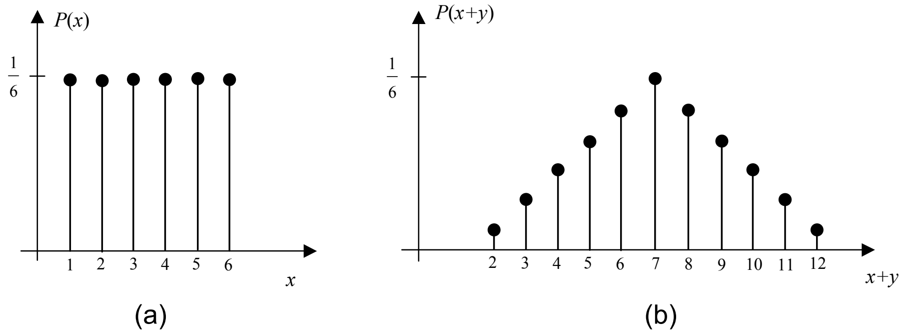

Figure 2.4 The probability mass function (pmf) of: (a) a die, and (b) the sum of a pair of dice.

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.

Example 2.3 *Probability of the sum of two random variables*. Figure 2.4(a) shows the pmf of a die. Now, let the variables (x, y) represent the outcomes of throwing a pair of dice. The probability that the sum of the two variables is A is given by

$$P(x+y=A)=\sum_{i=1}^{6} P(x=i)\,P(y=A-i) \tag{2.7}$$

The pmf of the sum of two dice is plotted in Figure 2.4(b). Note from Equation (2.7) that the probability of the sum of two random variables is the convolutional sum of the probability functions of the individual variables.

2.2.2 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:

$$F_X(x) = Prob\left(X \le x\right) \tag{2.8}$$

where *Prob*(·) 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

$$\begin{aligned}\frac{1}{\varDelta}Prob(x-\varDelta/2 \le X \le x+\varDelta/2) &= \frac{1}{\varDelta}\,[Prob(X \le x+\varDelta/2) - Prob(X \le x-\varDelta/2)]\\ &= \frac{1}{\varDelta}\,[F_X(x+\varDelta/2) - F_X(x-\varDelta/2)] \qquad (2.9)\end{aligned}$$

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

$$\begin{aligned}f_X(x) &= \lim_{\varDelta\to 0}\frac{1}{\varDelta}[F_X(x+\varDelta/2)-F_X(x-\varDelta/2)]\\ &= \frac{\partial F_X(x)}{\partial x}\end{aligned} \qquad (2.10)$$

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. $f_X(x) \ge 0$. The integral of the pdf of a random variable *X* in the range $\pm\infty$ is unity:

$$\int_{-\infty}^{\infty} f_X(x)dx = 1 \qquad (2.11)$$

2.3 Information Models

Information is the state or the outcome of a random process. In the news media a particular outcome of an information process about which there may be some uncertainty is treated as "news" when it happens. Information has a binary nature, hence any information can be represented as a sequence of binary numbers and stored on a computer memory. The concepts of information, randomness and probability models are closely related. For a signal to have information it must satisfy the following conditions:

(a) Posses two or more states or values.
(b) Move between the states in a random manner.

For example, the outcome of tossing a coin is an unpredictable binary (head/tail) event, the outcome of a weather forecast states can be one or a

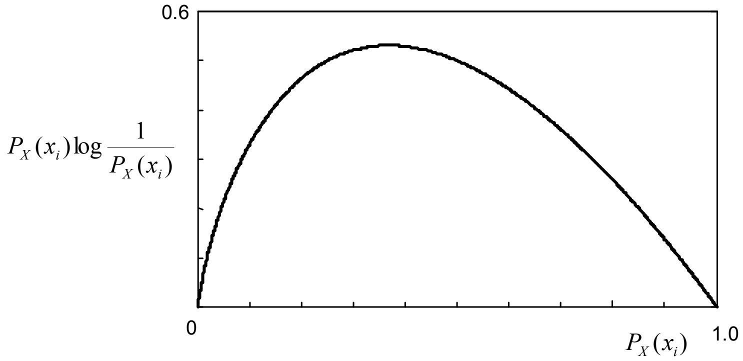

Figure 2.5 The information value of a random variable x_i as a function of its probability $P_X(x_i)$.

mutually nonexclusive combination of number of, the following states: {cold, warm, hot, cloud, sun, rain, snow, storm), and the DNA state sequence information determines a person’s physical characteristics.

Random signals are modelled with probability functions. It is therefore natural that information is also modelled as a function of probability. The information *content* of an event is modelled by

$$I(x_i)=P_X(x_i)\log\frac{1}{P_X(x_i)} \tag{2.12}$$

Note that information content of a variable has a value of zero for a certain event whose probability is one and hence contains no information, its value also approaches zero for an extremely unlikely event Fig. (2.5).

2.3.1 Entropy and Mutual Information

Consider a random variable X with M states $[x_1, x_2, \ldots, x_M]$ and state probabilities $[p_1, p_2, \ldots, p_M]$ where $P_X(x_i)=p_i$. The Entropy of X is given by

$$H(X)=\sum_{i=1}^{M}P_X(x_i)\log\frac{1}{P_X(x_i)} \tag{2.13}$$

The base of logarithm is 2 and the entropy is measured in units of bits. It represents the minimum number of bits required for binary coding of different values of X. The entropy of a set attains a maximum values of N bits for a *uniformly* distributed N-valued variable.

Example 2.4 Entropy of the English Alphabet. Calculate the information content of the set of English alphabet (A, B, C, D … Z), assuming that all letters are equally likely, and hence calculate the number of bits required to encode a text of 2000 words with an average of 5 letters per word.
For the English alphabet the number of symbols N=26, and assuming that all symbols are equally likely the probability of each symbol becomes p_i=1/26. Using Equation () we have The information content of an N-symbol random set is given by

$$H(X)=\sum_{i=1}^{26}-\frac{1}{26}\log_2\frac{1}{26}=4.7 \quad \text{bits}$$

Total number of bits for encoding 2000 words =4.7x2000x5=47 kbits.

Example 2.5 Entropy of English Phonemes. Spoken English language, speech, is based on the use of about 40 basic acoustic symbols, known as phonemes (or phonetic units), these are used to construct words, sentences etc. Assuming that all phonetic units are equi-probable, and that the average speaking rate is 10 phonemes/second, calculate the minimum number of bits per second required to encode speech at the average speaking rate.
For speech N=40 assume p_{i}=1/40

$$H(X)=\sum_{i=1}^{40}-\frac{1}{40}\log_2\frac{1}{40}=5.3 \quad \text{bits}$$

Number of bits/sec=5.3*10=53 bps.

2.3 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 2.6). The stationary property implies that all the statistical 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.

Example 2.6 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

$$Mean\left(A\sin\omega t\right) = \frac{1}{T}\int_T A\sin\omega t\, dt = 0, \qquad \text{constant} \qquad (2.14)$$

$$Power\left(A\sin\omega t\right) = \frac{1}{T}\int_T A^2\sin^2\omega t\, dt = \frac{A^2}{2}, \qquad \text{constant} \qquad (2.15)$$

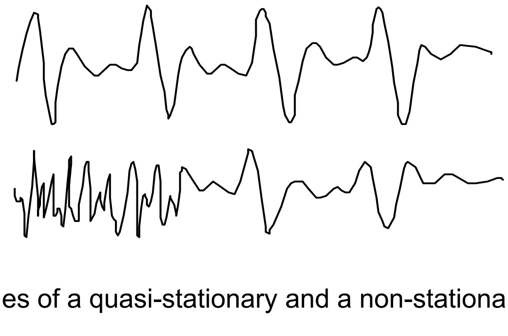

Figure 2.6 Examples of a quasi-stationary and a non-stationary speech segment.

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

$$Mean(Ae^{-\alpha t}) = \frac{1}{T}\int_{t}^{t+T} Ae^{-\alpha\tau}\,d\tau = \frac{A}{\alpha T}(1-e^{-\alpha T})e^{-\alpha t}, \qquad \text{time-varying} \tag{2.16}$$

$$Power\left(Ae^{-\alpha t}\right) = \frac{1}{T}\int_{t}^{t+T} A^2\,e^{-2\alpha\tau}\,d\tau = \frac{A^2}{2\alpha T}\left(1-e^{-2\alpha T}\right)e^{-2\alpha t}, \ \text{time-varying} \tag{2.17}$$

In Equations (2.16) and (2.17), the signal mean and power are exponentially decaying functions of the time variable t.

Example 2.7 Consider a non-stationary signal $y(m)$ generated by a binary-state random process described by the following equation:

$$y(m) = \bar{s}(m)x_0(m) + s(m)x_1(m) \tag{2.18}$$

where $s(m)$ is a binary-valued state indicator variable and $\bar{s}(m)$ denotes the binary complement of $s(m)$. From Equation (2.18), we have

$$y(m) = \begin{cases} x_0(m) & \text{if } s(m) = 0 \\ x_1(m) & \text{if } s(m) = 1 \end{cases} \tag{2.19}$$

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

$$\begin{aligned} \mathcal{E}[y(m)|s(m)] &= \bar{s}(m)\mathcal{E}[x_0(m)] + s(m)\mathcal{E}[x_1(m)] \\ &= \bar{s}(m)\mu_{x_0} + s(m)\mu_{x_1} \end{aligned} \tag{2.20}$$

In Equation (2.20), 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

$$\mathcal{E}\left[y^2(m)\middle|s(m)\right]=\bar{s}(m)\mathcal{E}\left[x_0^2(m)\right]+s(m)\mathcal{E}\left[x_1^2(m)\right] \\ =\bar{s}(m)P_{x_0}+s(m)P_{x_1} \tag{2.21}$$

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.

2.2.1 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. For a strict-sense stationary process, the mean, correlation and power spectrum are time-invariant and given by

$$\mathcal{E}[x(m)]=\mu_x \tag{2.22}$$

$$\mathcal{E}[x(m)x(m+k)]=r_{xx}(k) \tag{2.23}$$

and

$$\mathcal{E}[|X(f,m)|^2]=\mathcal{E}[|X(f)|^2]=P_{XX}(f) \tag{2.24}$$

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 $x(m)$.

2.2.2 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:

$$\mathcal{E}[x(m)] = \mu_x \tag{2.25}$$

$$\mathcal{E}[x(m)x(m+k)] = r_{xx}(k) \tag{2.26}$$

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.

2.2.3 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 2.7. In Figure 2.7(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 2.7(b) a time-varying process is modelled by a chain of time-invariant states, with each state having a different set of statistics or probability distributions. Finite state statistical models for time-varying processes are discussed in detail in Chapter 5.

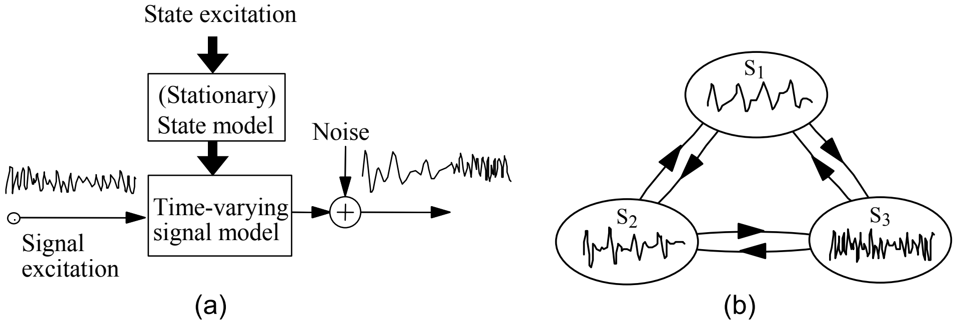

Figure 2.7 Two models for non-stationary processes: (a) a stationary process drives the parameters of a continuously time-varying model; (b) a finite-state model with each state having a different set of statistics.

2.4 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.
The most important, and widely used, expected values are the mean value, the correlation, the covariance, and the power spectrum.

2.4.1 The Mean Value

The mean value of a process plays an important part in signal processing and parameter estimation from noisy observations. For a segment of N samples of signal $x(m)$, the mean is obtained as

$$r_{xx}(m)=\frac{1}{N-1}\sum_{k=0}^{N-1}x(m) \tag{2.27}$$

2.4.2 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

$$r_{xx}(m_1,m_2) = \mathcal{E}[x(m_1)x(m_2)] \tag{2.28}$$

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 $m= m_1-m_2$:

$$r_{xx}(m_1+\tau, m_2+\tau) = r_{xx}(m_1,m_2) = r_{xx}(m_1-m_2) = r_{xx}(m) \tag{2.29}$$

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

$$r_{xx}(-m) = r_{xx}(m) \tag{2.30}$$

$$r_{xx}(m) \le r_{xx}(0) \tag{2.31}$$

Note that for a zero-mean signal, $r_{xx}(0)$ is the signal power. For a segment of N samples of signal $x(m)$, the autocorrelation function is obtained as

$$r_{xx}(m) = \frac{1}{N-1}\sum_{k=0}^{N-1-k} x(m)x(m+k) \tag{2.32}$$

2.4.4 Power Spectral Density

The power spectral density (PSD) function, also called the power spectrum, of a process gives the spectrum of the distribution of power along the frequency axis. 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:

$$\begin{aligned} P_{XX}(f) &= E[X(f)X^*(f)] \\ &= \sum_{m=-\infty}^{\infty} r_{xx}(k)\, e^{-j2\pi fm} \end{aligned} \tag{2.33}$$

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

$$P_{XX}(f) = r_{xx}(0) + \sum_{m=1}^{\infty} 2r_{xx}(m)\cos(2\pi fm) \tag{2.34}$$

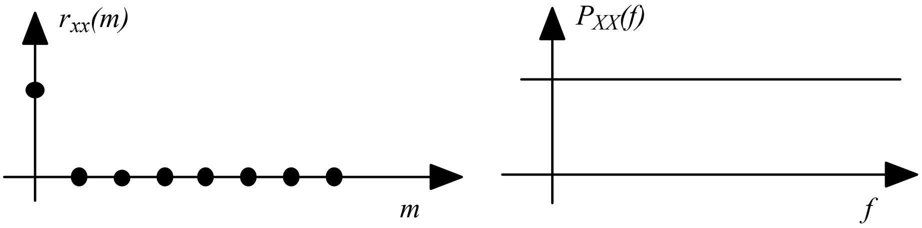

Figure 2.8 Autocorrelation and power spectrum of white noise.

The power spectral density is a real-valued non-negative function, expressed in units of watts per hertz. From Equation (2.34), the autocorrelation sequence of a random process may be obtained as the inverse Fourier transform of the power spectrum as

$$r_{xx}(m) = \int_{-1/2}^{1/2} P_{XX}(f)\, e^{j2\pi fm}\, df \tag{2.35}$$

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

Example 2.8 Power spectrum and autocorrelation of white noise (Figure2.8). 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:

$$r_{nn}(k) = \mathcal{E}[n(m)n(m+k)] = \begin{cases} \text{Noisepower} & k = 0 \\ 0 & k \neq 0 \end{cases} \tag{2.36}$$

Equation (2.36) 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 $r_{nn}(k)$:

$$P_{NN}(f) = \sum_{k=-\infty}^{\infty} r_{nn}(k) e^{-j2\pi fk} = r_{nn}(0) = \text{noise power} \tag{2.37}$$

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

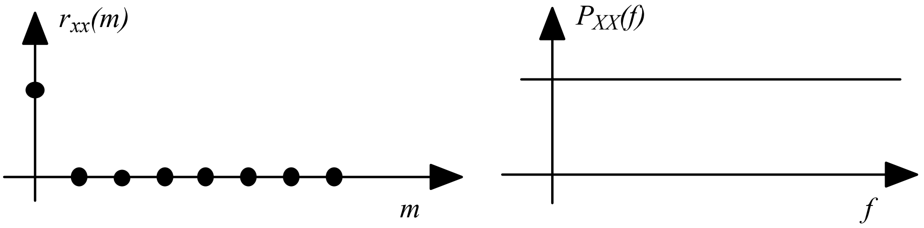

Figure 2.6 Autocorrelation and power spectrum of a discrete-time white noise.

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 2.7 Power spectrum and autocorrelation of a discrete-time impulse. The autocorrelation of an impulse with amplitude A, $A\delta(m)$, is defined as:

$$r_{\delta\delta}(k) = \mathcal{E}\ [\mathrm{A}^2\delta(m)\delta(m+k)] = \begin{cases} \mathrm{A}^2 & k = 0 \\ 0 & k \neq 0 \end{cases} \tag{2.38}$$

The power spectrum of the impulse is the obtained by taking the Fourier transform of $r_{\delta\delta}(k)$:

$$P_{\Delta\Delta}(f) = \sum_{k=-\infty}^{\infty} r_{\delta\delta}(k) e^{-j2\pi fk} = A^2 \tag{2.39}$$