Preface

The purpose of this book is to give the reader two things, to paraphrase Mark Twain: Roots to know the basics of modeling networks and Wings to fly away and attempt modeling other proposed systems of interest.

The Internet phenomenon is affecting us all in the way we communicate, conduct business, and access information and entertainment. More unforeseen applications are still to come. All of this is due to the existence of an efficient global high-performance network that connects millions of users and moves information at a high rate with small delay.

High-Performance Networks

A high-performance network is characterized by two performance measures *bandwidth* and *delay*. Traditional network design focused mainly on bandwidth planning; the solution to network problems was to add more bandwidth. Nowadays, we have to consider message delay particularly for delay-sensitive applications such as voice and real-time video. Both bandwidth and delay contribute to the performance of the network. Bandwidth can be easily increased by compressing the data, by using links with higher speed, or by transmitting several bits in parallel using sophisticated modulation techniques. Delay, however, is not so easily improved. It can only be reduced by the use of good scheduling protocols, very fast hardware and switching equipment throughout the network. The increasing use of optical fibers means that the transmission channel is close to ideal with extremely high bandwidth and low delay (speed of light). The areas that need optimization are the interfaces and devices that connect the different links together such as hubs, switches, routers, and bridges. The goal of this book is to explore the design and analysis techniques of these devices. There are indications, however, that the optical fiber channel is becoming less than ideal due to the increasing bit rates. Furthermore, the use of wireless mobile networking is becoming very popular. Thus new and improved techniques for transmitting across the noisy, and band-limited, channel become very essential. The work to be done to optimize the physical level of communication is devising algorithms and hardware for adaptive data coding and compression. Thus digital signal processing is finding an increasing and pivotal role in the area of networking.

Scope

The three main building blocks of high-performance networks are the links, the switching equipment connecting the links together, and the software employed at the nodes and switches. The purpose of this book is to provide the basic techniques for modeling and analyzing the last two components: the software and the switching equipment. The book also reviews the design options used to build efficient switching equipment. For this purpose, different topics are covered in the book such as Markov chains and queuing analysis, traffic modeling, interconnection networks, and switch architectures and buffering strategies.

There are many books and articles dealing with continuous-time Markov chains and queuing analysis. This is because continuous-time systems are thought to be easily modeled and analyzed. However, digital communications are discrete in nature. Luckily, discrete-time Markov chains are simple, if not even easier, to analyze. The approach we chose to present Markov chains and queuing analysis is to start with explaining the basic concepts, then explain the analytic and numerical techniques that could be used to study the system. We introduce many worked examples throughout to get a feel as to how to apply discrete-time Markov chains to many communication systems.

We employ MATLAB ® throughout this book due to its popularity among engineers and engineering students. There are many equally useful mathematical packages available nowadays on many workstations and personal computers such as Maple ® and Mathematica ®.

Organization

This book covers the mathematical theory and techniques necessary for analyzing telecommunication systems. Queuing and Markov chain analyses are provided for many protocols that are used in networking. The book then discusses in detail applications of Markov chains and queuing analysis to model over 15 communications protocols and hardware components. Several appendices are also provided that round up the discussion and provide a handy reference for the necessary background material.

Chapter 1 discusses probability theory and random variables. There is discussion of sample spaces and how to count the number of outcomes of a random experiment. Also discussed is probability density function and expectations. Important distributions are discussed since they will be used for describing traffic in our analysis. The Pareto distribution is discussed in this chapter, which is usually not discussed in standard engineering texts on probability. Perhaps what is new in this chapter is the review of techniques for generating random numbers that obey a desired probability distribution. Inclusion of this material rounds up the chapter and helps the designer or researcher to generate the network traffic data needed to simulate a switch under specified conditions.

Chapter 2 discusses random processes and in particular Poisson and exponential processes. The chapter also discusses concepts associated with random processes such as ensemble average, time average, autocorrelation function, and cross-correlation function.

Chapter 3 discusses discrete-time Markov chains. Techniques for constructing the state transition matrix are explored in detail as well as how the time step is determined since all discrete-time Markov chains require awareness of the time step value. The chapter also discusses transient behavior of Markov chains and explains the various techniques for studying it such as diagonalization, expansion of the initial distribution vector, Jordan canonic form, and using the z-transform.

Chapter 4 discusses Markov chains at equilibrium, or steady state. Analytic techniques for finding the equilibrium distribution vector are explained such as finding the eigenvalues and eigenvectors of the state transition matrix, solving difference equations, and the z-transform technique. Several numerical techniques for finding the steady-state distribution are discussed such as use of forward- and backward-substitution, and iterative equations. The concepts of balance equations and flow balance are also explained.

Chapter 5 discusses reducible Markov chains and explains the concept of closed and transient states. The transition matrix for a reducible Markov chain is partitioned into blocks and the closed and transient states are related to each partitioning block. An expression is derived for the state of a Markov chain at any time instant n and also at equilibrium. The chapter also discusses how a reducible Markov chain could be identified by studying its eigenvalues and eigenvectors. It is shown that the eigenvectors enable us to identify all sets of closed and transient states.

Chapter 6 discusses periodic Markov chains. Two types of periodic Markov chains are identified and discussed separately. The eigenvalues of periodic Markov chains are discussed and related to the periodicity of the system. Transient analysis of a periodic Markov chain is discussed in detail and asymptotic behavior is analyzed.

Chapter 7 discusses discrete-time queues and queuing analysis. Kendall's notation is explained and several discrete-time queues are analyzed such as the infinite-sized $M/M/1$ queue and the finite-sized $M/M/1/B$ queue. Equally important queues encountered in this book are also considered such as $M^m/M/1/B$ and $M/M^m/1/B$ queues. The important performance parameters considered for each queue are the throughput, delay, average queue size, loss probability, and efficiency. The chapter also discusses how to analyze networks of queues using two techniques: the flow balance approach and the merged approach.

Chapter 8 discusses the modeling of several flow control protocols using Markov chains and queuing analysis. Three traffic management protocols are considered: leaky bucket, token bucket, and the virtual scheduling (VS) algorithm.

Chapter 9 discusses the modeling of several error control protocols using Markov chains and queuing analysis. Three error control using automatic repeat request algorithms are considered: stop-and-wait (SW ARQ), go-back-N (GBN ARQ), and selective repeat protocol (SRP ARQ).

Chapter 10 discusses the modeling of several medium access control protocols using Markov chains and queuing analysis. Several media access protocols are discussed: IEEE Standard 802.1p (static priority), pure and slotted ALOHA, IEEE Standard 802.3 (CSMA/CD, Ethernet), Carrier sense multiple access with collision avoidance (CSMA/CA), IEEE Standard 802.4 (token bus) & 802.5 (Token ring), IEEE Standard 802.6 (DQDB), IEEE Standard 802.11 distributed coordination function for ad hoc networks, and IEEE Standard 802.11 point coordination function for infrastructure networks (1-persistent and p-persistent cases are considered).

Chapter 11 discusses the different models used to describe telecommunication traffic. The topics discussed deal with describing the data arrival rates, data destinations, and packet length variation. The interarrival time for Poisson traffic is discussed in detail and a realistic model for Poisson traffic is proposed. Extracting the parameters of the Poisson traffic model is explained given a source average and burst rates. The interarrival time for Bernoulli sources is similarly treated and a realistic model is proposed together with a discussion on how to determine the Bernoulli model parameters. Self-similar traffic is discussed and the Pareto model is discussed. Extracting the parameters of the Pareto traffic model is explained given a source average and burst rates. Modulated Poisson traffic models are also discussed such as the on–off model and the Markov modulated Poisson process. In addition to modeling data arrival processes, the chapter also discusses the traffic destination statistics for uniform, broadcast, and hot-spot traffic types. The chapter finishes by discussing packet length statistics and how to model them.

Chapter 12 discusses scheduling algorithms. The differences and similarities between scheduling algorithms and media access protocols are discussed. Scheduler performance measures are explained and scheduler types or classifications are explained. The concept of max–min fairness is explained since it is essential for the discussion of scheduling algorithms. Twelve scheduling algorithms are explained and analyzed: first-in/first-out (FIFO), static priority, round robin (RR), weighted round robin (WRR), processor sharing (PS), generalized processor sharing (GPS), fair queuing (FQ), packet-by-packet GPS (PGPS), weighted fair queuing (WFQ), frame-based fair queuing (FFQ), core-stateless fair queuing (CSFQ), and finally random early detection (RED).

Chapter 13 discusses network switches and their design options. Media access techniques are first discussed since networking is about sharing limited resources using a variety of multiplexing techniques. Circuit and packet-switching are discussed and packet switching hardware is reviewed. The basic switch components are explained and the main types of switches are discussed: input queuing, output queuing, shared buffer, multiple input queue, multiple output queue, multiple input and output queue, and virtual routing/virtual queuing (VRQ). A qualitative discussion of the advantages and disadvantages of each switch type is provided. Detailed quantitative analyses of the switches is discussed in Chapter 15.

Chapter 14 discusses interconnection networks. Time division networks are discussed and random assignment time division multiple access (TDMA) is analyzed. Several space division networks are studied: crossbar network, generalized cube network (GCN), banyan network, augmented data manipulator network (ADMN), and improved logical neighborhood network (ILN). For each network, a detailed explanation is provided for how a path is established and, equally important, the packet acceptance probability is derived. This last performance measure will prove essential to analyze the performance of switches.

Chapter 15 discusses modeling techniques for input buffer, output buffer, and shared buffer switches. Equations for the performance of each switch are obtained to describe packet loss probability, average delay within the switch, the throughput, and average queue size.

Chapter 16 discusses the design of two next-generation high-performance network switches. The first Promina 4000 switch developed by N.E.T. Inc. The second is the VRQ switch which was developed at the University of Victoria and is being continually improved. The two designs are superficially similar and a comparative study is reported to show how high-performance impacted the design decisions in each switch.

Appendix A provides a handy reference for many formulas that are useful while modeling the different queues considered here. The reader should find this information handy since it was difficult to find all the formulas in a single source.

Appendix B discusses techniques for solving difference equations or recurrence relations. These recurrence relations crop up in the analysis of queues and Markov chains.

Appendix C discusses how the z-transform technique could be used to find a closed-form expression for the distribution vector $\mathbf{s}(n)$ at any time value through finding the z-transform of the transition matrix $\mathbf{P}$.

Appendix D discusses vectors and matrices. Several concepts are discussed such as matrix inverse, matrix nullspace, rank of a matrix, matrix diagonalization, and eigenvalues and eigenvectors of a matrix. Techniques for solving systems of linear equations are discussed since these systems are encountered in several places in the book. Many special matrices are discussed such as circulant matrix, diagonal matrix, echelon matrix, Hessenberg matrix, identity matrix, nonnegative matrix, orthogonal matrix, plane rotation, stochastic (Markov) matrix, substochastic matrix, and tridiagonal matrix.

Appendix E discusses the use of MATLAB in engineering applications. A brief introduction to MATLAB is provided since it is one of the more common mathematical packages used.

Appendix F discusses design of databases. A database is required in a switch to act as the lookup table for important properties of transmitted packets. Hashing and B-trees are two of the main techniques used to construct the fast routing or lookup tables used in switches and routers. The performance of the hashing function and average lookup delay are analyzed. The B-tree data structure is discussed and the advantages of B-trees over regular binary trees and multiway trees are explained.

Advanced Topics

I invested special effort in making this book useful to practicing engineers and students. There are many interesting examples and models throughout the book. However, I list here some interesting topics:

- Chapter 1 discusses heavy-tailed distribution in Section 1.20 and generation of random numbers in Section 1.35.
- Chapter 3 discusses techniques for finding higher powers for Markov chain state transition matrix in Sections 3.13 and 3.14.
- Chapter 5 discusses reducible Markov chains at steady state in Section 5.7 and transient analysis of reducible Markov chains in Section 5.6. Also, there is a discussion on how to identify a reducible Markov chain by examining its eigenvalues and eigenvectors.
- Chapter 6 discusses transient analysis of periodic Markov chains in Section 6.15 and asymptotic behavior of periodic Markov chains in Section 6.15. Also, there is a discussion on how to identify a periodic Markov chain and how to determine its period by examining its eigenvalues.
- Chapter 7 discusses developing performance metrics for the major queue types.
- Chapter 8 discusses how to model three flow control protocols dealing with traffic management.
- Chapter 9 discusses how to model three flow control protocols dealing with error control.
- Chapter 10 discusses how to model three flow control protocols dealing with medium access control.
- Chapter 11 discusses developing realistic models for source traffic using Poisson description (Section 11.3.2), Bernoulli (Section 11.4.3), and Pareto traffic (Section 11.8). There is also discussion on packet destination and length modeling.
- Chapter 12 discusses 12 scheduling algorithms and provides Markov chain analysis for many of them.
- Chapter 13 discusses seven types of switches based on their buffering strategies and the advantages and disadvantages of each choice.
- Chapter 14 discusses many types of interconnection networks and also provides, for the first time, analysis of the performance of each network.

Web Resource

The website *http://www.ece.uvic.ca/∼fayez/Book*, www.springer.com/978-0-387-74437-7 contains information about the textbook and any related web resources.

Errors

This book covers a wide range of topics related to communication networks and provides an extensive set of analyses and worked examples. It is "highly probable" that it contains errors and omissions. Other researchers and/or practicing engineers might have other ideas about the content and organization of this book. We welcome receiving any constructive comments and suggestions for inclusion in the next edition. If you find any errors, we would appreciate hearing from you. We also welcome ideas for examples and problems (along with their solutions if possible) to include in the next edition with proper citation.

You can send your comments and bug reports electronically to fayez@uvic.ca, or you can fax or mail the information to

Dr. Fayez Gebali
Elec. and Comp. Eng. Dept.
University of Victoria Victoria, B.C., Canada V8W 3P6
Tel: (250) 721-6509
Fax: (250)721-6052.

Chapter 2
Random Processes

2.1 Introduction

We saw in Section 1.11 on page 10 that many systems are best studied using the concept of random variables where the outcome of a random experiment was associated with some numerical value. Next, we saw in Section 1.27 on page 30 that many more systems are best studied using the the concept of multiple random variables where the outcome of a random experiment was associated with multiple numerical values. Here we study *random processes* where the outcome of a random experiment is associated with a *function of time* [1]. Random processes are also called *stochastic processes*. For example, we might study the output of a digital filter being fed by some random signal. In that case, the filter output is described by observing the output waveform at random times.

Thus a *random process* assigns a random *function of time* as the outcome of a random experiment. Figure 2.1 graphically shows the sequence of events leading to assigning a function of time to the outcome of a random experiment. First we run the experiment, then we observe the resulting outcome. Each outcome is associated with a time function $x(t)$.

A random process $X(t)$ is described by

- the *sample space* S which includes all possible outcomes s of a random experiment
- the *sample function* $x(t)$ which is the time function associated with an outcome s. The values of the sample function could be discrete or continuous
- the *ensemble* which is the set of all possible time functions produced by the random experiment
- the time parameter t which could be continuous or discrete
- the statistical dependencies among the random processes $X(t)$ when t is changed.

Based on the above descriptions, we could have four different types of random processes:

1. Discrete time, discrete value: We measure time at discrete values $t = nT$ with $n = 0, 1, 2, \ldots$. As an example, at each value of n we could observe the number of cars on the road $x(n)$. In that case, $x(n)$ is an integer between 0 and 10, say.

F. Gebali, *Analysis of Computer and Communication Networks*,
DOI: 10.1007/978-0-387-74437-7_2,

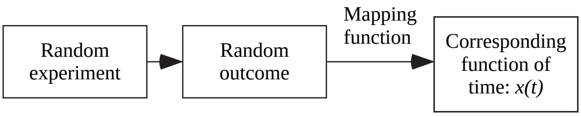

Fig. 2.1 The sequence of events leading to assigning a time function $x(t)$ to the outcome of a random experiment

Each time we perform this experiment, we would get a totally different sequence for $x(n)$.

2. Discrete time, continuous value: We measure time at discrete values $t = nT$ with $n = 0, 1, 2, \ldots$. As an example, at each value of n we measure the outside temperature $x(n)$. In that case, $x(n)$ is a real number between -30° and $+45^\circ$, say. Each time we perform this experiment, we would get a totally different sequence for $x(n)$.
3. Continuous time, discrete value: We measure time as a continuous variable t. As an example, at each value of t we store an 8-bit digitized version of a recorded voice waveform $x(t)$. In that case, $x(t)$ is a binary number between 0 and 255, say. Each time we perform this experiment, we would get a totally different sequence for $x(t)$.
4. Continuous time, continuous value: We measure time as a continuous variable t. As an example, at each value of t we record a voice waveform $x(t)$. In that case, $x(t)$ is a real number between 0 V and 5 V, say. Each time we perform this experiment, we would get a totally different sequence for $x(t)$.

Figure 2.2 shows a discrete time, discrete value random process for an observation of 10 samples where only three random functions are generated. We find that for $n = 2$, the values of the functions correspond to the random variable $X(2)$.

Therefore, random processes give rise to random variables when the time value t or n is fixed. This is equivalent to sampling all the random functions at the specified time value, which is equivalent to taking a vertical slice from all the functions shown in Fig. 2.2.

Example 1 A time function is generated by throwing a die in three consecutive throws and observing the number on the top face after each throw. Classify this random process and estimate how many sample functions are possible.

This is a discrete time, discrete value process. Each sample function will be have three samples and each sample value will be from the set of integers 1 to 6. For example, one sample function might be 4, 2, 5. Using the multiplication principle for probability, the total number of possible outputs is $6^3 = 216$. ■

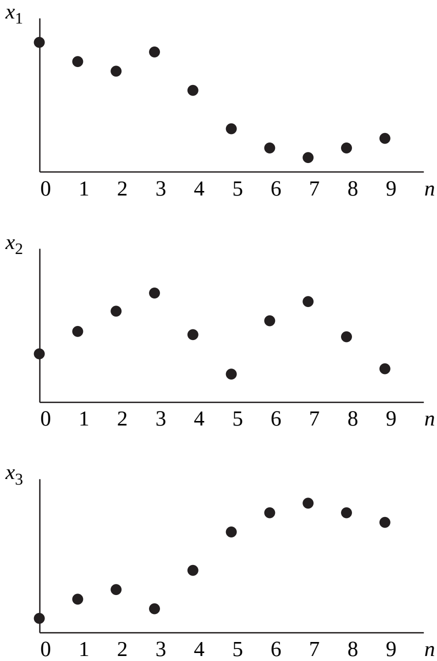

Fig. 2.2 An example of a discrete time, discrete value random process for an observation of 10 samples where only three random functions are possible.

2.2 Notation

We use the notation $X(t)$ to denote a continuous-time random process and also to denote the random variable measured at time t. When $X(t)$ is continuous, it will have a pdf $f_X(x)$ such that the probability that $x \leq X \leq x + \varepsilon$ is given by

$$p(X = x) = f_X(x)\,dx \tag{2.1}$$

When $X(t)$ is discrete, it will have a pmf $p_X(x)$ such that the probability that $X = x$ is given by

$$p(X = x) = p_X(x) \tag{2.2}$$

Likewise, we use the notation $X(n)$ to denote a discrete-time random process and also to denote the random variable measured at time n. That random variable is statistically described by a pdf $f_X(x)$ when it is continuous, or it is described by a pmf $p_X(x)$ when it is discrete.

2.3 Poisson Process

We shall encounter Poisson processes when we describe communication traffic. A Poisson process is a stochastic process in which the number of events occurring in a given period of time depends only on the length of the time period [2]. This number of events k is represented as a random variable K that has a Poisson distribution given by

$$p(k) = \frac{(\lambda\, t)^k e^{-\lambda\, t}}{k!} \tag{2.3}$$

where $\lambda > 0$ is a constant representing the rate of arrival of the events and t is the length of observation time.

2.4 Exponential Process

The exponential process is related to the Poisson process. The exponential process is used to model the interarrival time between occurrence of random events. Examples that lead to an interarrival time include the time between bus arrivals at a bus stop, the time between failures of a certain component, and the time between packet arrival at the input of a router.

The random variable T could be used to describe the interarrival time. The probability that the interarrival time lies in the range $t \leq T \leq t + dt$ is given by

$$\lambda e^{-\lambda\, t}\, dt \tag{2.4}$$

where λ is the average rate of the event under consideration.

2.5 Deterministic and Nondeterministic Processes

A *deterministic* process is one where future values of the sample function are known if the present value is known. An example of a deterministic process is the modulation technique known as quadrature amplitude modulation (QAM) for transmitting groups of binary data. The transmitted analog waveform is given by

$$v(t) = a \cos(\omega t + \phi) \tag{2.5}$$

where the signal amplitude a and phase angle ϕ change their value depending on the bit pattern that has been received. The analog signal is transmitted for the time period $0 \leq t < T_0$. Since the arriving bit pattern is random, the values of the corresponding two parameters a and ϕ are random. However, once a and ϕ are determined, we would be able to predict the shape of the resulting waveform.

A *nondeterministic* random process is one where future values of the sample function cannot be known if the present value is known. An example of a nondeterministic random process is counting the number of packets that arrive at the input of a switch every one second and this observation is repeated for a certain time. We would not be able to predict the pattern even if we know the present number of arriving packets.

2.6 Ensemble Average

The random variable $X(n_1)$ represents all the possible values x obtained when time is frozen at the value n_1. In a sense, we are sampling the ensemble of random functions at this time value.

The expected value of $X(n_1)$ is called the *ensemble average* or statistical average $\mu(n_1)$ of the random process at n_1. The ensemble average is expressed as

$$\mu_X(t) = E[X(t)] \qquad \text{continuous-time process} \tag{2.6}$$

$$\mu_X(n) = E[X(n)] \qquad \text{discrete-time process} \tag{2.7}$$

The ensemble average could itself be another random variable since its value could change at random with our choice of the time value t or n.

Example 2 The modulation scheme known as frequency-shift keying (FSK) can be modeled as a random process described by

$$X(t) = a \cos \omega t$$

where a is a constant and ω corresponds to the random variable Ω that can have one of two possible values ω_1 and ω_2 that correspond to the input bit being 0 or 1, respectively. Assuming that the two frequencies are equally likely, find the expected value $\mu(t)$ of this process.

Our random variable Ω is discrete with probability 0.5 when $\Omega = \omega_1$ or $\Omega = \omega_2$. The expected value for $X(t)$ is given by

$$\begin{aligned} E[X(t)] &= 0.5\, a \cos \omega_1 t + 0.5\, a \cos \omega_2 t \\ &= a \cos\left[\frac{(\omega_1 + \omega_2)\, t}{2}\right] \times \cos\left[\frac{(\omega_1 - \omega_2)\, t}{2}\right] \end{aligned}$$

■

Example 3 The modulation scheme known as pulse amplitude modulation (PAM) can be modeled as a random process described by

$$X(n) = \sum_{i=0}^{\infty} g(n)\, \delta(n-i)$$

where $g(n)$ is the amplitude of the input signal at time n. $g(n)$ corresponds to the random variable G that is uniformly distributed in the range 0–A. Find the expected value $\mu(t)$ of this process.

This is a discrete time, continuous value random process. Our random variable G is continuous and the expected value for $X(n)$ is given by

$$\begin{aligned} E\left[X(n)\right] &= \frac{1}{A} \int_0^A g\, dg \\ &= \frac{A}{2} \end{aligned}$$

■

2.7 Time Average

Figure 2.2 helps us find the *time average* of the random process. The time average is obtained by finding the average value for *one* sample function such as $X_1(n)$ in the figure. The time average is expressed as

$$\overline{X} = \tfrac{1}{T} \int_0^T X(t)\, dt \qquad \text{continuous-time process} \tag{2.8}$$

$$\overline{X} = \tfrac{1}{N} \sum_0^{N-1} X(n)] \qquad \text{discrete-time process} \tag{2.9}$$

In either case we assumed we sampled the function for a period T or we observed N samples. The time average $\overline{X}$ could itself be a random variable since its value could change with our choice of the random function under consideration.

2.8 Autocorrelation Function

Assume a discrete-time random process $X(n)$ which produces two random variables $X_1 = X(n_1)$ and $X_2 = X(n_2)$ at times n_1 and n_2 respectively. The *autocorrelation function* for these two random variables is defined by the following equation:

$$r_{XX}(n_1, n_2) = E\left[X_1\ X_2\right] \tag{2.10}$$

In other words, we consider the two random variables X_1 and X_2 obtained from the same random process at the two different time instances n_1 and n_2.

Example 4 Find the autocorrelation function for a second-order finite-impulse response (FIR) digital filter, sometimes called moving average (MA) filter, whose output is given by the equation

$$y(n) = a_0 x(n) + a_1 x(n-1) + a_2 x(n-2) \tag{2.11}$$

where the input samples $x(n)$ are assumed to be zero mean independent and identically distributed (iid) random variables.

We assign the random variable Y_n to correspond to output sample $y(n)$ and X_n to correspond to input sample $x(n)$. Thus we can have the following autocorrelation function

$$r_{YY}(0) = E\left[Y_n Y_n\right] = a_0^2 E\left[X_0^2\right] + a_1^2 E\left[X_1^2\right] + a_2^2 E\left[X_2^2\right] E\left[X_0^2\right] \tag{2.12}$$

$$= \left(a_0^2 + a_1^2 + a_2^2\right)\sigma^2 \tag{2.13}$$

Similarly, we can write

$$r_{YY}(1) = E\left(Y_n Y_{n+1}\right) = 2a_0 a_1\, \sigma^2 \tag{2.14}$$

$$r_{YY}\,(2) = E\left(Y_n Y_{n+2}\right) = a_0 a_2\, \sigma^2 \tag{2.15}$$

$$r_{YY}\,(k) = 0; \qquad k > 2 \tag{2.16}$$

where σ^2 is the input sample variance. Figure 2.3 shows a plot of the autocorrelation assuming all the filter coefficients are equal. ■

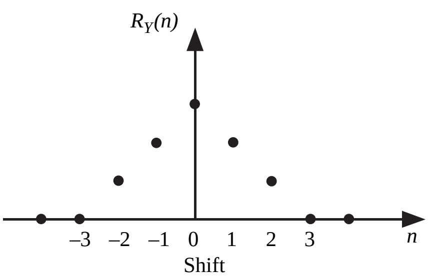

Fig. 2.3 Autocorrelation function of a second-order digital filter whose input is uncorrelated samples

2.9 Stationary Processes

A **wide-sense stationary** random process has the following two properties [3]

$$E\left[X(t)\right] = \mu = \text{ constant} \tag{2.17}$$

$$E\left[X(t)\ X(t+\tau)\right] = r_{XX}\left(t, t+\tau\right) = r_{XX}(\tau) \tag{2.18}$$

Such a process has a constant expected value and the autocorrelation function depends on the time difference between the two random variables.

The above equations apply to a continuous time random process. For a discrete-time random process, the equations for a wide-sense stationary random process become

$$E\left[X(n)\right] = \mu = \text{ constant} \tag{2.19}$$

$$E\left[X\left(n_1\right)\ X\left(n_1+n\right)\right] = r_{XX}\left(n_1, n_1+n\right) = r_{XX}(n) \tag{2.20}$$

The autocorrelation function for a wide-sense stationary random process exhibits the following properties [1].

$$r_{XX}(0) = E\left[X^2(n)\right] \geq 0 \tag{2.21}$$

$$|r_{XX}(n)| \leq r_{XX}(0) \tag{2.22}$$

$$r_{XX}(-n) = r_{XX}(n) \qquad \text{even symmetry} \tag{2.23}$$

A stationary random process is *ergodic* if all time averages equal their corresponding statistical averages [3]. Thus if $X(n)$ is an ergodic random process, then we could write

$$\overline{X} = \mu \tag{2.24}$$

$$\overline{X^2} = r_{XX}(0) \tag{2.25}$$

Example 5 The modulation scheme known as phase-shift keying (PSK) can be modeled as a random process described by

$$X(t) = a\cos(\omega t + \phi)$$

where a and ω are constant and ϕ corresponds to the random variable Φ with two values 0 and π which are equally likely. Find the autocorrelation function $r_{XX}(t)$ of this process.

The phase pmf is given by

$$p(0) = 0.5$$
$$p(\pi) = 0.5$$

The autocorrelation is found as

$$\begin{aligned} r_{XX}(\tau) &= E\left[a \cos(\omega t + \Phi)\, a \cos(\omega t + \omega\tau + \Phi)\right] \\ &= 0.5\, a^2 \; \cos(\omega\tau)\; E\left[\cos(2\omega t + \omega\tau + 2\Phi)\right] \\ &= 0.5\, a^2 \; \cos(\omega\tau)\; \cos(2\omega t + \omega\tau) \end{aligned}$$

We notice that this process is not wide-sense stationary since the autocorrelation function depends on t. ■

2.10 Cross-Correlation Function

Assume two discrete-time random processes $X(n)$ and $Y(n)$ which produce two random variables $X_1 = X(n_1)$ and $Y_2 = Y(n_2)$ at times n_1 and n_2, respectively. The *cross-correlation function* is defined by the following equation.

$$r_{XY}(n_1, n_2) = E\left[X_1\, Y_2\right] \tag{2.26}$$

If the cross-correlation function is zero, i.e. $r_{XY} = 0$, then we say that the two processes are *orthogonal*. If the two processes are *statistically independent*, then we have

$$r_{XY}(n_1, n_2) = E\left[X(n_1)\right] \times E\left[Y(n_2)\right] \tag{2.27}$$

Example 6 Find the cross-correlation function for the two random processes

$$X(t) = a \cos \omega t$$
$$Y(t) = b \sin \omega t$$

where a and b are two independent and identically distributed random variables with mean μ and variance σ^2.

The cross-correlation function is given by

$$\begin{aligned} r_{XY}(t, t+\tau) &= E\left[a \cos \omega t \; b \sin(\omega t + \omega\tau)\right] \\ &= 0.5[\sin \omega\tau + \sin(2\omega t + \omega\tau)]\; E[a]\; E[b] \\ &= 0.5\, \mu^2 \left[\sin \omega\tau + \sin(2\omega t + \omega\tau)\right] \end{aligned}$$

■

2.11 Covariance Function

Assume a discrete-time random process $X(n)$ which produces two random variables $X_1 = X(n_1)$ and $X_2 = X(n_2)$ at times n_1 and n_2, respectively. The *autocovariance function* is defined by the following equation:

$$c_{XX}(n_1, n_2) = E\left[(X_1 - \mu_1)(X_2 - \mu_2)\right] \tag{2.28}$$

The autocovariance function is related to the autocorrelation function by the following equation:

$$c_{XX}(n_1, n_2) = r_X(n_1, n_2) - \mu_1 \mu_2 \tag{2.29}$$

For a wide-sense stationary process, the autocovariance function depends on the difference between the time indices $n = n_2 - n_1$.

$$c_{XX}(n) = E\left[(X_1 - \mu)(X_2 - \mu)\right] = r_{XX}(n) - \mu^2 \tag{2.30}$$

Example 7 Find the autocovariance function for the random process $X(t)$ given by

$$X(t) = a + b\cos\omega t$$

where ω is a constant and a and b are iid random variables with zero mean and variance σ^2.

We have

$$\begin{aligned}
c_{XX} &= E\{(A + B\cos\omega t)[A + B\cos\omega(t+\tau)]\} \\
&= E\left[a^2\right] + E[ab]\,\left[\cos\omega t + \cos\omega(t+\tau)\right] + E\left[b^2\right]\cos^2\omega(t+\tau) \\
&= \sigma^2 + E[a]\,E[b]\,\left[\cos\omega t + \cos\omega(t+\tau)\right] + \sigma^2\cos^2\omega(t+\tau) \\
&= \sigma^2\left[1 + \cos^2\omega(t+\tau)\right]
\end{aligned}$$

■

The *cross-covariance function* for two random processes $X(n)$ and $Y(n)$ is defined by

$$\begin{aligned}
c_{XY}(n) &= E\left[(X(n_1) - \mu_X)(Y(n_1 + n) - \mu_Y)\right] \\
&= r_{XY}(n) - \mu_X \mu_Y
\end{aligned} \tag{2.31}$$

Two random processes are called *uncorrelated* when their cross-covariance function vanishes.

$$c_{XY}(n) = 0 \tag{2.32}$$

Example 8 Find the cross-covariance function for the two random processes $X(t)$ and $Y(t)$ given by

$$X(t) = a + b\cos\omega t$$
$$Y(t) = a + b\sin\omega t$$

where ω is a constant and a and b are iid random variables with zero mean and variance σ^2.

We have

$$\begin{aligned}
c_{XY}(n) &= E\left\{(A + B\cos\omega t)[A + B\sin\omega(t+\tau)]\right\} \\
&= E\left[A^2\right] + E[AB]\ [\cos\omega t + \sin\omega(t+\tau)] + E\left[B^2\right]\cos\omega t\ \sin\omega(t+\tau) \\
&= \sigma^2 + E[A]\ E[B]\ [\cos\omega t + \sin\omega(t+\tau)] + \sigma^2\cos\omega t\ \sin\omega(t+\tau) \\
&= \sigma^2\,[1 + \cos\omega t\ \sin\omega(t+\tau)]
\end{aligned}$$

■

2.12 Correlation Matrix

Assume we have a discrete-time random process $X(n)$. At each time step i we define the random variable $X_i = X(i)$. If each sample function contains n components, it is convenient to construct a vector representing all these random variables in the form

$$\mathbf{x} = \begin{bmatrix} X_1 & X_2 & \cdots & X_n \end{bmatrix}^t \tag{2.33}$$

Now we would like to study the correlation between each random variable X_i and all the other random variables. This would give us a comprehensive understanding of the random process. The best way to do that is to construct a *correlation matrix*.

We define the $n \times n$ correlation matrix $\mathbf{R}_X$, which gives the correlation between all possible pairs of the random variables as

$$\mathbf{R}_X = E\left[\mathbf{x}\,\mathbf{x}^t\right] = E\begin{bmatrix} X_1X_1 & X_1X_2 & \cdots & X_1X_n \\ X_2X_1 & X_2X_2 & \cdots & X_2X_n \\ \vdots & \vdots & \ddots & \vdots \\ X_nX_1 & X_nX_2 & \cdots & X_nX_n \end{bmatrix} \tag{2.34}$$

We can express $\mathbf{R}_X$ in terms of the individual correlation functions

$$\mathbf{R}_X = \begin{bmatrix} r_{XX}(1,1) & r_{XX}(1,2) & \cdots & r_{XX}(1,n) \\ r_{XX}(1,2) & r_{XX}(2,2) & \cdots & r_{XX}(2,n) \\ \vdots & \vdots & \ddots & \vdots \\ r_{XX}(1,n) & r_{XX}(2,n) & \cdots & r_{XX}(n,n) \end{bmatrix} \tag{2.35}$$

Thus we see that the correlation matrix is symmetric. For a wide-sense stationary process, the correlation functions depend only on the difference in times and we get an even simpler matrix structure:

$$\mathbf{R}_X = \begin{bmatrix} r_{XX}(0) & r_{XX}(1) & \cdots & r_{XX}(n-1) \\ r_{XX}(1) & r_{XX}(0) & \cdots & r_{XX}(n-2) \\ \vdots & \vdots & \ddots & \vdots \\ r_{XX}(n-1) & r_{XX}(n-2) & \cdots & r_{XX}(0) \end{bmatrix} \tag{2.36}$$

Each diagonal in this matrix has identical elements and our correlation matrix becomes a *Toeplitz matrix*.

Example 9 Assume the autocorrelation function for a stationary random process is given by

$$r_{XX}(\tau) = 5 + 3e^{-|\tau|}$$

Find the autocorrelation matrix for $\tau = 0$, 1, and 2.

The autocorrelation matrix is given by

$$\mathbf{R}_{XX} = \begin{bmatrix} 8 & 6.1036 & 5.4060 \\ 6.1036 & 8 & 6.1036 \\ 5.4060 & 6.1036 & 6 \end{bmatrix}$$

■

2.13 Covariance Matrix

In a similar fashion, we can define the *covariance matrix* for many random variables obtained from the same random process as

$$\mathbf{C}_{XX} = E\left[(\mathbf{x} - \overline{\mu})(\mathbf{x} - \overline{\mu})^t\right] \tag{2.37}$$

where $\overline{\mu} = \left[\, \mu_1 \ \mu_2 \ \cdots \ \mu_n \,\right]^t$ is the vector whose components are the expected values of our random variables. Expanding the above equation we can write

$$\mathbf{C}_{XX} = E\left[\mathbf{XX^t}\right] - \overline{\mu}\,\overline{\mu}^t \tag{2.38}$$
$$= \mathbf{R}_X - \overline{\mu}\,\overline{\mu}^t \tag{2.39}$$

When the process has zero mean, the covariance matrix equals the correlation matrix:

$$\mathbf{C}_{XX} = \mathbf{R}_{XX} \tag{2.40}$$

The covariance matrix can be written explicitly in the form

$$\mathbf{C}_{XX} = \begin{bmatrix} C_{XX}(1,1) & C_{XX}(1,2) & \cdots & C_{XX}(1,n) \\ C_{XX}(1,2) & C_{XX}(2,2) & \cdots & C_{XX}(2,n) \\ \vdots & \vdots & \ddots & \vdots \\ C_{XX}(1,n) & C_{XX}(2,n) & \cdots & C_{XX}(n,n) \end{bmatrix} \tag{2.41}$$

Thus we see that the covariance matrix is symmetric. For a wide-sense stationary process, the covariance functions depend only on the difference in times and we get an even simpler matrix structure:

$$\mathbf{C}_{XX} = \begin{bmatrix} C_{XX}(0) & C_{XX}(1) & \cdots & C_{XX}(n-1) \\ C_{XX}(1) & C_{XX}(0) & \cdots & C_{XX}(n-2) \\ \vdots & \vdots & \ddots & \vdots \\ C_{XX}(n-1) & C_{XX}(n-2) & \cdots & C_{XX}(0) \end{bmatrix} \tag{2.42}$$

Using the definition for covariance in (1.114) on page 35, we can write the above equation as

$$\mathbf{C}_{XX} = \sigma_X^2 \begin{bmatrix} 1 & \rho(1) & \rho(2) & \cdots & \rho(n-1) \\ \rho(1) & 1 & \rho(1) & \cdots & \rho(n-2) \\ \rho(2) & \rho(1) & 1 & \cdots & \rho(n-3) \\ \vdots & & & & \vdots \\ \rho(n-1) & \rho(n-2) & \rho(n-3) & \cdots & 1 \end{bmatrix} \tag{2.43}$$

Example 10 Assume the autocovariance function for a wide-sense stationary random process is given by

$$c_{XX}(\tau) = 5 + 3e^{-|\tau|}$$

Find the autocovariance matrix for $\tau = 0$, 1, and 2.

Since the process is wide-sense stationary, the variance is given by

$$\sigma^2 = c_{XX}(0) = 8$$

The autocovariance matrix is given by

$$\mathbf{C}_{XX} = 8 \begin{bmatrix} 1 & 0.7630 & 0.6758 \\ 0.7630 & 1 & 0.7630 \\ 0.6758 & 0.7630 & 1 \end{bmatrix}$$

■

Problems

2.1 Define deterministic and nondeterministic processes. Give an example for each type.

2.2 Let X be the random process corresponding to observing the noon temperature throughout the year. The number of sample functions are 365 corresponding to each day of the year. Classify this process.

2.3 Let X be the random process corresponding to reporting the number of defective lights reported in a building over a period of one month. Each month we would get a different pattern. Classify this process.

2.4 Let X be the random process corresponding to measuring the total tonnage (weight) of ships going through the Suez canal in one day. The data is plotted for a period of one year. Each year will produce a different pattern. Classify this process.

2.5 Let X be the random process corresponding to observing the number of cars crossing a busy intersection in one hour. The number of sample functions are 24 corresponding to each hour of the day. Classify this process.

2.6 Let X be the random process corresponding to observing the bit pattern in an Internet packet. Classify this process.

2.7 Amplitude-shift keying (ASK) can be modeled as a random process described by

$$X(t) = a \cos \omega t$$

where ω is constant and a corresponds to the random variable A with two values a_0 and a_1 which occur with equal probability. Find the expected value $\mu(t)$ of this process.

2.8 A modified ASK uses two bits of the incoming data to generate a sinusoidal waveform and the corresponding random process is described by

$$X(t) = a \cos \omega t$$

where ω is a constant and a is a random variable with four values a_0, a_1, a_2, and a_3. Assuming that the four possible bit patterns are equally likely find the expected value $\mu(t)$ of this process.

2.9 Phase-shift keying (PSK) can be modeled as a random process described by

$$X(t) = a\cos(\omega t + \phi)$$

where a and ω are constant and ϕ corresponds to the random variable Φ with two values 0 and π which occur with equal probability. Find the expected value $\mu(t)$ of this process.

2.10 A modified PSK uses two bits of the incoming data to generate a sinusoidal waveform and the corresponding random process is described by

$$X(t) = a\cos(\omega t + \phi)$$

where a and ω are constants and ϕ is a random variable Φ with four values $\pi/4$, $3\pi/4$, $5\pi/4$, and $7\pi/4$ [4]. Assuming that the four possible bit patterns occur with equal probability, find the expected value $\mu(t)$ of this process.

2.11 A modified frequency-shift keying (FSK) uses three bits of the incoming data to generate a sinusoidal waveform and the random process is described by

$$X(t) = a\cos\omega t$$

where a is a constant and ω corresponds to the random variable Ω with eight values $\omega_0, \omega_1, \ldots, \omega_7$. Assuming that the eight frequencies are equally likely, find the expected value $\mu(t)$ of this process.

2.12 A discrete-time random process $X(n)$ produces the random variable $X(n)$ given by

$$X(n) = a^n$$

where a is a uniformly distributed random variable in the range 0–1. Find the expected value for this random variable at any time instant n.

2.13 Define a wide-sense stationary random process.

2.14 Prove (2.23) on page 56.

2.15 Define an ergodic random process.

2.16 Explain which of the following functions represent a valid autocorrelation function.

$$r_{XX}(n) = a^n \quad 0 \le a < 1 \qquad r_{XX}(n) = |a|^n \quad 0 \le a < 1$$
$$r_{XX}(n) = a^{n^2} \quad 0 \le a < 1 \qquad r_{XX}(n) = |a|^{n^2} \quad 0 \le a < 1$$
$$r_{XX}(n) = \cos n \qquad r_{XX}(n) = \sin n$$

2.17 A random process described by

$$X(t) = a\cos(\omega t + \phi)$$

where a and ω are constant and ϕ corresponds to the random variable Φ which is uniformly distributed in the interval 0 to 2π. Find the autocorrelation function $r_{XX}(t)$ of this process.

2.18 Define what is meant by two random processes being orthogonal.

2.19 Define what is meant by two random processes being statistically independent.

2.20 Find the cross-correlation function for the following two random processes.

$$X(t) = a \cos \omega t$$
$$Y(t) = \alpha\, a \cos(\omega t + \theta)$$

where a and θ are two zero mean random variables and α is a constant.

2.21 Given two random processes X and Y, when are they uncorrelated?

References

1. R. D. Yates and D. J. Goodman, *Probability and Stochastic Processes*, John Wiley, New York, 1999.
2. Thesaurus of Mathematics. http://thesaurus.maths.org/dictionary/map/word/1656.
3. P. Z. Peebles, *Probability, Random Variables, and Random Signal Principles*, McGraw-Hill, New York, 1993.
4. S. Haykin, *An Introduction to Analog and Digital Communications*, John Wiley, New York, 1989.