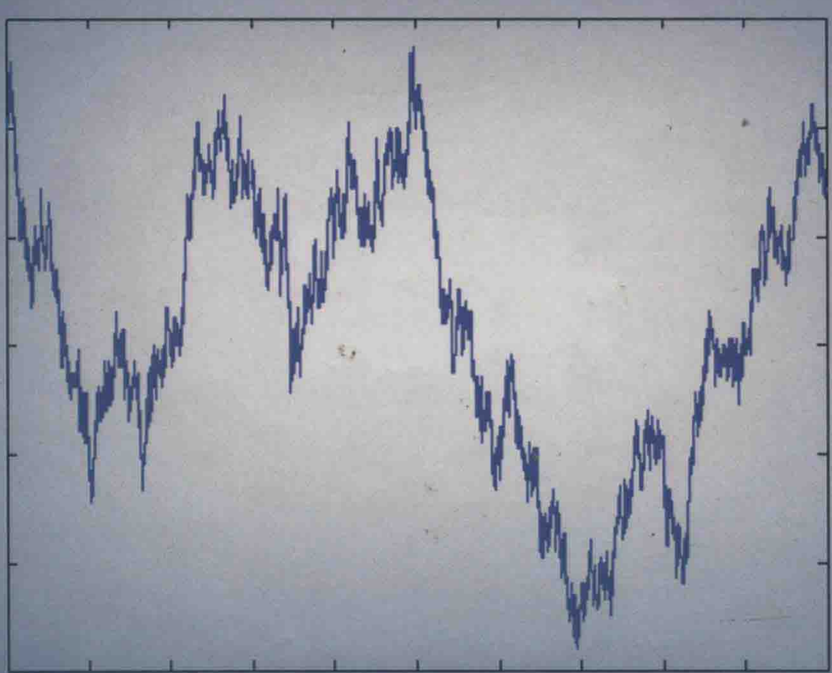
Introduction to Probability Theory and Stochastic Processes

John Chiasson



WILEY

Introduction to Probability Theory and Stochastic Processes

John Chiasson
Boise State University
Boise, Idaho

常州大学图书馆
藏书章

WILEY

Copyright © 2013 by John Wiley & Sons, Inc. All rights reserved.

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470, or on the web at www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at http://www.wiley.com/go/permission.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representation or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print, however, may not be available in electronic formats. For more information about Wiley products, visit our web site at www.wiley.com.

Library of Congress Cataloging-in-Publication Data:

Chiasson, John Nelson.
Introduction to probability theory and stochastic processes / John Chiasson, Boise State University, Idaho.
pages cm
Includes bibliographical references and index.
ISBN 978-1-118-38279-0 (pbk.)
1. Probabilities. 2. Stochastic processes. I. Title.
QA273.C547 2013
519.2—dc23 2012047234

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

Introduction to Probability Theory and Stochastic Processes

试读结束：需要全本请在线购买： www.ertongbook.com

To My Mother

Preface

This book is an introductory course on *Probability Theory and Stochastic Processes* for first-year graduate students in Engineering or perhaps applied students in Mathematics. There are several such books in print today including Papoulis and Pillai [1], Fine [2], Gubner [3], Leon-Garcia [4], Stark and Woods [5] as well as the advanced undergraduate books by Bertsekas and Tsitsiklis [6], and Helstrom [7]. The obvious question is, "Why another such book?". Well, the "Holy Grail" of teaching probability is to explain what it is all about without getting bogged down in measure theory. The attempt here is to do this with an approach based on the binary expansion of real numbers in the unit interval as a model of tossing a fair coin an infinite number of times [8][9][10]. Such an approach provides a means to explicitly construct the stochastic processes (random variables) used in the modeling of the typical random phenomena encountered in engineering and science. As a result, the student understands why the mathematical structure and notation of probability is setup the way it is. By structure is simply meant that there is a single underlying probability space on which all the random variables are defined (done via the binary expansion model). However, it must be emphasized that the presentation here is *not* a measure-theoretic approach to probability! Rather, we explain and emphasize that in applying probability theory, the user specifies the *probability distributions* F_{X_i} of the random variables X_i (not the abstract functions $X_i(\omega)$ for $\omega \in \Omega$, where Ω is the abstract sample space) in order to model random phenomena. The standard (and correct) definition of a random variable as a function $X_i : \Omega \to \mathbb{R}$ gives no insight into why they are useful for modeling random phenomena. In the author's opinion, this is especially confusing when modeling random phenomena with an uncountable number of outcomes (continuous random variables). In contrast, here a single underlying probability space is constructed with a uniform probability function P along with a sequence of independent random variables (functions) $X_i : \Omega \to \mathbb{R}$ such that $F_{X_i}(x) = P(\{\omega \in \Omega |\ X_i(\omega) \leq x\})$.

This basic structure of probability theory for uncountable sample spaces is developed in Chapters 4, 5, 6, and 7. The half-open interval $(0, 1]$ is taken as the underlying probability space with the uniform distribution as the probability function. Each $\omega \in \Omega = (0, 1]$ is written as a binary expansion, i.e., $\omega = 0.\omega_1\omega_2\omega_3...$, where $\omega_i = 0$ or 1 is the i^{th} term in the binary expansion of ω. The random variable $X_i : \Omega \to \mathbb{R}$ is defined by $X_i(\omega) = \omega_i$, where it is shown that these random variables (functions) are *independent* and

$$P(\{\omega \in \Omega | X_i = 0\}) = P(\{\omega \in \Omega | X_i = 1\}) = 1/2.$$

This is an explicit construction of a stochastic process modeling the tossing of a fair coin an infinite number of times. These coin tossing random variables X_i are then used to construct an infinite sequence U_i of *independent uniformly distributed* random variables all defined on $\Omega = (0, 1]$. Finally, using these uniformly distributed random variables U_i, it is shown how to construct an infinite sequence of independent random variables with a Gaussian, exponential, or any other desired distribution. The point of going through this explicit construction is to give the reader a concrete view of sequences of independent random variables. More importantly, these constructions are then used to drive home the point that the explicit functional dependence of the random variables $X_i(\omega)$ on the outcome ω is of *no* importance in modeling random phenomena; rather in applications one (typically) wants a sequence of random variables that are independent and of a given distribution to model a particular random phenomena. This modeling issue is hopefully made clear by the end of Chapter 7.

The Polish mathematician Mark (Marek) Kac pointed out in his book [10] (see pages 10,11) that the early development of probability was not embraced by mathematicians because it was not about numbers. Specifically, tossing a coin an infinite number of times was represented as an infinite sequence of Hs and Ts with the important probabilistic notion being that any toss is independent of any other toss. However, as Professor Kac explained, most mathematicians remained "aloof" because it was not clear "what the objects were to which the formalism [independence and probability] was applicable". Then in 1909 Émile Borel showed that the binary expansion digits of the numbers in $(0, 1]$ were independent in the probabilistic sense and thus mathematicians had these actual real-valued functions from the interval $(0, 1]$ to the set $\{0, 1\}$ to model the tossing of a fair coin an infinite number of times. As Professor Kac went on to say:

> At long last, there were well-defined objects to which probability theory for independent events could be applied without fear of getting involved with coins, events, tosses, and experiments.

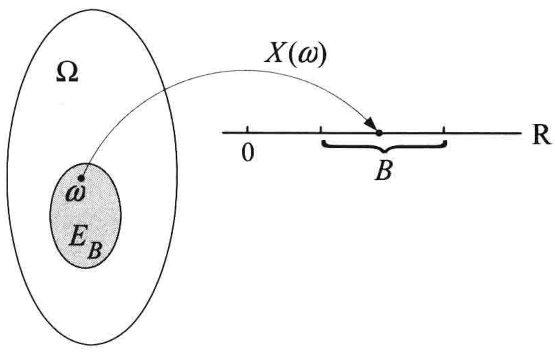
These "well-defined objects" were the binary expansion functions which are now called random variables.[1] As already mentioned, using these "well-defined objects", one is then able to construct an independent sequence of random variables with any prescribed probability distribution. It is hoped

---

[1] Of course, the later development in the 1930s of the use of measure theory by A. N. Kolmogorov as a general setting for probability theory showed that one could get involved with "coins, events, tosses, and experiments" rigorously. See *Foundations of the Theory of Probabilty* by A. N. Kolmogorov, 1933.

that this use of the binary expansion as a tool to explain the mathematical setup of probability theory is as helpful to the reader as it was to the author and, apparently, the mathematical community of the early part of the 20th century!

This book divides the presentation of probability theory according to whether the sample space has a countable number of outcomes or an uncountable number of outcomes. The first three chapters develop probability theory for countable (finite and countably infinite) sample spaces. The mathematics required for countable sample spaces is elementary with the physical meaning of outcomes clear and the notion of random variables straightforward. Chapters 4, 5, 6, and 7 then develop probability theory for sample spaces with an *uncountable* number of outcomes. In this case, typically no physical meaning is attached to the outcomes ω. A standard presentation is to say there is an experiment $\mathcal{H}$ with sample space Ω whose elements are the outcomes ω. This is often indicated pictorially as in the figure below. Here the large ellipse denotes the sample space Ω with the points inside this ellipse denoting the outcomes ω. A random variable $X(\omega)$ is (essentially) a function that maps each outcome to a number on the real line $\mathbb{R}$. An event E_B is a subset of Ω and the outcomes in E_B are mapped to some subset $B \subset \mathbb{R}$ of the real line by the random variable X.



However, the author has always found this explanation unsatisfying. What is this function $X(\omega)$? Apparently one doesn't care about the explicit functional relationship from ω to $X(\omega)$. Why? What is this so-called "experiment $\mathcal{H}$"? Again, this is where the binary expansion comes in to make everything much more concrete. One can take comfort in the words of the mathematician Joseph L. Doob that there is a "wide gap between accepting a definition and taking it seriously". In fact, quoting from an interview of J. L. Doob given by J. Laurie Snell [11][2]:

---

[2] Or go to the webpage *www.dartmouth.edu/~chance/Doob/conversation.html*

Kolmogorov's 1933 monograph on the foundations of (mathematical) probability appeared just when I was desperately trying to find out what the subject was all about. He gave measure theoretic definitions of probability, of random variables and their expectations, and conditional expectations. He also constructed probability measures in infinite dimensional coordinate spaces. Kolmogorov did not state that the set of coordinate variables of such a space constitutes a model for a collection of random variables with given compatible joint distributions, and I am ashamed to say that I completely missed the point of that section of his monograph, only realizing it after I had constructed some infinite dimensional product measures in the course of my own research. Kolmogorov defined a random variable as a measurable function on a probability measure space. But there is a wide gap between accepting a definition and taking it seriously. It was a shock for probabilists to realize that a function is glorified into a random variable as soon as its domain is assigned a probability distribution with respect to which the function is measurable. In a 1934 class discussion of bivariate normal distributions Hotelling remarked that zero correlation of two jointly normally distributed random variables implied independence, but it was not known whether the random variables of an uncorrelated pair were necessarily independent. Of course he understood me at once when I remarked after class that the interval $[0, 2\pi]$ when endowed with Lebesgue measure divided by 2π is a probability measure space, and that on this space the sine and cosine functions are uncorrelated but not independent random variables. He had not digested the idea that a trigonometric function is a random variable relative to any Borel probability measure on its domain. The fact that nonprobabilists commonly denote functions by f, g, and so on whereas probabilists tend to call functions random variables and use the notation x, y and so on at the other end of the alphabet helped to make nonprobabilists suspect that mathematical probability was hocus pocus rather than mathematics. And the fact that probabilists called some integrals "expectations" and used the letters E or M instead of integral signs strengthened the suspicion.

J. L. Doob is pointing out the not so obvious notion of an infinite sequence of random variables on a single probability space (infinite dimensional product measures), the difference between the mathematical definition of a random variable and its actual use (i.e., its "digestion"), and that expectation is simply integration over the underlying probability space. It is precisely

试读结束：需要全本请在线购买： www.ertongbook.com

this type of confusion of the structure of probability theory when one is first learning the subject that has motivated the approach in this textbook. As previously stated, Chapters 4, 5, 6, and 7 construct the probability space $(\Omega = (0,1], \beta_{(0,1]}, P)$ where the sample space Ω is simply the half-open interval $(0,1]$, $\beta_{(0,1]}$ is the Borel sigma field of subsets of $(0,1]$ (which is simply explained to be the set of subsets of $(0,1]$ that one can integrate over), and P is the uniform distribution. An infinite sequence of independent Bernoulli random variables with parameter $p = 1/2$ is constructed from the (nonterminating) binary expansion of points in $(0,1]$. Using these random variables, it is then shown how an infinite sequence of independent uniformly distributed random variables (functions) are constructed on this same probability space. From this it is then shown straightforwardly how to obtain an *independent identically distributed* sequence of random variables with any given probability distribution. After going through this development, it is hoped that the reader realizes that random variables are just functions defined on the probability space; but in modeling random phenomena, it is the distributions of the random variables (Gaussian, exponential, etc.) and their specified properties (e.g., independence, uncorrelatedness, etc.) that are important in applications, not the functional dependence of the random variable $X(\omega)$ on ω. Also, it hopefully clarifies the rather abstract notation[3] $P(\{\omega \in \Omega | x_1 < X(\omega) \le x_2\})$ where $\{\omega \in \Omega | x_1 < X(\omega) \le x_2\}$ is a set in the underlying probability space, P is the probability function on the underlying probability space and the probability of the set is actually computed using the induced probability distribution of the random variable X as $P(\{\omega \in \Omega | x_1 < X(\omega) \le x_2\}) = \int_{x_1}^{x_2} f_X(x)dx$. Further, as shown in Chapter 6, this development allows us to show that the expectation of a random variable X can be computed in the usual way using the distribution of X, i.e., as $E[X] = \int_{-\infty}^{\infty} x f_X(x)dx$ or (conceptually) as integrating over the underlying probability space, i.e., $E[X] = \int_\Omega X dP = \int_0^1 X(\omega)d\omega$.

This explicit construction of the random variables is also used to clarify the statement[4] of the Strong Law of Large Numbers (SLLN). Specifically, the SLLN states that

$$\lim_{n\to\infty} \frac{\sum_{i=1}^n X_i(\omega)}{n} = E[X_i] \quad \text{(almost everywhere)},$$

where the X_i are independent identically distributed random variables. Students are used to the notion of (pointwise) convergence using *explicit* functions from their first calculus course. For example, with $S_n(\omega) = (\sum_{i=1}^n X_i(\omega))/n$, suppose that S_n can written explicitly as (say) $S_n(\omega) =$

---

[3] And the equivalent shorthand notations $P(\{\omega | x_1 < X(\omega) \le x_2\})$, $P(\{x_1 < X \le x_2\})$ and $P(x_1 < X \le x_2)$.

[4] Even a probability space with a *finite* number of outcomes requires the construction of an infinite sequence of independent identically distributed random variables to state the SLLN!

$\mu(1 - e^{-n^2(\omega - 1/n)})$ for $0 < \omega \leq 1$. Then, for all $\omega \in (0,1]$, we have

$$\lim_{n\to\infty} S_n(\omega) = \lim_{n\to\infty} \mu(1 - e^{-n^2(\omega-1/n)}) = \mu.$$

However, as shown by the construction of an independent sequence of (for example) Gaussian random variables, the reader is made well aware that the explicit dependence of $S_n(\omega)$ on the outcome "ω" is *not* known making it impossible to carry out such a limit. Hence it is explained that the SLLN must be proven using only the knowledge that the X_i are independent, that the distribution functions of all the X_i are the same, and that the expected value $E[X_i]$ is finite. The emphasis here is on what the strong law says and what are the issues in proving it. Similar comments hold in regards to the central limit theorem (CLT). It is this explicitly constructed underlying probability space $((0,1], \mathcal{B}_{(0,1]}, P)$ that is exploited to give a more concrete explanation of these abstract things called random variables.

Chapters 8, 9, and 10 present the methods on computing the distribution of functions of random variables as well as the use of transform techniques (characteristic functions and moment generating functions). This is standard material and provides important mathematical tools for dealing with random variables. What is different from the standard pedagogy is that the Poisson process is defined and its probability distribution is derived in Chapter 9. This is simply because the idea of a stochastic process consisting of an infinite sequence of i.i.d. exponential random variables has already been developed in Chapter 7 so that the Poisson process can be defined. Further, using the techniques developed in the early part of Chapter 9, we are also able to derive its probability distribution. Also, the central limit theorem is discussed in Chapter 9.

Chapter 11 develops the notion of conditional probability for jointly continuous random variables. Conditional expectation is developed and shown to be the minimum mean square error estimator. The special case of conditional expectation applied to jointly Gaussian random variables is presented in detail. The section on the (simplistic) example of navigation using a GPS measurement is presented primarily to give some idea on how to take a problem and put it into a standard mathematical formulation for analysis. The orthogonality of the conditional expectation is also proven and discussed. Linear mean square estimation is then developed including its relationship to the orthogonality principle.

Chapter 12 generalizes the ideas developed in Chapter 10 for two random variables to n random variables. In particular, the general definition and properties of n jointly Gaussian random variables are given. The chapter ends with a section on the general case of linear mean square estimation.

Chapter 13 continues on from the brief look at the Bernoulli and geometric processes in Chapter 7 and the brief look at the Poisson process in Chapter 9 to give a more in depth study of these processes. The approach here is to do the modeling and analysis of arrivals in discrete time (which is easier conceptually and mathematically) and then take the limit to continuous time. In particular, the "waiting time paradox" in discrete time is completely worked out in a very simple manner using the Bernoulli model. The "fresh start" property of both the geometric and Bernoulli processes are derived in a simple manner as well. The second section in this chapter then shows how a binomial process S_n where $P(\{\omega|S_n = k\}) = \binom{n}{k}p^k(1-p)^{n-k}$ with n interpreted as time can be viewed as a discrete time version of the Poisson process. The advantage of this approach is that it very easy to show that the binomial process has independent increments, that its increments are also binomial random variables, and further, letting $n \to \infty$ with $np = \lambda > 0$ (λ fixed), that S_n converges in distribution to the Poisson distribution. This development is used to motivate the corresponding properties for the Poisson process which are not proven. After a discussion on the properties of the Poisson process (independent increments, the waiting time paradox, etc.), some interesting examples using these properties are presented. A section on the "Order Statistics" is presented as enough background has been developed to study this application. Finally, the last section covers shot noise as a classical application of the Poisson process. In particular, Campbell's theorem is derived by using a discrete-time Bernoulli process to approximate the Poisson process.

Chapter 14 uses the random walk process to motivate the Brownian motion (Wiener) process and also to develop the white noise process. In particular, the random walk is developed as a discrete-time Bernoulli process with parameters n and $p = 1/2$, where n is interpreted as discrete time and $p = 1/2$ is the probability of heads/tails. The central limit theorem is then invoked to show the distribution of the random walk goes to a Wiener (Gaussian) process. The independent increment property is also motivated and discussed. White noise is first motivated as a finite difference increment of the Wiener process, which is then used to motivate the idea that the autocorrelation of white noise is a delta function. Thermal white noise (voltage noise due to random motion of electrons in a metal) is presented next using the explicit autocorrelation function determined by J. B. Johnson and H. Nyquist in their classic papers published in 1928. Based on their representation of this noise as a zero-mean Gaussian process with their *experimentally* determined autocorrelation function, it is then shown why in typical engineering applications its autocorrelation function can be represented by a Dirac delta function. Care is taken to show when this representation is valid and when it is not. The explanation of white noise in the frequency domain is put off until the notion of power spectral density is developed in Chapter 15.

Chapter 15 develops the concept of a stationary random process which is then used to develop the idea of the power spectral density of a process. The first section of the chapter goes through the standard definitions for stationary random processes along with some (hopefully) clarifying examples. An overview of discrete-time linear time-invariant (LTI) systems is then presented including convolution, transfer functions, and state space representations. Also covered is stability, causality, and discrete-time Fourier transforms along with the deterministic versions of autocorrelation, energy and power spectral densities. Then the notion of power spectral density for discrete-time stationary random processes is discussed with quite a bit of effort given to developing its interpretation. After working in discrete time, an overview of continuous-time LTI systems is presented followed by the development of the notion of power spectral density for continuous-time random processes. There is then a section covering thermal white noise. The effort here concentrates on presenting an explanation that the standard approach of modeling the autocovariance of this noise as a Dirac delta function follows from the Physics of thermal noise and the way it is used in applications.

Chapter 16 presents the basic limit theorems of probability and the various notions of convergence. As an application of these ideas, a more in-depth development of Brownian motion is presented at the end of the chapter.

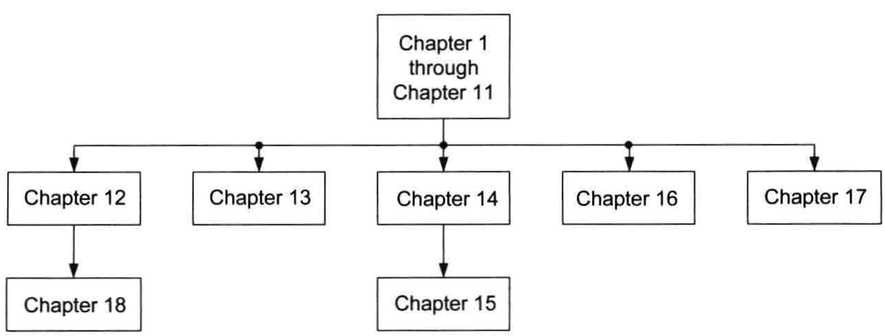
Chapter 17 develops the basic ideas of Statistics specifically including estimation of the mean and variance along with the development of confidence intervals. This chapter is based on the unpublished lecture notes of R. B. Ash [12].

Chapter 18 presents an introduction to Kalman filtering using Anderson and Moore's book [13] *Optimal Filtering* as the primary reference.

Finally, it must be said that this text has borrowed ideas and examples from many books and I have tried to be careful to cite the references from which I obtained such material.[5]

Logical Dependence of the Chapters

The logical dependence of the chapters is given in the figure below.



[5] That old paraphrase of Picasso is perhaps more appropriate: *Good ideas are inspired, great ideas are stolen!*

Use of the Book for a One Semester Course

The prerequisites for this book are an elementary undergraduate probability course, a signals and systems course (convolution and Fourier transforms), and some elementary matrix theory. This background is in a typical undergraduate electrical engineering program.

In using this book in a one-semester graduate-level course, the first 11 chapters can be covered as follows:

Chapter 1, Sections 1.1 through 1.6
Chapter 2, Sections 2.1 through 2.4
Chapter 3, Sections 3.1 through 3.4
Chapter 4, Sections 4.1 through 4.4
Chapter 5, Sections 5.1 through 5.4
Chapter 6, Sections 6.1 and 6.2
Chapter 7, Sections 7.1 through 7.3
Chapter 8, Sections 8.1 through 8.3
Chapter 9, Sections 9.1 through 9.5
Chapter 10, Sections 10.1 through 10.3
Chapter 11, Sections 11.1 through 11.6

The first few versions of the manuscript of this book were set up to skip Chapters 2 and 3 (they didn't exist). If this is done, there is time to cover

Chapter 13, Sections 13.1, 13.2, 13.3, and 13.5

or

Chapter 14, Sections 14.1, 14.2, 14.3, and 14.5
Chapter 15, Sections 15.1, 15.3, 15.4, and 15.6 through 15.8.

Computer Programs

There is a set of computer programs written in `Matlab` from MATHWORKS, INC. that accompany this book (for reference in the book, look under *Simulation* in the Index). They can be downloaded from the book's webpage which is found by going to the website *www.wiley.com* and searching for this book under its title and author. For the reader who does not have access to `Matlab`, the open source software `Octave` (www.octave.org) should be able to run these programs as well.

Acknowledgments

Many students endured the preliminary versions of this book. Among them, I would like to acknowledge Mark Wicks, Chellury Sastry, Samer Saab, Ahmed Oteafy, Hidayatullah Ahsan, Qawi Harvard, Charley Lester, Uri Rogers, Xia Li and Shane Taylor for their feedback.

I am grateful to George Pearson and John MacKendrick of the technical support team at MacKichan Software Inc. for answering my questions on `Scientific Word`, which was used to typeset this book in LaTeX.

At John Wiley & Sons, Inc., I would like to thank my editor Simone Taylor for getting my manuscript through the review and contract-signing process, my production editor Kellsee Chu for getting the book through the publication process, my copy editor Bob Golden for correcting many errors and inconsistencies in the manuscript, and senior editorial assistant Diana Gialo for all of her help.

The anonymous reviewers of the book provided very helpful comments which I have tried to incorporate into this final version.

Most importantly I want to express my deep gratitude to Professor Saul B. Gelfand for his many discussions with me on probability theory and clarifying for me many issues I had on the subject.

Feedback

Any comments, criticisms, or corrections are welcome and may be sent to the author at *chiasson@ieee.org*.

John Chiasson

试读结束：需要全本请在线购买： www.ertongbook.com