William J. Stewart

Probability, Markov Chains, Queues, and Simulation

The Mathematical Basis of Performance Modeling

概率论、马尔科夫链、排队和模拟

PRINCETON UNIVERSITY PRESS

世界图书出版公司
www.wpcbj.com.cn

PROBABILITY, MARKOV CHAINS, QUEUES, AND SIMULATION

The Mathematical Basis of
Performance Modeling

William J. Stewart

PRINCETON UNIVERSITY PRESS

PRINCETON AND OXFORD

Published by Princeton University Press, 41 William street, Princeton, New Jersey 08540
In the United Kingdom: Princeton University Press, 6 Oxford Street,
Woodstock, Oxfordshire OX20 1TW

Library of Congress Cataloging-in-Publication Data
Stewart, William J., 1946–
Probability, Markov chains, queues and simulation: the mathematical basis of performance modeling/William J. Stewart. – 1st ed.
p. cm.
ISBN 978-0-691-14062-9 (cloth : alk. paper) 1. Probability–Computer simulation.
2. Markov processes. 3. Queueing theory. I. Title.
QA273.S7532 2009
519.201′13–dc22 2008041122

British Library Cataloging-in-Publication Data is available

This book has been composed in Times
Printed on acid-free paper. ∞
press.princeton.edu

图书在版编目（CIP）数据

概率论、马尔科夫链、排队和模拟 = Probability, Markov Chains, Queues, and Simulation：英文/（美）斯图尔特（Stewart, W. J.）著. —影印本. —北京：世界图书出版公司北京公司，2013.5
ISBN 978-7-5100-6178-3

Ⅰ.①概… Ⅱ.①斯… Ⅲ.①概率论—高等学校—教材—英文②马尔柯夫过程—高等学校—教材—英文 Ⅳ.①O211②O211.62

中国版本图书馆 CIP 数据核字（2013）第 103752 号

书　　名：Probability, Markov Chains, Queues, and Simulation
作　　者：William J. Stewart
中 译 名：概率论、马尔科夫链、排队和模拟
责任编辑：高蓉　刘慧

出 版 者：世界图书出版公司北京公司
印 刷 者：三河市国英印务有限公司
发　　行：世界图书出版公司北京公司（北京朝内大街 137 号 100010）
联系电话：010-64021602，010-64015659
电子信箱：kjb@wpcbj.com.cn

开　　本：16 开
印　　张：49
版　　次：2013 年 10 月
版权登记：图字：01-2013-2914

书　　号：978-7-5100-6178-3　　　　定　　价：139.00 元

PROBABILITY, MARKOV CHAINS, QUEUES, AND SIMULATION

This book is dedicated to all those whom I love, especially

My dear wife, Kathie,
and my wonderful children
Nicola, Stephanie, Kathryn, and William

My father, William J. Stewart and
the memory of my mother, Mary (Marshall) Stewart

Preface and Acknowledgments

This book has been written to provide a complete, yet elementary and pedagogic, treatment of the mathematical basis of systems performance modeling. Performance modeling is of fundamental importance to many branches of the mathematical sciences and engineering as well as to the social and economic sciences. Advances in methodology and technology have now provided the wherewithal to build and solve sophisticated models. The purpose of this book is to provide the student and teacher with a modern approach for building and solving probability based models with confidence.

The book is divided into four major parts, namely, "Probability," "Markov Chains," "Queueing Models," and "Simulation." The eight chapters of Part I provide the student with a comprehensive and thorough knowledge of probability theory. Part I is self-contained and complete and should be accessible to anyone with a basic knowledge of calculus. Newcomers to probability theory as well as those whose knowledge of probability is rusty should be equally at ease in their progress through Part I. The first chapter provides the fundamental concepts of set-based probability and the probability axioms. Conditional probability and independence are stressed as are the laws of total probability and Bayes' rule. Chapter 2 introduces combinatorics—the art of counting—which is so important for the correct evaluation of probabilities. Chapter 3 introduces the concepts of random variables and distribution functions including functions of a random variable and conditioned random variables. This chapter prepares the ground work for Chapters 4 and 5: Chapter 4 introduces joint and conditional distributions and Chapter 5 treats expectations and higher moments. Discrete distribution functions are the subject of Chapter 6 while their continuous counterparts, continuous distribution functions, are the subject of Chapter 7. Particular attention is paid to phase-type distributions due to the important role they play in modeling scenarios and the chapter also includes a section on fitting phase-type distributions to given means and variances. The final chapter in Part I is devoted to bounds and limit theorems, including the laws of large numbers and the central limit theorem.

Part II contains two rather long chapters on the subject of Markov chains, the first on theoretical aspects of Markov chains, and the second on their numerical solution. In Chapter 9, the basic concepts of discrete and continuous-time Markov chains and their underlying equations and properties are discussed. Special attention is paid to irreducible Markov chains and to the potential, fundamental, and reachability matrices in reducible Markov chains. This chapter also contains sections on random walk problems and their applications, the property of reversibility in Markov chains, and renewal processes. Chapter 10 deals with numerical solutions, from Gaussian elimination and basic iterative-type methods for stationary solutions to ordinary differential equation solvers for transient solutions. Block methods and iterative aggregation-disaggregation methods for nearly completely decomposable Markov chains are considered. A section is devoted to matrix geometric and matrix analytic methods for structured Markov chains. Algorithms and computational considerations are stressed throughout this chapter.

Queueing models are presented in the five chapters that constitute Part III. Elementary queueing theory is presented in Chapter 11. Here an introduction to the basic terminology and definitions is followed by an analysis of the simplest of all queueing models, the *M/M/*1 queue. This is then generalized to birth-death processes, which are queueing systems in which the underlying Markov chain matrix is tridiagonal. Chapter 12 deals with queues in which the arrival process need no longer be Poisson and the service time need not be exponentially distributed. Instead, interarrival times and service times can be represented by phase-type distributions and the underlying Markov chain is now block tridiagonal. The following chapter, Chapter 13, explores the z-transform approach for solving similar types of queues. The *M/G/1* and *G/M/1* queues are the subject of Chapter 14. The approach used is that of the embedded Markov chain. The Pollaczek-Khintchine mean value and transform equations are derived and a detailed discussion of residual time and busy period follows. A thorough discussion of nonpreemptive and preempt-resume scheduling policies as well as shortest-processing-time-first scheduling is presented. An analysis is also provided for the case in which only a limited number of customers can be accommodated in both the *M/G/1* and *G/M/1* queues. The final chapter of Part III, Chapter 15, treats queueing networks. Open networks are introduced via Burke's theorem and Jackson's extensions to this theorem. Closed queueing networks are treated using both the convolution algorithm and the mean value approach. The "flow-equivalent server" approach is also treated and its potential as an approximate solution procedure for more complex networks is explored. The chapter terminates with a discussion of product form in queueing networks and the BCMP theorem for open, closed, and mixed networks.

The final part of the text, Part IV, deals with simulation. Chapter 16 explores how uniformly distributed random numbers can be applied to obtain solutions to probabilistic models and other time-independent problems—the "Monte Carlo" aspect of simulation. Chapter 17 describes the modern approaches for generating uniformly distributed random numbers and how to test them to ensure that they are indeed uniformly distributed and independent of each other. The topic of generating random numbers that are not uniformly distributed, but satisfy some other distribution such as Erlang or normal, is dealt with in Chapter 18. A large number of possibilities exist and not all are appropriate for every distribution. The next chapter, Chapter 19, provides guidelines for writing simulation programs and a number of examples are described in detail. Chapter 20 is the final chapter in the book. It concerns simulation measurement and accuracy and is based on sampling theory. Special attention is paid to the generation of confidence intervals and to variance reduction techniques, an important means of keeping the computational costs of simulation to a manageable level.

The text also includes two appendixes; the first is just a simple list of the letters of the Greek alphabet and their spellings; the second is a succinct, yet complete, overview of the linear algebra used throughout the book.

Genesis and Intent

This book saw its origins in two first-year graduate level courses that I teach, and have taught for quite some time now, at North Carolina State University. The first is entitled "An Introduction to Performance Evaluation;" it is offered by the Computer Science Department and the Department of Electrical and Computer Engineering. This course is required for our networking degrees. The second is a course entitled "Queues and Stochastic Service Systems" and is offered by the Operations Research Program and the Industrial and Systems Engineering Department. It follows then that this book has been designed for students from a variety of academic disciplines in which stochastic processes constitute a fundamental concept, disciplines that include not only computer science and engineering, industrial engineering, and operations research, but also mathematics, statistics, economics, and business, the social sciences—in fact all disciplines in which stochastic performance modeling plays a primary role. A calculus-based probability course is a prerequisite for both these courses so it is expected that students taking these classes are already familiar with probability theory. However, many of the students who sign up for these courses are returning students, and it is often the case that it has been several years and in some cases a decade or more, since they last studied probability. A quick review of probability is hardly sufficient to bring them up to the required level. Part I of the book has been designed with them in mind. It provides the prerequisite probability background needed to fully understand and appreciate the material in the remainder of the text. The presentation, with its numerous examples and exercises, is such that it facilitates an independent review so the returning student in a relatively short period of time, preferably prior to the beginning of class, will once again have mastered probability theory. Part I can then be used as a reference source as and when needed.

The entire text has been written at a level that is suitable for upper-level undergraduate students or first-year graduate students and is completely self-contained. The entirety of the text can be covered in a two-semester sequence, such as the stochastic processes sequence offered by the Industrial Engineering (IE) Department and the Operations Research (OR) Program at North Carolina State University. A two-semester sequence is appropriate for classes in which students have limited (or no) exposure to probability theory. In such cases it is recommended that the first semester be devoted to the Chapters 1–8 on probability theory, the first five sections of Chapter 9, which introduce the fundamental concepts of discrete-time Markov chains, and the first three sections of Chapter 11, which concern elementary queueing theory. With this background clearly understood, the student should have no difficulty in covering the remaining topics of the text in the second semester.

The complete content of Parts II–IV might prove to be a little too much for some one-semester classes. In this case, an instructor might wish to omit the later sections of Chapter 10 on the numerical solution of Markov chains, perhaps covering only the basic direct and iterative methods. In this case the material of Chapter 12 should also be omitted since it depends on a knowledge of the matrix geometric method of Chapter 10. Because of the importance of computing numerical solutions, it would be a mistake to omit Chapter 10 in its entirety. Some of the material in Chapter 18 could also be eliminated: for example, an instructor might include only the first three sections of this chapter. In my own case, when teaching the OR/IE course, I concentrate on covering all of the Markov chain and queueing theory chapters. These students often take simulation as an individual course later on. When teaching the computer science and engineering course, I omit some of the material on the numerical solution of Markov chains so as to leave enough time to cover simulation.

Numerous examples with detailed explanations are provided throughout the text. These examples are designed to help the student more clearly understand the theoretical and computational aspects of the material and to be in a position to apply the acquired knowledge to his/her own areas of interest. A solution manual is available for teachers who adopt this text for their courses. This manual contains detailed explanations of the solution of all the exercises.

Where appropriate, the text contains program modules written in Matlab or in the Java programming language. These programs are not meant to be robust production code, but are presented so that the student may experiment with the mathematical concepts that are discussed. To free the student from the hassle of copying these code segments from the book, a listing of all of the code used can be freely downloaded from the web page:
`http://press.princeton.edu/titles/8844.html`

Acknowledgments

As mentioned just a moment ago, this book arose out of two courses that I teach at North Carolina State University. It is, therefore, ineluctable that the students who took these courses contributed immeasurably to its content and form. I would like to express my gratitude to them for their patience and input. I would like to cite, in particular, Nishit Gandhi, Scott Gerard, Rong Huang, Kathryn Peding, Amirhosein Norouzi, Robert Shih, Hui Wang, Song Yang, and Shengfan Zhang for their helpful comments. My own doctoral students, Shep Barge, Tugrul Dayar, Amy Langville, Ning Liu, and Bin Peng, were subjected to different versions of the text and I owe them a particular expression of thanks. One person who deserves special recognition is my daughter Kathryn, who allowed herself to be badgered by her father into reading over selected probability chapters.

I would like to thank Vickie Kearn and the editorial and production staff at Princeton University Press for their help and guidance in producing this book. It would be irresponsible of me not to mention the influence that my teachers, colleagues, and friends have had on me. I owe them a considerable debt of gratitude for helping me understand the vital role that mathematics plays, not only in performance modeling, but in all aspects of life.

Finally, and most of all, I would like to thank my wife Kathie and our four children, Nicola, Stephanie, Kathryn, and William, for all the love they have shown me over the years.

Contents