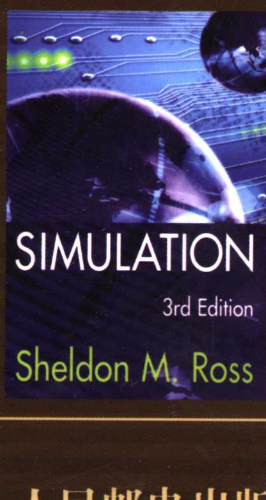
图灵原版数学·统计学系列

Simulation

统计模拟

（英文版·第3版）

[美] Sheldon M. Ross 著



人民邮电出版社
POSTS & TELECOM PRESS

TURING 图灵原版数学·统计学系列

Simulation

统计模拟

(英文版·第3版)

[美] Sheldon M. Ross 著

人民邮电出版社
POSTS & TELECOM PRESS

图书在版编目（CIP）数据

统计模拟／（美）罗斯著．—北京：人民邮电出版社，2006.1
（图灵原版数学·统计学系列）
ISBN 7-115-14125-8

Ⅰ．统… Ⅱ．罗… Ⅲ．统计学—模拟实验—英文 Ⅳ．C8-33

中国版本图书馆 CIP 数据核字（2005）第 127046 号

内 容 提 要

本书介绍了统计模拟的一些实用方法和技术。在对概率的基本知识进行了简单的回顾之后，介绍了如何利用计算机产生随机数以及如何利用这些随机数产生任意分布的随机变量、随机过程等。然后介绍一些分析统计数据的方法和技术，如 Bootstrap、方差缩减技术等。接着介绍了如何利用统计模拟来判断所选的随机模型是否拟合实际的数据。最后介绍了 MCMC 及一些最新发展的统计模拟技术和论题。

本书可作为统计学、计算数学、保险学、精算学等专业本科生教材，也可供相关专业人士参考。

图灵原版数学·统计学系列

统 计 模 拟（英文版·第 3 版）

◆ 著　[美] Sheldon M. Ross
　责任编辑　王丽萍

◆ 人民邮电出版社出版发行　北京市崇文区夕照寺街 14 号
　邮编 100061　电子函件 315@ptpress.com.cn
　网址 http://www.ptpress.com.cn
　北京顺义振华印刷厂印刷
　新华书店总店北京发行所经销

◆ 开本：800×1000 1/16
　印张：18
　字数：382 千字　2006 年 1 月第 1 版
　印数：1－3 000 册　2006 年 1 月北京第 1 次印刷

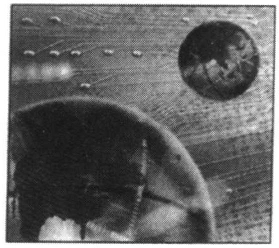
著作权合同登记号　图字：01-2004-4005 号

ISBN 7-115-14125-8/TP·5050

定价：39.00 元

读者服务热线：（010）88593802　印装质量热线：（010）67129223

Preface



Overview

In formulating a stochastic model to describe a real phenomenon, it used to be that one compromised between choosing a model that is a realistic replica of the actual situation and choosing one whose mathematical analysis is tractable. That is, there did not seem to be any payoff in choosing a model that faithfully conformed to the phenomenon under study if it were not possible to mathematically analyze that model. Similar considerations have led to the concentration on asymptotic or steady-state results as opposed to the more useful ones on transient time. However, the relatively recent advent of fast and inexpensive computational power has opened up another approach — namely, to try to model the phenomenon as faithfully as possible and then to rely on a simulation study to analyze it.

In this text we show how to analyze a model by use of a simulation study. In particular, we first show how a computer can be utilized to generate random (more precisely, pseudorandom) numbers, and then how these random numbers can be used to generate the values of random variables from arbitrary distributions. Using the concept of discrete events we show how to use random variables to generate the behavior of a stochastic model over time. By continually generating the behavior of the system we show how to obtain estimators of desired quantities of interest. The statistical questions of when to stop a simulation and what confidence to place in the resulting estimators are considered. A variety of ways in which one can improve on the usual simulation estimators are presented. In addition, we show how to use simulation to determine whether the stochastic model chosen is consistent with a set of actual data.

New to this Edition

Expositional and notational changes throughout the text.
New exercises in almost all chapters.

A new section (4.6) on generating random vectors, with an example illustrating how to generate a multinomial vector.

A new section (6.6) on using discrete events to simulate an insurance risk model.

A new section (8.7) on using simulation to efficiently evaluate the expected price of an exotic option. Various variance reduction methods are combined to obtain an efficient procedure for evaluating these options, which are of importance in finance and insurance.

A new section (11.4) on using simulation to estimate first passage time distributions of a Markov chain, a problem with a variety of applications in almost all areas of applied probability. The method utilized is also applied to deriving the tail probabilities of joint distributions, such as the bivariate normal distribution.

A new section (11.5) on coupling from the past, a techique that allows one to simulate a random variable whose distribution is that of the stationary distribution of a Markov chain.

A new example (8m) on using simulation to estimate tail probabilities of compound random variables. Such random variables are of great importance in insurance.

There is new material in Section 8.5 on using importance sampling to estimate tail probabilities.

Section 10.3 on the Gibbs sampler has been rewritten. A new example concerned with generating a multinomial vector conditional on the event that all outcomes occur at least once is presented.

Chapter Descriptions

The successive chapters in this text are as follows. **Chapter 1** is an introductory chapter which presents a typical phenomenon that is of interest to study. **Chapter 2** is a review of probability. Whereas this chapter is self-contained and does not assume the reader is familiar with probability, we imagine that it will indeed be a review for most readers. **Chapter 3** deals with random numbers and how a variant of them (the so-called pseudorandom numbers) can be generated on a computer. The use of random numbers to generate discrete and then continuous random variables is considered in Chapters 4 and 5.

Chapter 6 presents the discrete event approach to track an arbitrary system as it evolves over time. A variety of examples — relating to both single and multiple server queueing systems, to an insurance risk model, to an inventory system, to a machine repair model, and to the exercising of a stock option — are presented. **Chapter 7** introduces the subject matter of statistics. Assuming that our average reader has not previously studied this subject, the chapter starts with very basic concepts and ends by introducing the bootstrap statistical method, which is quite useful in analyzing the results of a simulation.

Chapter 8 deals with the important subject of variance reduction. This is an attempt to improve on the usual simulation estimators by finding ones having the same mean and smaller variances. The chapter begins by introducing the technique of using antithetic variables. We note (with a proof deferred to the chapter's appendix) that this always results in a variance reduction along with a computational savings when we are trying to estimate the expected value of a function that is monotone in each of its variables. We then introduce control variables and illustrate their usefulness in variance reduction. For instance, we show how control variables can be effectively utilized in analyzing queueing systems, reliability systems, a list reordering problem, and blackjack. We also indicate how to use regression packages to facilitate the resulting computations when using control variables. Variance reduction by use of conditional expectations is then considered. Its use is indicated in examples dealing with estimating π, and in analyzing finite capacity queueing systems. Also, in conjunction with a control variate, conditional expectation is used to estimate the expected number of events of a renewal process by some fixed time. The use of stratified sampling as a variance reduction tool is indicated in examples dealing with queues with varying arrival rates and evaluating integrals. The relationship between the variance reduction techniques of conditional expectation and stratified sampling is explained and illustrated in the estimation of the expected return in video poker. The technique of importance sampling is next considered. We indicate and explain how this can be an extremely powerful variance reduction technique when estimating small probabilities. In doing so, we introduce the concept of tilted distributions and show how they can be utilized in an importance sampling estimation of a small convolution tail probability. Applications of importance sampling to queueing, random walks, and random permutations, and to computing conditional expectations when one is conditioning on a rare event are presented. The final variance reduction technique of Chapter 8 relates to the use of a common stream of random numbers. An application to valuing an exotic stock option that utilizes a combination of variance reduction techniques is presented in Section 8.7.

Chapter 9 is concerned with statistical validation techniques, which are statistical procedures that can be used to validate the stochastic model when some real data are available. Goodness of fit tests such as the chi-square test and the Kolmogorov–Smirnov test are presented. Other sections in this chapter deal with the two-sample and the n-sample problems and with ways of statistically testing the hypothesis that a given process is a Poisson process.

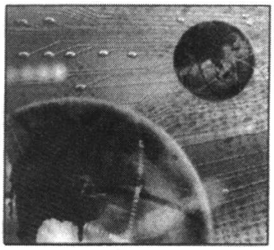
Chapter 10 is concerned with Markov chain Monte Carlo methods. These are techniques that have greatly expanded the use of simulation in recent years. The standard simulation paradigm for estimating $\theta = E[h(\mathbf{X})]$, where $\mathbf{X}$ is a random vector, is to simulate independent and identically distributed copies of $\mathbf{X}$ and then use the average value of $h(\mathbf{X})$ as the estimator. This is the so-called "raw" simulation estimator, which can then possibly be improved upon by using one or more of the variance reduction ideas of Chapter 8. However, in order to employ this approach it is necessary both that the distribution of $\mathbf{X}$ be specified and also that we be able to simulate from this distribution. Yet, as we see in Chapter 10, there are many examples where the distribution of **X** is known but we are not able to directly simulate the random vector **X**, and other examples where the distribution is not completely known but is only specified up to a multiplicative constant. Thus, in either case, the usual approach to estimating θ is not available. However, a new approach, based on generating a Markov chain whose limiting distribution is the distribution of **X**, and estimating θ by the average of the values of the function h evaluated at the successive states of this chain, has become widely used in recent years. These Markov chain Monte Carlo methods are explored in Chapter 10. We start, in Section 10.2, by introducing and presenting some of the properties of Markov chains. A general technique for generating a Markov chain having a limiting distribution that is specified up to a multiplicative constant, known as the Hastings–Metropolis algorithm, is presented in Section 10.3, and an application to generating a random element of a large "combinatorial" set is given. The most widely used version of the Hastings–Metropolis algorithm is known as the Gibbs sampler, and this is presented in Section 10.4. Examples are discussed relating to such problems as generating random points in a region subject to a constraint that no pair of points are within a fixed distance of each other, to analyzing product form queueing networks, to analyzing a hierarchical Bayesian statistical model for predicting the numbers of home runs that will be hit by certain baseball players, and to simulating a multinomial vector conditional on the event that all outcomes occur at least once. An application of the methods of this chapter to deterministic optimization problems, called simulated annealing, is presented in Section 10.5, and an example concerning the traveling salesman problem is presented. The final section of Chapter 10 deals with the sampling importance resampling algorithm, which is a generalization of the acceptance–rejection technique of Chapters 4 and 5. The use of this algorithm in Bayesian statistics is indicated.

Chapter 11 deals with some additional topics in simulation. In Section 11.1 we learn of the alias method which, at the cost of some setup time, is a very efficient way to generate discrete random variables. Section 11.2 is concerned with simulating a two-dimensional Poisson process. In Section 11.3 we present an identity concerning the covariance of the sum of dependent Bernoulli random variables and show how its use can result in estimators of small probabilities having very low variances. Applications relating to estimating the reliability of a system, which appears to be more efficient that any other known estimator of a small system reliability, and to estimating the probability that a specified pattern occurs by some fixed time, are given. Section 11.4 presents an efficient technique to employ simulation to estimate first passage time means and distributions of a Markov chain. An application to computing the tail probabilities of a bivariate normal random variable is given. Section 11.5 presents the *coupling from the past* approach to simulating a random variable whose distribution is that of the stationary distribution of a specified Markov chain.

Thanks

We are indebted to David Butler (Oregon State University), Matt Carlton (California Polytechnic State University), James Daniel (University of Texas, Austin), William Frye (Ball State University), Mark Glickman (Boston University), Chuanshu Ji (University of North Carolina), Yonghee Kim-Park (California State University, Long Beach), Donald E. Miller (St. Mary's College), Krzysztof Ostaszewski (Illinois State University), Erol Peköz, Boston University, Yuval Peres (University of California, Berkeley), and Esther Portnoy, (University of Illinois, Urbana-Champaign) for their many helpful comments. We would like to thank those text reviewers who wish to remain anonymous.

Contents



1 Introduction 1

Exercises 3

2 Elements of Probability 5

2.1 Sample Space and Events 5
2.2 Axioms of Probability 6
2.3 Conditional Probability and Independence 7
2.4 Random Variables 8
2.5 Expectation 10
2.6 Variance 13
2.7 Chebyshev's Inequality and the Laws of Large Numbers 15
2.8 Some Discrete Random Variables 17
Binomial Random Variables 17
Poisson Random Variables 18
Geometric Random Variables 20
The Negative Binomial Random Variable 20
Hypergeometric Random Variables 21
2.9 Continuous Random Variables 22
Uniformly Distributed Random Variables 22
Normal Random Variables 23
Exponential Random Variables 25
The Poisson Process and Gamma Random Variables 27
The Nonhomogeneous Poisson Process 29

2.10 Conditional Expectation and Conditional Variance 30
Exercises 32
References 36

3 Random Numbers 37

Introduction 37
3.1 Pseudorandom Number Generation 37
3.2 Using Random Numbers to Evaluate Integrals 38
Exercises 42
References 44

4 Generating Discrete Random Variables 45

4.1 The Inverse Transform Method 45
4.2 Generating a Poisson Random Variable 50
4.3 Generating Binomial Random Variables 52
4.4 The Acceptance–Rejection Technique 53
4.5 The Composition Approach 55
4.6 Generating Random Vectors 56
Exercises 57

5 Generating Continuous Random Variables 63

Introduction 63
5.1 The Inverse Transform Algorithm 63
5.2 The Rejection Method 67
5.3 The Polar Method for Generating Normal Random Variables 73
5.4 Generating a Poisson Process 76
5.5 Generating a Nonhomogeneous Poisson Process 77
Exercises 81
References 85

6 The Discrete Event Simulation Approach 87

Introduction 87
6.1 Simulation via Discrete Events 87
6.2 A Single-Server Queueing System 88
6.3 A Queueing System with Two Servers in Series 91
6.4 A Queueing System with Two Parallel Servers 93
6.5 An Inventory Model 96
6.6 An Insurance Risk Model 97
6.7 A Repair Problem 99
6.8 Exercising a Stock Option 102

6.9 Verification of the Simulation Model 103
Exercises 105
References 108

7 Statistical Analysis of Simulated Data 109

Introduction 109
7.1 The Sample Mean and Sample Variance 109
7.2 Interval Estimates of a Population Mean 115
7.3 The Bootstrapping Technique for Estimating Mean Square Errors 118
Exercises 124
References 127

8 Variance Reduction Techniques 129

Introduction 129
8.1 The Use of Antithetic Variables 131
8.2 The Use of Control Variates 139
8.3 Variance Reduction by Conditioning 147
Estimating the Expected Number of Renewals by Time t 155
8.4 Stratified Sampling 157
8.5 Importance Sampling 166
8.6 Using Common Random Numbers 180
8.7 Evaluating an Exotic Option 181
Appendix: Verification of Antithetic Variable Approach When Estimating the Expected Value of Monotone Functions 185
Exercises 188
References 195

9 Statistical Validation Techniques 197

Introduction 197
9.1 Goodness of Fit Tests 197
The Chi-Square Goodness of Fit Test for Discrete Data 198
The Kolmogorov–Smirnov Test for Continuous Data 200
9.2 Goodness of Fit Tests When Some Parameters Are Unspecified 205
The Discrete Data Case 205
The Continuous Data Case 208
9.3 The Two-Sample Problem 208
9.4 Validating the Assumption of a Nonhomogeneous Poisson Process 215
Exercises 219
References 221

10 Markov Chain Monte Carlo Methods 223

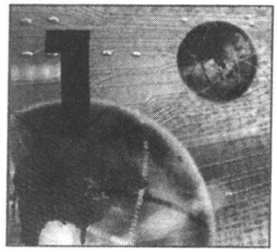
Introduction 223
10.1 Markov Chains 223
10.2 The Hastings–Metropolis Algorithm 226
10.3 The Gibbs Sampler 228
10.4 Simulated Annealing 239
10.5 The Sampling Importance Resampling Algorithm 242
Exercises 246
References 249

11 Some Additional Topics 251

Introduction 251
11.1 The Alias Method for Generating Discrete Random Variables 251
11.2 Simulating a Two-Dimensional Poisson Process 255
11.3 Simulation Applications of an Identity for Sums of Bernoulli Random Variables 258
11.4 Estimating the Distribution and the Mean of the First Passage Time of a Markov Chain 262
11.5 Coupling from the Past 267
Exercises 269
References 271

Index 272

Introduction



Consider the following situation faced by a pharmacist who is thinking of setting up a small pharmacy where he will fill prescriptions. He plans on opening up at 9 A.M. every weekday and expects that, on average, there will be about 32 prescriptions called in daily before 5 P.M. His experience indicates that the time that it will take him to fill a prescription, once he begins working on it, is a random quantity having a mean and standard deviation of 10 and 4 minutes, respectively. He plans on accepting no new prescriptions after 5 P.M., although he will remain in the shop past this time if necessary to fill all the prescriptions ordered that day. Given this scenario the pharmacist is probably, among other things, interested in the answers to the following questions:

1. What is the average time that he will depart his store at night?
2. What proportion of days will he still be working at 5:30 P.M.?
3. What is the average time it will take him to fill a prescription (taking into account that he cannot begin working on a newly arrived prescription until all earlier arriving ones have been filled)?
4. What proportion of prescriptions will be filled within 30 minutes?
5. If he changes his policy on accepting all prescriptions between 9 A.M. and 5 P.M., but rather only accepts new ones when there are fewer than five prescriptions still needing to be filled, how many prescriptions, on average, will be lost?
6. How would the conditions of limiting orders affect the answers to questions 1 through 4?

In order to employ mathematics to analyze this situation and answer the questions, we first construct a probability model. To do this it is necessary to make some reasonably accurate assumptions concerning the preceding scenario. For instance, we must make some assumptions about the probabilistic mechanism that describes the arrivals of the daily average of 32 customers. One possible assumption might be that the arrival rate is, in a probabilistic sense, constant over the day, whereas a second (probably more realistic) possible assumption is that the arrival rate depends on the time of day. We must then specify a probability distribution (having mean 10 and standard deviation 4) for the time it takes to service a prescription, and we must make assumptions about whether or not the service time of a given prescription always has this distribution or whether it changes as a function of other variables (e.g., the number of waiting prescriptions to be filled or the time of day). That is, we must make probabilistic assumptions about the daily arrival and service times. We must also decide if the probability law describing a given day changes as a function of the day of the week or whether it remains basically constant over time. After these assumptions, and possibly others, have been specified, a probability model of our scenario will have been constructed.

Once a probability model has been constructed, the answers to the questions can, in theory, be analytically determined. However, in practice, these questions are much too difficult to determine analytically, and so to answer them we usually have to perform a simulation study. Such a study programs the probabilistic mechanism on a computer, and by utilizing "random numbers" it simulates possible occurrences from this model over a large number of days and then utilizes the theory of statistics to estimate the answers to questions such as those given. In other words, the computer program utilizes random numbers to generate the values of random variables having the assumed probability distributions, which represent the arrival times and the service times of prescriptions. Using these values, it determines over many days the quantities of interest related to the questions. It then uses statistical techniques to provide estimated answers — for example, if out of 1000 simulated days there are 122 in which the pharmacist is still working at 5:30, we would estimate that the answer to question 2 is 0.122.

In order to be able to execute such an analysis, one must have some knowledge of probability so as to decide on certain probability distributions and questions such as whether appropriate random variables are to be assumed independent or not. A review of probability is provided in Chapter 2. The bases of a simulation study are so-called random numbers. A discussion of these quantities and how they are computer generated is presented in Chapter 3. Chapters 4 and 5 show how one can use random numbers to generate the values of random variables having arbitrary distributions. Discrete distributions are considered in Chapter 4 and continuous ones in Chapter 5. After completing Chapter 5, the reader should have some insight into the construction of a probability model for a given system and also how to use random numbers to generate the values of random quantities related to this model. The use of these generated values to track the system as it evolves continuously over time — that is, the actual simulation of the system — is discussed in Chapter 6, where we present the concept of "discrete events" and indicate how to utilize these entities to obtain a systematic approach to simulating systems. The discrete event simulation approach leads to a computer program, which can be written in whatever language the reader is comfortable in, that simulates the system a large number of times. Some hints concerning the verification of this program — to ascertain that it is actually doing what is desired — are also given in Chapter 6. The use of the outputs of a simulation study to answer probabilistic questions concerning the model necessitates the use of the theory of statistics, and this subject is introduced in Chapter 7. This chapter starts with the simplest and most basic concepts in statistics and continues toward the recent innovation of "bootstrap statistics," which is quite useful in simulation. Our study of statistics indicates the importance of the variance of the estimators obtained from a simulation study as an indication of the efficiency of the simulation. In particular, the smaller this variance is, the smaller is the amount of simulation needed to obtain a fixed precision. As a result we are led, in Chapter 8, to ways of obtaining new estimators that are improvements over the raw simulation estimators because they have reduced variances. This topic of variance reduction is extremely important in a simulation study because it can substantially improve its efficiency. Chapter 9 shows how one can use the results of a simulation to verify, when some real-life data are available, the appropriateness of the probability model (which we have simulated) to the real-world situation. Chapter 10 introduces the important topic of Markov chain Monte Carlo methods. The use of these methods has, in recent years, greatly expanded the class of problems that can be attacked by simulation. Chapter 11 considers a variety of additional topics.

Exercises

1. The following data yield the arrival times and service times that each customer will require, for the first 13 customers at a single server system. Upon arrival, a customer either enters service if the server is free or joins the waiting line. When the server completes work on a customer, the next one in line (i.e., the one who has been waiting the longest) enters service.

Arrival Times:	12	31	63	95	99	154	198	221	304	346	411	455	537
Service Times:	40	32	55	48	18	50	47	18	28	54	40	72	12

(a) Determine the departure times of these 13 customers.
(b) Repeat (a) when there are two servers and a customer can be served by either one.
(c) Repeat (a) under the new assumption that when the server completes a service, the next customer to enter service is the one who has been waiting the least time.

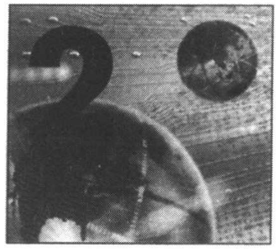
2. Consider a service station where customers arrive and are served in their order of arrival. Let A_n, S_n, and D_n denote, respectively, the arrival time, the service time, and the departure time of customer n. Suppose there is a single server and that the system is initially empty of customers.

(a) With $D_0 = 0$, argue that for $n > 0$

$$D_n - S_n = \text{Maximum}\,\{A_n, D_{n-1}\}$$

(b) Determine the corresponding recursion formula when there are two servers.
(c) Determine the corresponding recursion formula when there are k servers.
(d) Write a computer program to determine the departure times as a function of the arrival and service times and use it to check your answers in parts (a) and (b) of Exercise 1.

Elements of Probability



2.1 Sample Space and Events

Consider an experiment whose outcome is not known in advance. Let S, called the sample space of the experiment, denote the set of all possible outcomes. For example, if the experiment consists of the running of a race among the seven horses numbered 1 through 7, then

$$S = \{\text{all orderings of } (1, 2, 3, 4, 5, 6, 7)\}$$

The outcome (3, 4, 1, 7, 6, 5, 2) means, for example, that the number 3 horse came in first, the number 4 horse came in second, and so on.

Any subset A of the sample space is known as an event. That is, an event is a set consisting of possible outcomes of the experiment. If the outcome of the experiment is contained in A, we say that A has occurred. For example, in the above, if

$$A = \{\text{all outcomes in } S \text{ starting with } 5\}$$

then A is the event that the number 5 horse comes in first.

For any two events A and B we define the new event $A \cup B$, called the union of A and B, to consist of all outcomes that are either in A or B or in both A and B. Similarly, we define the event AB, called the intersection of A and B, to consist of all outcomes that are in both A and B. That is, the event $A \cup B$ occurs if either A or B occurs, whereas the event AB occurs if both A and B occur. We can also define unions and intersections of more than two events. In particular, the union of the events $A_1, \ldots, A_n$ — designated by $\cup_{i=1}^{n} A_i$ — is defined to consist of all outcomes that are in any of the A_i. Similarly, the intersection of the events $A_1, \ldots, A_n$ — designated by $A_1 A_2 \cdots A_n$ — is defined to consist of all outcomes that are in all of the A_i.