



WILLIAM L. DUNN
J. KENNETH SHULTIS



EXPLORING MONTE CARLO METHODS

Praise for *Exploring Monte Carlo Methods*

This book by Dunn and Shultis provides a comprehensive mathematical formalism for the important topics of Monte Carlo. In doing so, it mitigates the existing scarcity of textbooks on Monte Carlo and its applications. The book targets a wide audience, from students to senior Monte Carlo experts. Many teachers have been eagerly awaiting such a pedagogical tool. I know that I will use this outstanding book and will recommend it to my students.

Pedro Vaz
ITN-Portugal

The authors' proficiency and experience with Monte Carlo is apparent on every page. Worked examples render crystal clear even the most complex concepts to the beginner. Rare inclusions like inverse Monte Carlo and adjoint transport will appeal to specialists. The link between Monte Carlo and the transport formalism provides a formal physical framework which is rarely found. I love this book!

Jorge E. Fernández
Professor
University of Bologna

What a terrific resource for practitioners of the Monte Carlo method! There has long been a gap in the published Monte Carlo literature between instructional material that was either too easy or too highly mathematical to be relevant to the everyday Monte Carlo user and university student. Dunn and Shultis now have hit a sweet spot in bridging that gap with this text.

Roger L. Martz
MCNP Developer
Los Alamos National Laboratory

Exploring Monte Carlo Methods

William L. Dunn
J. Kenneth Shultis



AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO
Academic Press is an imprint of Elsevier



Academic Press is an imprint of Elsevier
Radarweg 29, PO Box 211, 1000 AE Amsterdam, The Netherlands
The Boulevard, Langford Lane, Kidlington, Oxford, OX5 1GB, UK
30 Corporate Drive, Suite 400, Burlington, MA 01803, USA
525 B Street, Suite 1900, San Diego, CA 92101-4495, USA

Copyright © 2012 Elsevier B.V. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the Publisher. Details on how to seek permission, further information about the Publisher's permissions policies, and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notices

Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

Library of Congress Cataloging-in-Publication Data

Dunn, William L. (William Lee), 1944-

Exploring Monte Carlo methods / William L. Dunn, J. Kenneth Shultis.
p. cm.

ISBN 978-0-444-51575-9

1. Monte Carlo method. I. Shultis, J. Kenneth. II. Title.

QA298.D86 2012

518'.282-dc22

2010050137

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

For information on all Academic Press publications
visit our Web site at www.elsevierdirect.com

Typeset by: diacriTech, India

Printed in the United States of America

11 12 13 14 9 8 7 6 5 4 3 2 1

Working together to grow
libraries in developing countries

www.elsevier.com | www.bookaid.org | www.sabre.org

ELSEVIER

BOOK AID
International

Sabre Foundation

To the women in our lives

Preface

Monte Carlo methods for solving various numerical problems have been used for over a century, but saw relatively limited use until the past several decades when inexpensive digital computers with immense computational power became ubiquitous. Today, Monte Carlo methods dominate many areas of science and engineering, largely because of their ability to treat complex problems that previously could be approximated only by simplified deterministic methods. Various Monte Carlo methods now are used routinely by scientists, engineers, mathematicians, statisticians, economists, and others who work in a wide range of quantitative disciplines.

This book explores the central features and tools that allow Monte Carlo methods to be useful in such a wide variety of applications. The emphasis in this book is on the practical implementation of Monte Carlo rather than on strict mathematical rigor. Our references cite such rigor, but most Monte Carlo users need not wade through the arcane proofs that underlie Monte Carlo methods in order to use those methods effectively. The primary purpose of this book is to serve as an introductory textbook on Monte Carlo, at the undergraduate or graduate level. Another purpose is to provide a text that readers can use to teach themselves how to utilize Monte Carlo methods. The book also is intended to be of use to readers who explore new subjects for the joy that can be derived from learning. Finally, it is hoped that those experienced in Monte Carlo methods may find the book useful as a general reference.

The idea for a basic book on Monte Carlo was conceived many years ago, shortly after one of us (W.L.D.) learned how to apply Monte Carlo to radiation transport problems from reading various accounts written in the 1950s and 1960s. Events conspired to keep him from devoting the time necessary to complete the task, until he joined the faculty of the Mechanical and Nuclear Engineering Department at Kansas State University in 2002. There he found not only the opportunity but also a willing colleague (J.K.S.) to provide valuable assistance to the endeavor. Fortunately, new ideas have emerged in the interim and the book now serves a greater purpose than originally envisioned.

The first eight chapters of this book are generic, in that they apply to Monte Carlo methods that can be used in any field of application. The next two chapters focus on applications to radiation transport because we are nuclear engineers and have experience in this area. However, we have endeavored to make the entire book sufficiently generic and accessible so that it can be useful to practitioners in many fields. Thus, the chapters on radiation transport focus on basic issues and provide guidance as to how the generic Monte Carlo methods can be applied in a particular field.

Each of the chapters incorporates examples and is followed by a list of problems or exercises. Five appendices are included that provide useful supporting material, such

as descriptions of a number of probability distributions and a summary of several of the more popular general-purpose radiation transport simulation codes.

We cannot adequately recognize and thank all of those who have contributed to the creation of this work. Dr. Robin P. Gardner first exposed one of us (W.L.D.) to Monte Carlo in the late 1960s and has continued to be an inspiration in many ways. We gratefully acknowledge the reviews of substantial portions of drafts of this book that Richard E. Faw and Clell J. Solomon provided. We also thank Richard Hugtenburg who kindly reviewed the chapter on Markov chain Monte Carlo. Additionally, we note that many practitioners in the field of Monte Carlo provided substantial input to Appendix D. These people are true experts in Monte Carlo who not only have produced valuable general-purpose codes but also have taken the time to summarize their codes for the benefit of the readers of this book. The many students who have waded through early drafts and who pointed out deficiencies and places where clarifications were needed have been a tremendous help to us. Also, we thank the LAB for its hospitality during the many hours spent there revising early drafts.

The electronic files used by the publisher were prepared by the authors using $\text{\LaTeX}$. We thank Elsevier for an excellent job of copyediting the manuscript we submitted. We welcome comments and would like to be informed of typographical or other errors that readers may discover.

Finally, we acknowledge the support and inspiration that our wives and children have provided us throughout our professional careers.

William L. Dunn (dunn@mne.ksu.edu)

J. Kenneth Shultis (jks@ksu.edu)

Manhattan, KS

September 2010

1 Introduction

*Many things Comte Buffon did adore,
among them dropping needles galore.
He did this you see,
because to his glee,
he could find π as never before.*

Heads or tails; right or left; up or down; yes or no; zero or one: Can such basic alternatives really mean much? Well, look at what digital computers do with “bits.” Bits become bytes, which aggregate into kilobytes, gigabytes, petabytes, and even larger units. Basic bits of data can be manipulated to represent almost anything we can imagine. Simple combinations of binary data are used to represent integers, floating-point numbers, characters, colors, images, and many other things. The rich structure of the world in which we live and the cosmos in which our world exists results from combinations of basic entities, such as protons and electrons. Each snowflake is unique, although the constituent water molecules are identical. Macromizing the microscopic yields diversity and complexity.

The graininess of the atomic world is not apparent to us in our everyday lives. But the reflection we see when we look in a mirror is provided by many individual photons performing the dance choreographed by the optical laws of nature. The weather report we see on television is a result of huge calculations made on computers that are fed massive amounts of data from diverse instruments. Indeed, the images we see on older televisions are formed from the collective flow of immense numbers of electrons, each individually following the path prescribed by electromagnetic elements in the picture tube. Complex systems result from combinations of very simple elements. “It from bit,” as the eminent physicist John Wheeler eloquently stated [Wheeler 1998]. In some sense, life, perception, and knowledge exist because basic binary combinations contain information.

The simple expedient of imagining a continuum between two discrete alternatives enriches the possibilities. The number space between zero and one becomes infinite—it has unlimited values. Try to represent the rational number $1/3$ in a finite number of decimal bits; the threes just keep coming. The concept of one-third is simple enough, but an exact digital representation is problematic. In the continuous world, the space between zero and one is infinite.

1.1 What Is Monte Carlo?

The analysis technique called *Monte Carlo* is, in essence, a methodology to use sample means to estimate population means. The term Monte Carlo was coined when digital

computers were first used to implement the procedure that, until then, had been called statistical sampling. In its simplest form, Monte Carlo turns bits into meaningful rational numbers. Run an experiment that leads to either success (one) or failure (zero). Repeat the experiment N times. Call N the number of *histories* or *trials*. Then divide the number of successes S by the number of trials. There are $N + 1$ possible outcomes ($0, 1/N, 2/N, \dots, N/N = 1$). If N is large, then the quantity S/N gives a good approximation to the average or expected value of the experiment. In this form, each history is a bit (a zero or a one) and the result is a rational number. In the next section, it is shown how this simple process can be used to estimate the irrational number π .

The discrete formulation of Monte Carlo becomes much more powerful in the fuller, continuous framework. Monte Carlo allows one to create viable answers to complex questions by appropriate actions with repetitive histories. In this more usual form, each history yields a real number (or a floating-point number created on a digital computer) or a collection of such numbers. In other forms, histories can lead to algebraic functions. However, the result is still the sum of the histories divided by the number of trials.

More broadly, Monte Carlo is a widely used numerical procedure that allows people to estimate answers to a wide variety of problems [Dunn and Shultis 2009]. In essence, Monte Carlo is a highly flexible and powerful form of quadrature, or numerical integration, that can be applied to a very wide range of problems, both direct and inverse. One of the unique aspects of Monte Carlo is that it allows the user a high degree of flexibility. One can introduce intuition into the solution, as long as one follows certain simple rules. This means that there are many ways to solve a given problem by Monte Carlo, and the investigator can experiment with different ways. Because Monte Carlo involves simple techniques to “solve” complex problems, it is powerful; because the technique is flexible, it is interesting and even fun.

Partially because of its fanciful name and partially because it seems to rely on games of chance, the first impression one may get about Monte Carlo is that it is a somewhat mysterious hit-or-miss technique that really should not be trusted. Indeed, according to Halton [1970], Trotter and Tukey [1954] claimed “the only good Monte Carlos are dead Monte Carlos.” However, Monte Carlo simulation is a very sound technique based on two mathematical theorems. If you were to go to the city of Monte Carlo and try your hand at some games of chance offered there, you might consider yourself lucky to win more money than you lost; however, if you use the simulation procedure of Monte Carlo and play the game correctly, you almost always win.

How can Monte Carlo be used to solve “real-world” problems? What are the tools one needs to extend this technique to solve increasingly complex problems? Can Monte Carlo be used to solve design and optimization problems? How can one quantify the uncertainty in one’s estimate of a complex outcome? These and many other issues are explored in the remainder of this book.

1.2 A Brief History of Monte Carlo

This section is based on brief accounts of the background of Monte Carlo provided by several authors. We especially wish to acknowledge the review articles prepared

by Nicholas Metropolis (one of the founders of Monte Carlo) [Metropolis 1987], Chapter 1 of Hammersley and Handscomb [1964], the interesting biographical summaries available on the World Wide Web provided by the School of Mathematics and Statistics at the University of St. Andrews, Scotland, and Chapter 11 of the shielding book by Shultis and Faw [2000].

The history of Monte Carlo can be traced back to the interesting Bernoulli family. Nicolaus Bernoulli (1623–1708) was a Belgian merchant who took over the family spice business in Basel, Switzerland. (His father, a Protestant, had left Amsterdam for religious and political reasons and moved his spice business to Basel.) Among Nicolaus' many children were two brothers gifted in mathematics, Jacob (1654–1705) and Johann (1667–1748). Jacob (a.k.a. Jacques or James) earned his degrees in philosophy and theology but loved mathematics and, in 1687, was appointed professor of mathematics in Basel. He initially collaborated with his brother Johann, but the relationship soured as the two were fiercely competitive. Jacob produced, in 1689, the original statement of the law of large numbers, on which Monte Carlo is based (see Chapter 2). According to Maistrov [1974], this law was eventually published in Jacob's posthumous "Ars Conjectandi" [Bernoulli 1713] or "The Art of Conjecture." Whereas Jacob must have conducted some exploratory tests, the law does not appear to have been put to extensive use at the time to estimate expectations of complex phenomena.



Figure 1.1 Jacob Bernoulli (1654–1705).

Also, it is interesting to note that Johann tutored the French mathematician Guillaume François Antoine Marquis de l'Hôpital (1661–1704), and in fact, Johann first enunciated what is now known as l'Hôpital's rule, in notes he supplied (for pay) to l'Hôpital. Both Jacob and Johann made many important contributions to mathematics. In his history of calculus, Jacob was the first to use the term *integral* in its present sense. He was also the first to solve what is known as the *Bernoulli equation*, he gave us the *Bernoulli numbers*, and he introduced probability as a quantity and not just a concept.

Almost a century later, George-Louis Leclerc, Comte de Buffon (Fig. 1.2), proposed [Buffon 1777] that if needles or sticks of length L were dropped randomly onto a plane surface that had parallel lines at a spacing of $D \geq L$, then the probability P_{cut} that a needle would cut (overlap or intersect) one of the lines could be expressed as (see Example 1.1)

$$P_{cut} = \frac{2L}{\pi D}. \quad (1.1)$$

Pierre-Simon de Laplace showed [Laplace 1786] that such a procedure could be used to estimate the value of π , because L and D are known and one can estimate P_{cut} from the number of successes (needles cutting a line, n_c) over trials (needles dropped, n_d). Thus,

$$\pi = \frac{2L}{P_{cut}D} \simeq \frac{2L}{(n_c/n_d)D}. \quad (1.2)$$

This problem is an excellent example of analog Monte Carlo, in which one phenomenon (dropping sticks or needles) “simulates” another phenomenon (estimating the irrational number π). The situation is depicted in Fig. 1.3, in which ten needles of length $L = 0.6D$ units are shown. For this case, Eq. (1.2) gives

$$\pi \simeq \frac{2(0.6)}{0.4(1)} = 3.$$

A more accurate estimate of π is obtained with a larger number of needle drops.

The practice of estimating π by needle drops became somewhat fashionable in the nineteenth century and is recounted, for instance, by Hall [1873]. Some, however, may not have played the game fairly. For example, the mathematician Lazzarini stated in 1901 that he, using needles with $L/D = 5/6$, had obtained $n_c = 1808$ intersections in $n_d = 3408$ needle drops, thereby producing an estimate of $\pi \simeq 3.1415929$, which is accurate to seven significant figures! However, as pointed out by Rao [1989] (reported



Figure 1.2 George-Louis Leclerc, Comte de Buffon (1707–1788).

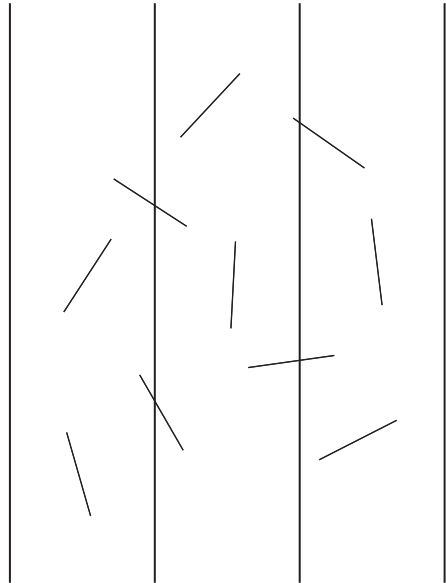


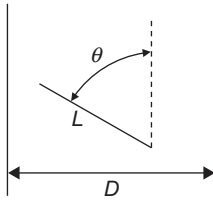
Figure 1.3 A grid of parallel lines with ten needles dropped at random.

by [Gentle \[1998\]](#)), this high accuracy is suspect. The best rational approximation for π with integers less than 16,000 is $\pi \simeq 355/113 \simeq 3.141592920 \dots$ ¹ From [Eq. \(1.2\)](#), with $P_{cut} \simeq n_c/n_d$,

$$n_d \simeq \pi \frac{D}{2L} n_c \simeq \frac{355}{113} \frac{6}{2 \times 5} n_c = \frac{213}{113} n_c.$$

Because n_d must be an integer, n_c must be a multiple of the prime number 113. If $n_c = 16 \times 113$, then $n_d = 3408$. Assuming Lazzarini actually obtained the stated results, he could have done so by dropping 213 needles k times, until he obtained precisely $113k$ successes, which he apparently did after $k = 16$ attempts.

Example 1.1 Derivation of [Eq. \(1.1\)](#)

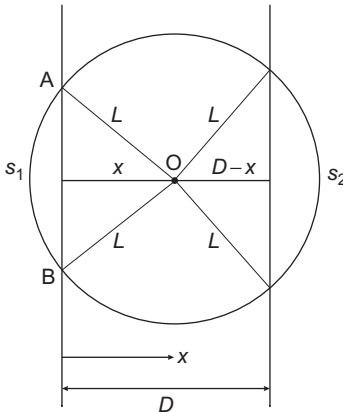


Consider the needle of length $L \leq D$ at an angle θ to the vertical shown to the left. The horizontal component of the needle is $L \sin \theta$, and, because the needle end has a uniform probability of being anywhere between the two lines, the probability a needle with angle θ cuts the left line is simply $P_{cut}(\theta) = (L \sin \theta)/D$ provided $L \leq D$. The probability a randomly dropped needle has an angle in $d\theta$ about θ is $d\theta/\pi$. Hence, the probability that a randomly dropped needle cuts a grid line is

$$P_{cut} = \int_0^\pi P_{cut}(\theta) \frac{d\theta}{\pi} = \int_0^\pi \frac{L \sin \theta}{D} \frac{d\theta}{\pi} = \frac{L}{\pi D} \int_0^\pi \sin \theta d\theta = \frac{2L}{\pi D}. \quad (1.3)$$

This derivation holds only for $L \leq D$. An alternative derivation, which can be generalized to needles of any length (see problem 5), is given in [Example 1.2](#).

Example 1.2 A More General Buffon's Needle Derivation



Consider the case shown in the figure to the left when one end of the needle is dropped at O, a distance x from the left line. Because all orientations of the needle are equally probable, the other end of the needle can be anywhere on the circle of radius L centered at O. The probability a needle intersects either the right or left line is just the sum of the arc lengths s_1 and s_2 outside the strip divided by the circle's circumference $2\pi L$. The arc length s_1 is L times the angle AOB, which can be expressed as

$$s_1(x) = \begin{cases} 2L \cos^{-1}(x/L), & 0 < x < L \\ 0, & x \geq L \end{cases}. \quad (1.4)$$

¹ The next rational approximation for π is $52,163/16,604 \simeq 3.141592387 \dots$, which has essentially the same accuracy as $\pi = 355/113$.

Similarly,

$$s_2(x) = \begin{cases} 2L \cos^{-1}((D-x)/L), & x > D-L \\ 0, & 0 < x \leq D-L \end{cases} \quad (1.5)$$

The end of the needle can land anywhere in the strip with equal probability, so the probability the needle end is dropped in dx about x is simply dx/D . The probability a needle with an end at x intersects one of the bounding lines is $[s_1(x) + s_2(x)]/(2\pi L)$. Hence, the probability a dropped needle intersects either bounding line is

$$P_{cut} = \int_0^D \frac{dx}{D} \frac{s_1(x) + s_2(x)}{2\pi L}. \quad (1.6)$$

Substitution for $s_1(x)$ and $s_2(x)$ and a change of variables yields

$$P_{cut} = \begin{cases} \frac{2}{\pi D} \int_0^D \cos^{-1}\left(\frac{x}{L}\right) dx & \text{for } D < L \\ \frac{2}{\pi D} \int_0^L \cos^{-1}\left(\frac{x}{L}\right) dx & \text{for } D \geq L \end{cases} \quad (1.7)$$

For $D \geq L$, Eq. (1.7) evaluates to Eq. (1.1), namely, $P_{cut} = 2L/(\pi D)$. Evaluation of Eq. (1.7) for $D < L$ is slightly more complicated (see problem 6).

Lord Rayleigh commented on Bernoulli's theorem in an article before the turn of the twentieth century [Rayleigh 1899]. Also, Lord Kelvin (born William Thomson, in 1824) would generate numbers at random by drawing bits of paper, from a glass jar, on which numbers had been written. He used these numbers for some of his studies of the kinetic theory of gases [Kelvin 1901]. This practice of picking random numbers from jars also was used by statisticians in Britain in the early twentieth century as a means of demonstrating statistical concepts to students. In addition, William S. Gosset, who published under the pseudonym "Student," used height and middle finger lengths from 3000 convicts to generate random samples from two correlated normal distributions. With such sampling, he studied the correlation coefficient and certain parameters of the t distribution [Student 1908].

What is not so well known is that the Italian physicist Enrico Fermi, in the early 1930s, used statistical sampling methods, before the name Monte Carlo had been coined, in his studies of the slowing down of neutrons. According to Emilio Segrè, Fermi's student and collaborator, Fermi wowed his Roman colleagues with his uncannily accurate predictions of their experimental results, which, unknown to them, were obtained by statistical sampling methods [Metropolis 1987]. The fact that Fermi had no computers to generate his random numbers and had to do his calculations with mechanical devices or even in his head, makes his accomplishments all the more remarkable. However, Fermi never published the results of these Monte Carlo calculations and, as a consequence, he often is neglected in the historical record of Monte Carlo.

Monte Carlo became a practical analysis technique only in the mid-twentieth century, with the advent of digital computers. Once one could generate pseudorandom numbers and compute long summations, it became evident that the Monte Carlo formalism was a great tool for estimating the behavior of neutrons and other radiation particles. Neutrons were of interest in the mid-twentieth century because of the effort to construct nuclear weapons. Just as the space program 20 years later gave great impetus to computing and materials research, so too the nuclear weapons program gave great impetus to the Monte Carlo method as a numerical procedure. In fact, it was during efforts to design and test the hydrogen bomb that Monte Carlo got its name.

Nicholas Metropolis gives a fascinating account [Metropolis 1987] of the contributions of Stanislaw Ulam, John von Neumann, and himself in resuscitating statistical mechanics methods using the Electronic Numerical Integrator and Computer (ENIAC), one of the first electronic computers. In 1945, von Neumann suggested to Metropolis that he develop a computational model to test the ENIAC. Metropolis, Stan Frankel, and Anthony Turkevich produced a model related to the hydrogen bomb, then being worked on by Edward Teller and others, and this model was run on the ENIAC. Ulam attended a session in early 1946 in which Metropolis presented the results of the ENIAC calculations. Ulam then suggested to von Neumann that the ENIAC could be used to conduct more general studies using statistical methods. All three (von Neumann, Metropolis, and Ulam) were gifted mathematicians who quickly realized the value of this suggestion. In early 1947, von Neumann formally proposed (in a letter to his superior Robert Richtmyer) that statistical methods be applied to solving neutron diffusion problems related to nuclear bomb research. It was Metropolis who, well aware of Ulam's interest in games of chance (for which the capital of Monaco, Monte Carlo, was well known), suggested the name *Monte Carlo*, and he and Ulam wrote a paper titled "The Monte Carlo Method" [Metropolis and Ulam 1949]. Several Monte Carlo calculations were performed on the ENIAC (after it had been moved to Aberdeen, MD, in 1947) and the results were so useful that a widespread interest in Monte Carlo methods was born.

In fact, the name *Monte Carlo* is most appropriate. In the games of chance popular at Monte Carlo and other gambling centers, the "house" assures itself, by a very small bias in the rules of each game, that it will win when averaged over a large number of players. However, any individual player also retains a chance to win in any particular game, hence providing the motivation and allure to play. In a Monte Carlo simulation, an individual history can go anywhere in the problem space, teasing out all the possibilities and their consequences; however, the investigator is assured that if enough histories are run (a large enough *sample* is chosen) and the game is played according to the rules, then the procedure will converge toward the desired result.

Within a few years, Herman Kahn wrote a research memorandum [Kahn 1956] that formally discussed the methods of Monte Carlo, including straightforward and rejection sampling, various probability distributions, evaluation of integrals, and variance reduction. Soon, several other books at least touched on Monte Carlo as a means of numerical integration. One book [Cashwell and Everett 1959] provided a practical manual on the use of Monte Carlo methods, and another [Hammersley and Handscomb 1964] discussed the use of Monte Carlo not only in radiation transport applications but

also in statistical mechanics, polymer science, the solution of linear operator equations, and so forth. The book *The Monte Carlo Method* [Shreider 1966] was translated from Russian by Tee and published in 1966.

A number of early monographs, chapters, or papers deal with Monte Carlo applications in radiation transport. Some, written for specialists in nuclear reactor calculations, include Goertzel and Kalos [1958], Kalos et al. [1968], and Spanier and Gelbard [1969]. More general treatments are available in the works of Bauer [1958], Hammersley and Handscomb [1964], Halton [1970], Carter and Cashwell [1975], Kalos and Whitlock [1986], and Lux and Koblinger [1991]. The compilation edited by Jenkins et al. [1988] considers coupled photon and electron transport. Practical information also can be obtained from the manuals for well-developed Monte Carlo computer codes. Especially recommended are those for the EGS4 code [Nelson et al. 1985] and its related EGSnrc code [Kawrakow et al. 2010], the GEANT4 toolkit [Agostinelli et al. 2003], and the MCNP code [Forster et al. 2004]. Other references for these and other widely used Monte Carlo codes are provided in Appendix D along with brief descriptions of the codes.

Other general works on Monte Carlo include Rubinstein [1981], Gentle [1998], Mooney [1997], Landau and Binder [2002], and Fishman [2003]. Markov Chain Monte Carlo is discussed in detail by Gilks et al. [1996]. As the twenty-first century emerged, Monte Carlo was ubiquitous in simulating nuclear and high-energy physics problems, weather, cosmological models, business and finance, transportation, and many other phenomena. The literature and the World Wide Web are replete with articles, books, and reports on the theory and application of Monte Carlo methods.

1.3 Monte Carlo as Quadrature

Our exploration of the Monte Carlo method begins by considering continuous functions of a single variable. Later, it is shown how the methodology developed here can be applied to discrete functions and to functions of several variables. Consider the function $z(x)$, which is dependent on a stochastic (or random) variable x . Its mean, or expected value,² is

$$\langle z \rangle = \int_a^b z(x) f(x) dx, \quad (1.8)$$

where³ $f(x)dx$ is the probability the random variable x has a value within dx about x . Here, it is assumed that values of x are in the range $[a, b]$. The Buffon needle problem of Example 1.2 is precisely of this form. In Eq. (1.6), $f(x) = (1/D)$ and the function $z(x) = [s_1(x) + s_2(x)]/(2\pi L)$, the expected value of which is the probability that the needle intersects a line when its end is a distance x from one line.

² The expected value $\langle z \rangle$ is sometimes denoted by $E(z)$, although the former notation is generally used in this book.

³ The function $f(x)$, which is called a *probability density function* (PDF), is discussed more thoroughly in Chapter 2.

A vast number of physical problems involve the evaluation of integral expressions such as Eq. (1.8) (or its multivariable or discrete analogs). Ideally, the right side of Eq. (1.8) can be evaluated analytically, as was done for Buffon's needle problem in Example 1.2 for $D > L$. However, in many situations, the integrand is too complex to allow analytic evaluation. Even more frequently, an explicit expression for the integrand is unknown! In both of these cases, Monte Carlo techniques can be used with great effectiveness to *estimate* the value of $\langle z \rangle$.

The evaluation of $\langle z \rangle$ from Eq. (1.8) is far from the most complex problem to which Monte Carlo can be applied. However, it serves as a useful simple problem that allows many of the essential features of Monte Carlo to be demonstrated, discussed, and understood. It is shown later how Monte Carlo can be applied to problems with much more complexity than that of Eq. (1.8).

In this section, the case is considered in which the form of the integral of Eq. (1.8) is known but the value of the integral is too difficult to evaluate analytically. In such cases, the integral must be evaluated using numerical techniques. Perhaps the most frequently used numerical technique is to apply *numerical quadrature* (or just quadrature, for short). Quadrature is nothing more than a numerical prescription for approximating integrals by finite summations. A quadrature scheme estimates the value of $\langle z \rangle$ by constructing a summation of weighted evaluations of the integrand, namely,

$$\langle z \rangle \simeq \sum_{i=1}^N w_i z(x_i) f(x_i), \quad (1.9)$$

where the x_i are called the quadrature *nodes* or *abscissas* and the w_i are called the *weights* of the quadrature scheme.

The very definition of a Riemann integral is expressed in this form. Break the interval $[a, b]$ into N contiguous subintervals, the i th subinterval width being denoted Δx_i . Identify x_i as the midpoint of the i th subinterval. Then the integral

$$I = \int_a^b z(x) f(x) dx \quad (1.10)$$

is defined (in the Riemann sense) as

$$I = \lim_{N \rightarrow \infty} \sum_{i=1}^N z(x_i) f(x_i) \Delta x_i. \quad (1.11)$$

Common quadrature schemes, such as Gauss-Legendre quadratures, are based on approximating functions by polynomials and their effectiveness depends on the forms of the functions $f(x)$ and $z(x)$ in Eq. (1.10). In such quadrature schemes, the abscissas and weights are prescribed and, thus, known beforehand.

Monte Carlo is a technique for estimating expected values, but because these can be expressed as integrals and because almost any integral can be interpreted as an expectation, Monte Carlo can be used to estimate the values of definite integrals.⁴

⁴ The simple estimation of integrals by Monte Carlo has sometimes been called *deterministic* Monte Carlo to distinguish it from *probabilistic* or *analog* Monte Carlo (see Section 1.4).

In the Monte Carlo approach to numerical quadrature, the abscissas are chosen randomly according to the probability density function (PDF) $f(x)$ and the expected value is estimated from a sum such as in Eq. (1.9). Explicitly, the straightforward Monte Carlo quadrature scheme proceeds as follows:

- generate N values x_i of the random variable x from the PDF $f(x)$;
- define the quadrature abscissas as the sampled values x_i ; and
- form the arithmetic average of the corresponding values of $z(x_i)$, i.e.,

$$\bar{z} \equiv \frac{1}{N} \sum_{i=1}^N z(x_i). \quad (1.12)$$

Note that the terms in the summation do not include the PDF. In effect, the weight factors in this basic Monte Carlo scheme are of the form $w_i = 1/(Nf(x_i))$. Later it is shown that $\lim_{N \rightarrow \infty} \bar{z} = \langle z \rangle$.

Example 1.3 Buffon's Needle Problem by Monte Carlo Quadrature

Consider the analysis of Buffon's needle problem given in Example 1.2. Suppose the integral of Eq. (1.6) could not be evaluated analytically. This integral is of the form of Eq. (1.8) with $a = 0$, $b = D$, $f(x) = 1/D$, and

$$z(x) = \frac{s_1(x) + s_2(x)}{2\pi L}, \quad (1.13)$$

where $s_1(x)$ and $s_2(x)$ are given in Example 1.2. The probability that a needle intersects a line is, from Eqs. (1.6) and (1.12),

$$P_{cut} \equiv \langle z \rangle \simeq \bar{z} = \frac{1}{N} \sum_{i=1}^N z(x_i), \quad (1.14)$$

where the x_i are uniformly distributed over the interval $(0, D)$. For the special case that $D = 2L$, the probability of line intersection $P_{cut} = 1/\pi \simeq 0.318310$. The value of $\bar{z}$ and the percent error are shown in Fig. 1.4 as a function of the number of samples N used to evaluate the right side of Eq. (1.14).

Now suppose one has to evaluate the integral

$$I = \int_a^b g(x) dx, \quad (1.15)$$

which is not in the form of an expected value. Such an integral can be converted into the same form as Eq. (1.8), for example, by rewriting the integral as

$$I = (b - a) \int_a^b \frac{g(x)}{b - a} dx = (b - a) \int_a^b g(x) f(x) dx, \quad (1.16)$$

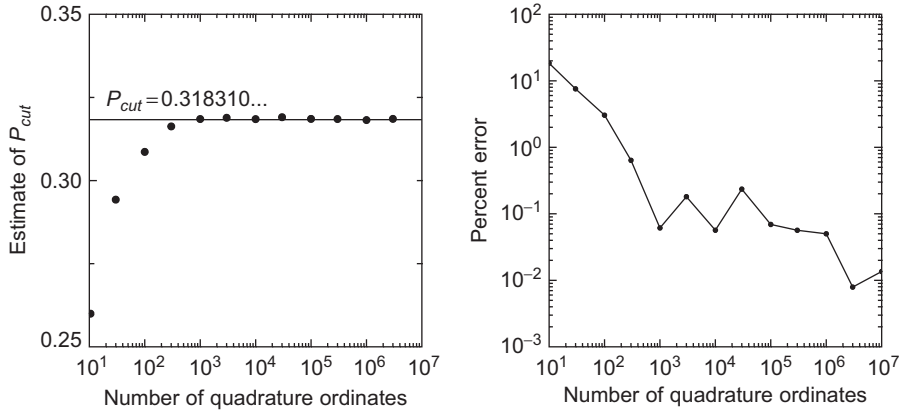


Figure 1.4 Evaluation of Buffon's needle intersection probability P_{cut} using Monte Carlo quadrature as described in [Example 1.3](#) (left). The corresponding absolute percentage error is shown in the right figure.

where $f(x) \equiv 1/(b-a)$ is, for reasons demonstrated in the next chapter, a PDF on the interval $[a, b]$. Then a Monte Carlo quadrature evaluation gives

$$I \simeq \bar{I} = \frac{b-a}{N} \sum_{i=1}^N g(x_i), \quad (1.17)$$

where the x_i are sampled from $f(x) = 1/(b-a)$, $a < x < b$.

[Equation \(1.8\)](#) is, in some sense, just a special case of a more general “expectation.” In general, the expected value of a function $z(x)$ with respect to a function $h(x)$ can be expressed as follows:

$$\langle z \rangle = \int_a^b z(x)h(x) dx / \int_a^b h(x) dx. \quad (1.18)$$

For instance, the x -coordinate of the center of gravity of a suitably symmetric body can be computed as

$$\langle x \rangle = \int_a^b xm(x) dx / \int_a^b m(x) dx, \quad (1.19)$$

where $m(x) dx$ is the mass in the interval dx about x .

Equations such as [Eq. \(1.18\)](#) can also be written in the form of [Eq. \(1.8\)](#) by a simple normalization procedure. First, let

$$M = \int_a^b h(x) dx. \quad (1.20)$$

Then $\langle z \rangle$ can be written as

$$\langle z \rangle = \int_a^b z(x) \frac{h(x)}{M} dx \equiv \int_a^b z(x) f(x) dx, \quad (1.21)$$

which is in the form of Eq. (1.8) and thus can be estimated by Monte Carlo. This procedure may seem forced, because, if the integral of Eq. (1.20) could be evaluated without Monte Carlo, then the integral in Eq. (1.21) most likely could also be evaluated without Monte Carlo. However, this example merely is intended to demonstrate that Monte Carlo can be applied, in principle, to the evaluation of any definite integral. Further, $\langle z \rangle$ could be estimated by using Monte Carlo to estimate both integrals in Eq. (1.18) separately, i.e., M could be evaluated by the basic Monte Carlo estimator

$$M \simeq \overline{M} = \frac{b-a}{N} \sum_{i=1}^N h(x_i), \quad (1.22)$$

where the x_i are sampled from the PDF $1/(b-a)$, and then $\langle z \rangle$ could be estimated as

$$\langle z \rangle \simeq \frac{1}{N} \sum_{i=1}^N z(x_i), \quad (1.23)$$

where the x_i are sampled from $f(x) = h(x)/\overline{M}$. Thus, Monte Carlo can be applied to a great many problems indeed.

Later, it is shown that Monte Carlo quadrature is not limited to estimating single integrals. In fact, Monte Carlo demonstrates its superiority to many other quadrature schemes when the integrals are highly multidimensional. It will be shown in Chapter 2 that, regardless of the number of dimensions m , the precision of the Monte Carlo estimator varies at worst as $N^{-1/2}$, where N is the number of histories. Traditional quadratures, with prescribed abscissas and weights, have precisions that vary as $N^{-1/m}$ [Mosegaard and Sambridge 2002]. Hence, Monte Carlo is superior to other quadrature schemes for integrals of dimension greater than two! Further, there are a host of ways, called *variance reduction techniques*, to reduce the Monte Carlo estimation errors even more rapidly (see Chapter 5).

Thus, it is seen that the variable x really need not be stochastic in order to apply the Monte Carlo quadrature procedure. As long as f has the properties of a PDF, then the integral of Eq. (1.8) is a number that can be treated as if it is an average or expected value. If the integral of Eq. (1.8) is not easily evaluated analytically, then quadrature is generally used. Monte Carlo provides a powerful form of quadrature, as is demonstrated in subsequent chapters.

1.4 Monte Carlo as Simulation

It is not necessary that a problem be explicitly posed as an integral in order for Monte Carlo to be applied. For example, Laplace [1786] and Hall [1873] showed that the value of π could be estimated by a “game” in which needles are dropped randomly

on a striped grid. This implementation of Monte Carlo is sometimes called *simulation*. In truth, all implementations of Monte Carlo can be viewed as simulations; but here is stressed the difference between problems formulated and identified explicitly as integrals from others in which the integral formulation is hidden and unspecified.

Consider a rectangular map that shows a country with an irregular boundary and of which the area inside the enclosed irregular boundary is sought. One way to estimate the area of this shape is to randomly throw darts at the map. The ratio of the number of darts that are within the region of interest to the total number that hit anywhere on the map, multiplied by the area of the map (which is trivially calculated), is a reasonable estimate of the area of the region. Of course, in a fundamental way, calculation of an area is an estimation of an integral, but, by throwing darts, the area can be estimated without ever formulating the problem as an integral. It is also interesting to note that in this example a regular grid of points could be overlayed on the map and the area estimated by counting the number of grid points within the country's border. The area estimate improves the finer the grid used. This procedure is an example of simulation without sampling and is akin to *systematic* sampling, which is discussed in Chapter 5.

There are numerous other examples of quantities that can be estimated by simulating the underlying processes without explicitly formulating an integral. For example, the number of "counts" that a radiation detector registers from cosmic radiation can be estimated by simulating the process of cosmic-ray interactions with the atmosphere, the laboratory, and the detector. In such an application, some assumption must be made about the incident distribution of cosmic rays, perhaps based on balloon and satellite measurements, but the rest of the problem can be simulated without having to go into a laboratory and make an actual measurement. In this sense, Monte Carlo is often viewed as a virtual experiment conducted on a computer (or by some other means, such as dropping needles on a striped plane or throwing darts at a map).

Similarly, high-energy physicists often design their specialized detectors using Monte Carlo simulation of the processes involved in accelerator interactions. Why would one want to simulate the process rather than just make the measurement? Well, high-energy physics detectors are very expensive to build. One would rather run some relatively inexpensive simulations to be assured that a detector will behave as hoped and that the detector design is optimal before investing large sums of money for its construction. Further, once the detector is built, Monte Carlo simulation can be used to interpret the data that the detector collects. In these applications one rarely sets up the problem as an expectation; rather, one simulates the various processes that the interacting particles experience. The ways this is done are discussed in later chapters.

One also can simulate various models for galaxy formation, or expected swings in the stock market, or traffic flow patterns on highways, or many other processes. In each case, of course, one must start with models of the processes involved (e.g., the laws of gravity) and often one must make assumptions (e.g., there is a Gaussian distribution of speeds about the posted speed limit by cars on a certain highway). The Monte Carlo estimates developed are only as good as the underlying models and assumptions. Nevertheless, these estimates are often of great utility to researchers, analysts, practitioners, planners, and others. For instance, Monte Carlo is now being used in treatment planning for cancer patients because the Monte Carlo dose estimates are, in general, superior to those obtained in other ways.

In such Monte Carlo simulations the quantity of interest x usually has a distribution $h(x)$ of possible values as determined by the underlying processes. Of course, if one knew $h(x)$ a priori, the desired expected value could be evaluated readily by quadrature, i.e., $\langle x \rangle = \int x h(x) dx / \int h(x) dx$. However, in the vast majority of problems, $h(x)$ is unknown. But if the underlying processes are known, then a Monte Carlo simulation can be used not only to estimate the expected or average value of the quantity, $\langle x \rangle$, but to also estimate the distribution $h(x)$ itself! An example follows.

Example 1.4 Buffon's Needle Problem with Monte Carlo Simulation

From [Example 1.2](#) it was found that the probability P_{cut} a dropped needle intersects a line is given by [Eq. \(1.7\)](#), which has the form

$$P_{cut} = \langle z \rangle = \int_0^D z(x) f(x) dx. \quad (1.24)$$

The end of the needles fall uniformly between the two lines, so that $f(x) = 1/D$. However, suppose we had no idea of the probability $z(x)$ that a needle, the end of which is a distance x from a line, intersects a line. A Monte Carlo simulation could then be used to estimate the function $z(x)$ and then find its average value, which is an estimate of P_{cut} .

The simulation begins by first dividing the distance between the two lines into m contiguous intervals, each with width $\Delta x = D/m$. A scoring bin is created for each interval. Then N needle drops are simulated, with each needle history calculated as follows:

- randomly select the needle's end location x_i uniformly in $(0, D)$;
- randomly select the angle θ_i between the needle and the x -axis uniformly in $(0, 2\pi)$;
- calculate the x -coordinate of the needle's other end $x_j^o = x_i + L \cos \theta_i$; and
- if $x_j^o < 0$ or $x_j^o > D$, then add a count to the scoring bin for the interval containing x_i^o .

After the simulation is completed, the function $z(x)$ can be estimated from the number of counts in each scoring bin. Bin k tallies the number n_k of needles with an end that fell between x_{k-1} and x_k and that intersected a line. Because the needle ends are uniformly distributed in $(0, D)$, the number starting in each bin is $n_o = N/m = N\Delta x/D$. Thus, if $x_{k-1/2}$ is the midpoint location for the bin, $z(x_{k-1/2}) \simeq n_k/n_o$. Two such estimated $z(x)$ distributions, for the case $D = 2L$, are shown in [Fig. 1.5](#) together with the exact distribution given by [Eq. \(1.13\)](#). The intersection probability

$$P_{cut} = \langle z \rangle = \int_0^L z(x) \frac{dx}{D}$$

is then estimated by

$$\bar{z} = \frac{1}{D} \sum_{k=1}^N \Delta x z(x_{k-1/2}) \simeq \frac{1}{D} \sum_{k=1}^m \Delta x \frac{n_k}{n_o} = \frac{1}{N} \sum_{k=1}^m n_k. \quad (1.25)$$

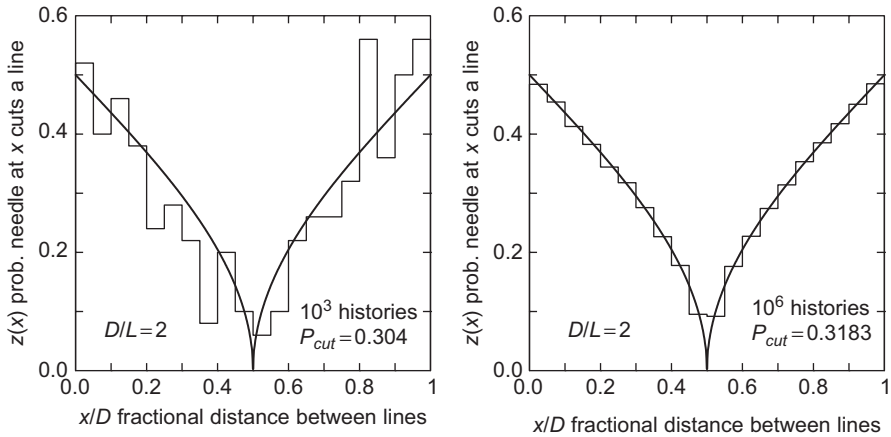


Figure 1.5 Monte Carlo estimates (histograms) of the Buffon intersection probability function $z(x)$ for 10^3 (left) and 10^6 (right) simulated needle drops. The heavy line shows the exact $z(x)$ of Eq. (1.13). Results are for the case $D = 2L$. The correct value is $P_{cut} = 0.318310$.

1.5 Preview of Things to Come

The overall purpose of this book is to explore, in some detail, many of the facets and variations of the Monte Carlo method. Key quantities (such as probability density functions, cumulative distribution functions, population means and variances, and sample means and variances) are defined in Chapter 2, and the two fundamental theorems (the law of large numbers and the central limit theorem) that form the theoretical basis of the Monte Carlo method are discussed there. Chapter 3 summarizes some of the methods for generating what are called pseudorandom numbers, which are used repeatedly in Monte Carlo analyses. Truly random numbers cannot be generated using deterministic algorithms; however, pseudorandom numbers that exhibit certain useful statistical properties can. The concepts of “sampling” and “scoring” are explained, for both discrete and continuous random variables, in Chapter 4. Monte Carlo calculations can be computationally expensive, and thus Chapter 5 reviews many of the “variance reduction” procedures that frequently are used to improve the efficiency of Monte Carlo simulations. These first five chapters present the introductory material that forms the general basis for use of Monte Carlo in a variety of applications.

Chapter 6 introduces Markov chain Monte Carlo (MCMC), which provides an alternative method for sampling distributions; MCMC often is applied when attempting to sample multidimensional and complex distributions, such as can arise in statistical inference and decision theory. Chapter 7 presents an overview of techniques for solving inverse and optimization problems using Monte Carlo. These inverse Monte Carlo techniques include both iterative simulation procedures and the noniterative “symbolic” Monte Carlo method. Chapter 8 reviews the use of Monte Carlo methods

for solving linear operator equations. Chapter 9 introduces the essential concepts of neutral particle transport, which are needed in Chapter 10 where Monte Carlo simulation of neutral particle transport is discussed. The field of radiation transport was chosen, in order to demonstrate applications of Monte Carlo, for two reasons. First, it is the field that sparked the greatest interest and growth in Monte Carlo (and indeed radiation transport specialists gave the Monte Carlo method its name, although the method had been used sparingly by others previously). Second, the authors are familiar with this particular field. However, Monte Carlo methods are widely used in the physical sciences, the social sciences, many engineering fields, medicine, finance, business, transportation, and many other disciplines.

Five appendices are provided. Appendix A discusses several of the commonly used probability distributions. Appendix B discusses the subtle differences between the weak and strong forms of the law of large numbers and, for the purists, provides a proof of the weak law of large numbers. Appendix C provides a proof of the central limit theorem. Appendix D gives summaries of nine of the general-purpose radiation transport simulation codes. Appendix E is included in order to identify the pseudorandom number generator that was used to solve the example problems. Readers may wish to use this pseudorandom number generator, with the specific initial seed 73907, in order to duplicate the results obtained in the example problems.

1.6 Summary

Monte Carlo is, in essence, a very powerful form of quadrature or numerical integration, in which finite summations are used to estimate definite integrals. Although Monte Carlo is inherently involved with the concept of probability, it can be applied, with much success, to problems that have no apparent connection with probabilistic phenomena. The methods are based on rigorous mathematics that began with the first statement of the law of large numbers, as enunciated by Jacob Bernoulli over 300 years ago. Monte Carlo is particularly valuable when considering multidimensional integrals, where it generally outperforms traditional quadrature methods. Significantly, it also can be applied to a great variety of problems for which the integral formulation is not posed explicitly. Often, the complex mathematics needed in many analytical applications can be avoided entirely by simulation. Thus, Monte Carlo methods provide extremely powerful ways to address realistic problems that are not amenable to solution by analytic techniques. Today, with the widespread availability of powerful and inexpensive computers, Monte Carlo methods are widely used in almost every discipline that requires quantitative analysis.

Bibliography

- AGOSTINELLI, S. ET AL. 2003, "GEANT4—A Simulation Toolkit," *Nucl. Instrum. Meth. B*, **506**, 250–303.
- BAUER, W.F. 1958, "The Monte Carlo Method," *J. Soc. Ind. Appl. Math.*, **6**(4), 438–451.

- BERNOULLI, J. 1713, *Ars Conjectandi*, Impensis Thurnisiorum, Fratrū, Basilea. This posthumous Latin work was subsequently published in German as *Wahrscheinlichkeitsrechnung* in two volumes by Englemann, Leipzig, 1899.
- BUFFON, G.-L. LECLERC 1777, "Essai d'Arithmétique Morale," *Supplément à l'Histoire Naturelle*, vol. 4.
- CARTER, L.L. AND E.D. CASHWELL 1975, *Particle-Transport Simulation with the Monte Carlo Method*, TID-26607, National Technical Information Service, U.S. Department of Commerce, Springfield, VA.
- CASHWELL E.D. AND C.J. EVERETT 1959, *A Practical Manual on the Monte Carlo Method for Random Walk Problems*, Pergamon Press, New York.
- DUNN, W.L. AND J.K. SHULTIS 2009, "Monte Carlo Methods for Design and Analysis of Radiation Detectors," *Radiat. Phys. Chem.*, **78**, 852–858.
- FISHMAN, G.S. 2003, *Monte Carlo Concepts, Algorithms, and Applications*, Springer, New York (corrected printing).
- FORSTER, R.A., L.J. COX, R.F. BARRETT, T.E. BOOTH, J.F. BRIESENMEISTER, F.B. BROWN, J.S. BULL, G.C. GEISLER, J.T. GOORLEY, R.D. MOSTELLER, S.E. POST, R.E. PRAEL, E.C. SELCOW, AND A. SOOD 2004, *MCNPTM Version 5*, *Nucl. Instrum. Meth. B*, **213**, 82–86.
- GENTLE, J.E. 1998, *Random Number Generation and Monte Carlo Methods*, Springer, New York.
- GILKS, W.R., S. RICHARDSON, AND D.J. SPIEGELHALTER, Eds., 1996, *Markov Chain Monte Carlo in Practice*, Chapman & Hall, New York.
- GOERTZEL, G. AND M.H. KALOS 1958, "Monte Carlo Methods in Transport Problems," *Progress in Nuclear Energy II*, Series 1, Physics and Mathematics, D.J. Hughes, J.E. Sanders, and J. Horowitz, Eds., 315–369, Pergamon Press, New York.
- HALL, A. 1873 "On an Experimental Determination of π ," *Messeng. Math.*, **2**, 113–114.
- HALTON, J.H. 1970, "A Retrospective and Prospective Survey of the Monte Carlo Method," *SIAM Rev.* **12**(1), 1–63.
- HAMMERSLEY, J.M. AND D.C. HANDSCOMB 1964, *Monte Carlo Methods*, Spottiswoode, Ballantyne & Co., London.
- JENKINS, T.M., W.R. NELSON, AND A. RINDI 1988, *Monte Carlo Transport of Electrons and Photons*, Plenum Press, New York.
- KAHN, H. 1956, *Applications of Monte Carlo*, AECU-3259, United States Atomic Energy Commission, Rand Corporation, Santa Monica, CA.
- KALOS, M.H., F.R. NAKACHE, AND J. CELNIK 1968, "Monte Carlo Methods in Reactor Computations," in *Computing Methods in Reactor Physics*, H. Greenspan, C.N. Kelber, and D. Okrent, Eds., Gordon and Breach, New York.
- KALOS, M.H. AND P.A. WHITLOCK 1986, *Monte Carlo Methods*, Wiley, New York.
- KAWRAKOW, I., E. MAINEGRA-HING., D.W.O. ROGERS, F. TESSIER, AND B.R.B. WALTERS 2010, "The EGSrc Code System: Monte Carlo Simulation of Electron and Photon Transport," NRCC Report PIRS-701, National Research Council of Canada, Ottawa, Canada.
- KELVIN, W. 1901, "Nineteenth Century Clouds over the Dynamical Theory of Heat and Light," *Phil. Mag.*, **2**(6), 1–40.
- LANDAU, D.P. AND K. BINDER 2002, *A Guide to Monte Carlo Simulations in Statistical Physics*, Cambridge University Press, Cambridge (originally published in 2000 and reprinted in 2002).
- LAPLACE, P.S. 1786, *Théorie Analytique des Probabilités*, Livre 2, 356–366, contained in *Oeuvres Complètes de Laplace, de L'Académie des Sciences*, **7**, part 2.

- LUX, I. AND L. KOBLINGER 1991, *Monte Carlo Particle Transport Methods: Neutron and Photon Calculations*, CRC Press, Boca Raton, FL.
- MAISTROV, L.E. 1974, *Probability Theory, A Historical Sketch*, Translated and Edited by S. Kotz, Academic Press, New York.
- METROPOLIS, N. 1987, "The Beginning of the Monte Carlo Method," *Los Alamos Science*, Special issue, 125–130.
- METROPOLIS, N. AND S. ULAM 1949, "The Monte Carlo Method," *J. Am. Stat. Assoc.*, **44**, 335–341.
- MOONEY, C.Z. 1997, *Monte Carlo Simulation*, Sage Publications, Thousand Oaks, CA.
- MOSEGAARD, K. AND M. SAMBRIDGE 2002, "Monte Carlo Analysis of Inverse Problems," *Inverse Probl.*, **18**, R29–R54.
- NELSON, W.R., H. HIRAYAMA, AND D.W.O. ROGERS 1985, *The EGS4 Code System*, SLAC-265 UC-32, Stanford Linear Accelerator Center, Stanford University, Palo Alto, CA.
- RAO, C.R. 1989, *Statistics and Truth*, Council of Scientific & Industrial Research, New Delhi, India.
- RAYLEIGH, L. 1899, "On James Bernoulli's Theorem in Probabilities," *Philos. Mag.* **47**, 246–251.
- RUBINSTEIN, R.Y. 1981, *Simulation and the Monte Carlo Method*, John Wiley & Sons, New York.
- SHREIDER, Y.A., Ed. 1966, *The Monte Carlo Method*, Pergamon Press, New York.
- SHULTIS, J.K. AND R.E. FAW 2000, *Radiation Shielding*, American Nuclear Society, La Grange Park, IL.
- SPANIER, J. AND E.M. GELBARD 1969, *Monte Carlo Principles and Neutron Transport Problems*, Addison-Wesley, Reading, MA.
- STUDENT 1908, "The Probable Error of a Mean," *Biometrika*, **6**, 1–25; also "Probable Error of a Correlation Coefficient," *Biometrika*, **6**, 302–310.
- TROTTER, H.F. AND J.W. TUKEY 1954, "Conditional Monte Carlo for Normal Samples," in *Symposium on Monte Carlo Methods*, University of Florida, H.A. Meyer, ed., John Wiley, New York.
- WHEELER, J.A. 1998, *Geons, Black Holes & Quantum Foam*, W.W. Norton & Company, New York.

Problems

1. Find a small stick and measure its length L . Draw parallel lines on a large piece of paper at an equal spacing of $D > L$. Drop the stick from a height of several feet onto the paper 20 times, each time noting whether or not the stick crosses one of the lines. Use different orientations of the stick and start from different positions above the paper. Disregard any drops for which the stick does not end up on the paper. Calculate the success rate from

$$p = \frac{S}{N},$$

where S is the number of successes (times that the stick overlapped one of the lines) and $N = 20$ is the number of trials. (a) Use the data to estimate the value of π . (b) Repeat the process for another 20 trials and use the cumulative results to estimate π . Is the result now closer to the actual value?

2. Consider the function

$$u(x) = 6x - 1$$

and the PDF

$$f(x) = 1, 0 \leq x \leq 1.$$

(a) What is the expected value of u with respect to this PDF? (b) Use the following sequence of “pseudorandom” numbers to form the basic Monte Carlo estimate $\bar{u}$ of the mean value $\langle u \rangle$: {0.985432, 0.509183, 0.054991, 0.641903, 0.390188, 0.719905, 0.876302, 0.251175, 0.160938, 0.487301}.

3. Consider the definite integral

$$I = \int_{-4}^4 e^{-x^2/2} dx. \quad (1.26)$$

Use Monte Carlo to estimate the value of this integral, using the ten pseudorandom numbers given in problem 2.

4. Consider a horizontal beam 20 feet long. The mass distribution $m(x)$ along the beam varies with the distance x from the left end as

$$m(x) = 40 - 3x + 0.05x^2,$$

where x is in feet. (a) Use Eq. (1.19) to evaluate analytically the center of gravity $\langle x \rangle$ of this beam. (b) Then use the following pseudorandom numbers to form a basic Monte Carlo estimate, $\bar{x}$, of the x -coordinate of the center of gravity: {19.70864, 10.18366, 1.09982, 7.80376, 14.39810, 17.52604, 5.02350, 3.21876, 12.83806, 9.74602}.

5. In Buffon’s needle problem, the probability a randomly dropped needle cuts a grid line was shown in Example 1.1 to be

$$P_{cut} = \frac{L}{\pi D} \int_0^\pi \sin \theta \, d\theta.$$

Use Monte Carlo quadrature to evaluate this integral for the case $2L = D$ and compare your result with the analytical result.

6. Consider Buffon’s needle problem for the case that the needle length L is greater than the distance D between the stripes. Following an analysis similar to that of Example 1.2, show the probability a needle intersects a strip is given by,

$$P_{cut} = \frac{2}{\pi D} \left[D \cos^{-1} \frac{D}{L} + L - \sqrt{L^2 - D^2} \right].$$

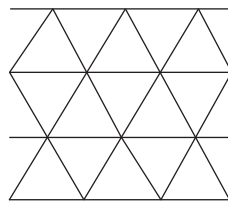
7. Consider an infinite rectangular grid of lines that form square cells with sides of length D . Onto this grid circular disks of radius R are dropped at random. Show that the probability P_{cut} that a disk lands on a grid line is

$$P_{cut} = \begin{cases} 4 \frac{R}{D} \left(1 - \frac{R}{D} \right), & 2R \leq D \\ 1, & 2R > D \end{cases}.$$

8. Consider Buffon’s needle problem in which n needles of length L are randomly dropped onto a square grid of lines with sides of length D . If n_1 are the number of needles cutting exactly one grid line and n_2 the number cutting two grid lines, show that for large n ,

$$\frac{n_1 + n_2}{n} \simeq \frac{1}{\pi} \frac{L}{D} \left(1 - \frac{L}{D} \right).$$

9. Consider a variant of Buffon's needle problem in which needles of length L are randomly dropped unto an equilateral triangular grid such as the one shown to the right. The perpendicular from a vertex to the opposite side has a length $D > L$. Derive an expression for the probability a randomly dropped needle cuts a grid line.



2 The Basis of Monte Carlo

*We all know that the concept of limit,
except in the mind of a dimwit,
is used to make sense
of things small or immense,
taming zero and the infinite.*

The Monte Carlo method is firmly based on two mathematical theorems: the *law of large numbers* and the *central limit theorem*. However, before these theorems are examined, some basic definitions and nomenclature must be established. The concepts of a random variable, a probability density function, a cumulative distribution function, marginal and conditional probabilities, population means and variances, and sample means and variances all must be understood. These are discussed in the following sections.

A variable is considered random (also called stochastic) if its value cannot be specified in advance of observing it. For instance, the maximum temperature in your town (as measured, for instance, by a certain thermometer at a certain location in town) varies from day to day and the maximum temperature for tomorrow can be predicted, based on detailed data analysis, but cannot be known exactly until tomorrow arrives. Similarly, it is assumed that a fair coin tossed fairly should yield heads 50% of the time and tails 50% of the time, but the exact outcome of the next toss cannot be predicted with certainty. In fact, most observable or measurable quantities show, *at some scale*, a random variation from one observation to the next. For example, a mint tries to make each coin of a given type have the same mass; but it is extremely unlikely that each coin has exactly the same number of atoms and, consequently, there is a very small but real random variation in mass among the coins produced by the mint. Nevertheless, as a practical matter, many variables can be treated as nonrandom. Thus, all coins from the mint can be treated as having the same mass macroscopically. Similarly, the position at which you hang a picture on a wall can be determined by measuring distance along the wall and then pounding a nail into the wall at a certain position. Once that position has been determined, it is “known,” and so the variable describing that position is not random; it can be specified now and be expected to be the same tomorrow. As another example, the time the sun will rise tomorrow can be predicted with very high accuracy and hence is considered nonrandom.

Many variables in nature, however, are random at a large scale. The exact time a radioactive nuclide will disintegrate cannot be predicted with certainty. (This is an example of a continuous random variable, because time to decay varies continuously between 0 and ∞ .) Also, the number of individual pieces of glass that will result if you drop a mirror on the floor cannot be specified in advance. (This is an example

of a discrete random variable because the number will be a finite integer.) A random “variable” also can be a vector of several individual random variables, just like position is a vector of three spatial variables. Each random variable has associated with it a probability density function, a cumulative distribution function, a population mean, a population variance, and other measures of the variable’s randomness.

2.1 Single Continuous Random Variables

Let x denote a single continuous random variable defined over some interval. The interval can be finite or infinite. The value of x on any observation cannot be specified in advance because the variable is random. Nevertheless, it is possible to talk in terms of probabilities. The notation $\text{Prob}\{x_i \leq X\}$ represents the probability that an observed value x_i will be less than or equal to some specified value X . More generally, $\text{Prob}\{E\}$ is used to represent the probability of an event E . The event E can be almost anything. For instance, event E can be the temperature T in Dallas exceeding 295 K on New Year’s Day, in which case one may write $\text{Prob}\{T > 295 \text{ K}\} = p$, where p is the estimated probability. Alternatively, event E might be the Chicago Cubs winning the World Series next year, or event E could be obtaining a measurement of 10^6 counts in a radiation detector under certain external conditions. Later, in Chapter 6, the concept of probability is examined in more detail when statistical inference and Markov chains are discussed. For now, only the essential concepts that are needed to implement Monte Carlo analyses are presented.

2.1.1 Probability Density Function

A *probability density function* (PDF) of a single stochastic variable is a function that has the following three properties: (1) it is defined on an interval $[a, b]$, where $b > a$, (2) it is non-negative on that interval, although it can be zero for some $x \in [a, b]$, and (3) it is normalized such that $\int_a^b f(x) dx = 1$. Here, a and b represent real numbers or infinite limits (i.e., $a \rightarrow -\infty$ and/or $b \rightarrow \infty$), and the intervals can be either closed or open.¹

A PDF is a *density* function, i.e., it specifies the probability *per unit of* x , so that $f(x)$ has units that are the inverse of the units of x . For a continuous random variable, $f(x)$ is not the probability of obtaining x (there are infinitely many values that x can assume and the probability of obtaining a single specific value is zero). Rather, the quantity $f(x) dx$ is the probability that a random sample x_i will assume a value within dx

¹ Strictly speaking, $x \in [a, b]$ implies that x can assume the values a and b and that f is defined for such values. However, the properties of PDFs also apply whether the intervals are closed, half open, or fully open. For simplicity, throughout this book $[a, b]$ is used in a generic sense to represent any of the following cases: the closed interval $[a, b]$, where both a and b are finite and the variable x can assume the values a and b ; the half-open intervals $(a, b]$ or $[a, b)$, or the fully open interval (a, b) , with a and b being either finite or tending to minus or plus infinity, respectively. From a probability standpoint, whether or not the endpoint is included or excluded is immaterial because the probability that x *exactly* equals the endpoint is zero.

about x . Often, this is stated in the form

$$f(x)dx = \text{Prob}\{x \leq x_i \leq x + dx\}. \quad (2.1)$$

2.1.2 Cumulative Distribution Function

The integral defined by

$$F(x) \equiv \int_a^x f(x') dx', \quad (2.2)$$

where $f(x)$ is a PDF over the interval $[a, b]$, is called the *cumulative distribution function* (CDF) of f . Note that, from this definition, a CDF has the following properties: (1) $F(a) = 0$, (2) $F(b) = 1$, and (3) $F(x)$ is monotone increasing, because f is always non-negative.

The CDF is a direct measure of probability. The value $F(x_i)$ represents the probability that a random sample of the stochastic variable x will assume a value between a and x_i , i.e., $\text{Prob}\{a \leq x \leq x_i\} = F(x_i)$. More generally,

$$\text{Prob}\{x_1 \leq x \leq x_2\} = \int_{x_1}^{x_2} f(x) dx = F(x_2) - F(x_1). \quad (2.3)$$

Later, in Chapter 4, what is meant by the term *random sample* is examined in more detail. For the present, it is sufficient to use the intuitive notion that a random sample is a value of x selected or measured according to the governing probability distribution.

2.1.3 Some Example Distributions

Three widely used distributions are presented here for illustration.

1. The *uniform*, or rectangular, PDF on the interval $[a, b]$ is given by

$$f(x) = \frac{1}{b-a}. \quad (2.4)$$

The CDF for this PDF is given by

$$F(x) = \int_a^x \frac{1}{b-a} dx' = \frac{x-a}{b-a}. \quad (2.5)$$

Note that either a or b (or both) can be negative but both must be finite. The uniform distribution is plotted in [Fig. 2.1](#) for the particular case $a = 15$ and $b = 25$.

2. The *exponential* PDF is given by

$$f(x) = f(x|\alpha) = \alpha e^{-\alpha x}, \quad (2.6)$$

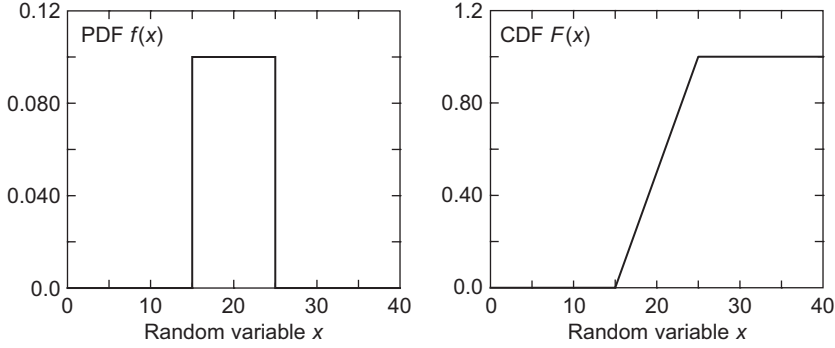


Figure 2.1 The PDF and CDF for the rectangular distribution on the interval $[15, 25]$.

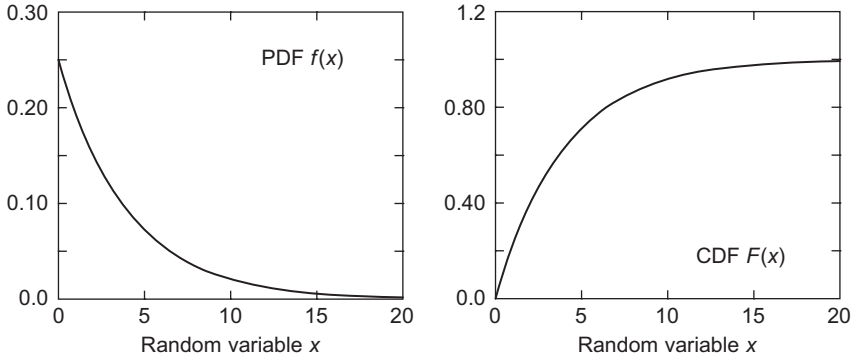


Figure 2.2 The PDF and CDF for the exponential distribution with parameter $\alpha = 0.25$.

which is defined on the interval $x \geq 0$ with a single parameter α , and its CDF is given by

$$F(x) = \int_0^x \alpha e^{-\alpha x'} dx' = 1 - e^{-\alpha x}. \quad (2.7)$$

The exponential distribution is plotted in Fig. 2.2 for the particular case $\alpha = 0.25$.

3. The *normal*, or *Gaussian*, PDF is given by

$$f(x) = f(x|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x - \mu)^2}{2\sigma^2}\right], \quad (2.8)$$

which is defined on the interval $[-\infty, \infty]$. In this case, the distribution has two parameters, the mean value μ and the standard deviation σ . Its CDF is given by

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x \exp\left[-\frac{(x' - \mu)^2}{2\sigma^2}\right] dx', \quad (2.9)$$

which must be evaluated numerically or from tabulations. This distribution is plotted in Fig. 2.3 for the particular case $\mu = 20$ and $\sigma = 5$.

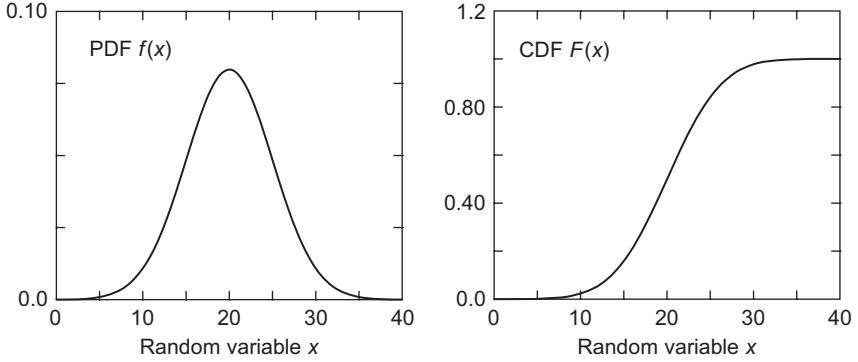


Figure 2.3 The PDF and CDF for the normal or Gaussian distribution with $\mu = 20$ and $\sigma = 5$.

These and several other commonly encountered probability distributions are summarized in Appendix A.

Note that f and F are defined only in the interval $[a, b]$. For finite values of a and b , they are not explicitly defined outside the interval. However, by setting them to constants outside $[a, b]$, $f(x)$ and $F(x)$ can be defined for all values of x . For instance, the uniform distribution of Eqs. (2.4) and (2.5) can be alternatively written in the form

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & x > b \text{ or } x < a \end{cases} \quad (2.10)$$

and

$$F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x \leq b. \\ 1, & x > b \end{cases} \quad (2.11)$$

This interpretation is actually used in Fig. 2.1, where the CDF is shown as zero for $x < 15$ and unity for $x > 25$. Using this convention, a CDF can be defined as $F(x) = \int_{-\infty}^x f(x') dx'$, because $f(x) = 0, x \leq a$.

2.1.4 Population Mean, Variance, and Standard Deviation

Two important measures of any PDF $f(x)$ are its mean μ and variance σ^2 . The mean is the expected or average value of x and is defined as

$$\langle x \rangle \equiv E(x) \equiv \mu(x) \equiv \int_a^b x f(x) dx. \quad (2.12)$$

The variance describes the spread of the random variable x from the mean and is defined as

$$\sigma^2(x) \equiv \langle [x - \langle x \rangle]^2 \rangle = \int_a^b [x - \langle x \rangle]^2 f(x) dx. \quad (2.13)$$

This result can be written equivalently as follows:

$$\begin{aligned} \sigma^2(x) &= \int_a^b [x^2 - 2x\langle x \rangle + \langle x \rangle^2] f(x) dx, \\ &= \int_a^b x^2 f(x) dx - 2\langle x \rangle \int_a^b x f(x) dx + \langle x \rangle^2 \int_a^b f(x) dx. \end{aligned} \quad (2.14)$$

By Eq. (2.12) the first integral is just the expected value of x^2 (i.e., $\langle x^2 \rangle$), the second integral is just $\langle x \rangle$, and because $f(x)$ is a PDF, the third integral is unity. Thus, Eq. (2.14) can be written as

$$\sigma^2(x) = \langle x^2 \rangle - \langle x \rangle^2. \quad (2.15)$$

The square root of the variance is called the *standard deviation* σ . Thus, one has the trivial result that the standard deviation is $\sigma(x) = \sqrt{\sigma^2(x)}$.

Consider a function $z(x)$, where x is a random variable described by a PDF $f(x)$. The function $z(x)$ itself is a random variable. Thus, the *expected* or mean value of $z(x)$ is defined as

$$\langle z \rangle \equiv \mu(z) \equiv \int_a^b z(x) f(x) dx. \quad (2.16)$$

The variance of a random variable $z(x)$, by analogy to Eq. (2.13), is defined as

$$\sigma^2(z) = \langle [z(x) - \langle z \rangle]^2 \rangle = \int_a^b [z(x) - \langle z \rangle]^2 f(x) dx, \quad (2.17)$$

which, by the same reduction used for Eq. (2.15), is simply

$$\sigma^2(z) = \langle z^2 \rangle - \langle z \rangle^2, \quad (2.18)$$

and the standard deviation is $\sigma(z) = \sqrt{\sigma^2(z)}$.

Example 2.1 Statistics of Radioactive Decay

Radionuclides decay spontaneously to more stable nuclides. The stochastic radioactive process is well described by the exponential distribution, and the *expected* number of radionuclides at time t is given by

$$n(t) = n_0 e^{-\lambda t},$$

where n_0 is the number of radionuclides at time $t = 0$. The parameter λ is called the *decay constant* and has a fixed value for each radionuclide species. The decay constant is the probability a radionuclide decays per unit differential time interval. The probability that a radionuclide will decay within dt about t is

$$[\text{probability it survives to time } t][\text{probability it decays in } dt] = (e^{-\lambda t})(\lambda dt).$$

Thus, the PDF for radioactive decay is $p(t) = \lambda e^{-\lambda t}$. From this, the mean lifetime of the radionuclides is given by

$$\langle t \rangle \equiv \int_0^\infty t p(t) dt = \lambda \int_0^\infty t e^{-\lambda t} dt = \lambda \left[\frac{e^{-\lambda t}}{(-\lambda)^2} (\lambda t - 1) \right]_0^\infty = \frac{1}{\lambda},$$

and the population variance of this process is given by

$$\sigma^2(t) = \int_0^\infty (t - \langle t \rangle)^2 p(t) dt = \langle t^2 \rangle - \langle t \rangle^2,$$

where

$$\langle t^2 \rangle = \int_0^\infty t^2 p(t) dt = \frac{2}{\lambda^2}.$$

Thus, the population standard deviation is

$$\sigma(t) = \sqrt{\frac{2}{\lambda^2} - \frac{1}{\lambda^2}} = \frac{1}{\lambda}.$$

2.1.5 Sample Mean, Variance, and Standard Deviation

The *sample mean* of a function z of a random variable x is an estimate of the population mean $\langle z \rangle$ obtained from a finite number of N samples or “histories” and is defined as

$$\bar{z} = \frac{1}{N} \sum_{i=1}^N z(x_i). \quad (2.19)$$

From Eq. (2.18) it is seen that the population variance $\sigma^2(z)$ depends on $\langle z \rangle$, which often is not known. Thus, one must estimate the population variance with what is called the *sample variance*. This is done as follows. Equation (2.17) is in the form of Eq. (2.16), with the function z replaced by the function $[z - \langle z \rangle]^2$. Thus, to estimate the variance, one naturally considers the quantity

$$\sigma^2(z) \cong \frac{1}{N} \sum_{i=1}^N [z(x_i) - \langle z \rangle]^2. \quad (2.20)$$

However, $\langle z \rangle$ is unknown—that is, after all, what is being sought. What is usually done is to approximate $\langle z \rangle$ by $\bar{z}$. This estimate of $\sigma^2(z)$ with $\langle z \rangle$ approximated by $\bar{z}$ is called the *sample variance* and is denoted by $s^2(z)$. However, in replacing $\langle z \rangle$ by $\bar{z}$, one *degree of freedom* has been used and, for reasons discussed below, the N in Eq. (2.20) is, by convention, replaced by $N - 1$. Thus, the *sample variance* is defined as

$$s^2(z) \equiv \frac{1}{N-1} \sum_{i=1}^N [z(x_i) - \bar{z}]^2. \quad (2.21)$$

The number of degrees of freedom ν is an often used but not always well-understood parameter. In applications such as that being considered, where there are N *independent* samples, the number of degrees of freedom is simply the number of independent variables less the number of constraints c , i.e., (see, for example [Dunn \[2005\]](#))

$$\nu = N - c. \quad (2.22)$$

The original N samples were unconstrained (they were drawn independently) and, thus, in calculating the sample mean $c = 0$ and $\nu = N$, N is used in the denominator of Eq. (2.19). However, in forming the sample variance, one of the values of $[z(x_i) - \bar{z}]^2$ is constrained ($c = 1$) by using $\langle z \rangle$ to estimate $\bar{z}$. The differences in $N - 1$ cases are independent of each other; they are random samples from which a constant has been subtracted. However, the remaining difference must assume a constrained value, that value being the value that assures that the samples produce $\bar{z}$. In this sense, the sample mean used one degree of freedom and so $c = 1$ and $\nu = N - 1$.

Special cases

Let us look at the simple cases of $N = 1$ and $N = 2$. If one were to draw a single sample z_1 from an unknown distribution then this sample would constitute the sample mean, i.e., $\bar{z} = z_1$. However, one would not have *any* information about the variance. A single sample is insufficient to indicate whether the variance is small or large. Application of Eq. (2.21) leads to the correct conclusion that $s^2(z)$ is unknown. However, putting N in the denominator would lead to the absurd conclusion that $s^2(z) = 0$ and that the population mean is known exactly. Now, draw a second sample z_2 from the same distribution. The sample average is $\bar{z} = (z_1 + z_2)/2$. However, both samples are equally removed from the sample average (i.e., $\bar{z} - z_1 = \Delta = z_2 - \bar{z}$), and hence there is only one *difference* from the mean that is available to be used in estimating the variance. Thus, $\nu = N - 1 = 1$, and the sample variance reduces to $s^2(z) = 2\Delta^2$.

The general case

Thus, use of $1/N$ in estimating the sample mean and of $1/(N - 1)$ in the estimation of the sample variance follows from consideration of degrees of freedom. As a practical matter, this point is largely irrelevant in most Monte Carlo calculations because N

is usually large and the difference between N and $N - 1$ is negligible. Nevertheless, Eq. (2.21) tends to avoid underestimating the sample variance and is the conventional way to express the sample variance in Monte Carlo calculations.

Equation (2.21) can be converted to a form that is easier to calculate. Note that

$$\begin{aligned} s^2(z) &= \frac{1}{N-1} \sum_{i=1}^N \left[z^2(x_i) - 2\bar{z}z(x_i) + \bar{z}^2 \right] \\ &= \frac{N}{N-1} \left[\frac{1}{N} \sum_{i=1}^N z^2(x_i) - \frac{2\bar{z}}{N} \sum_{i=1}^N z(x_i) + \bar{z}^2 \right] \end{aligned}$$

or

$$s^2(z) = \frac{N}{N-1} (\bar{z}^2 - \bar{z}^2).$$

(2.23)

The utility of this result is that in practice it is easy to estimate the sample variance by simply collecting two summations, $\sum_{i=1}^N z(x_i)$ and $\sum_{i=1}^N z^2(x_i)$, as is demonstrated more fully in Chapter 4. Finally, the *sample standard deviation* is simply $s(z) = \sqrt{s^2(z)}$.

2.2 Discrete Random Variables

Some probability density functions are defined only for discrete values of a random variable x . For example, in rolling a die the number of dots on the upward face has only six possible outcomes, namely, $x_1 = 1, x_2 = 2, \dots, x_6 = 6$. In such cases, $f_i \equiv f(x_i)$ represents the probability of event i , where $i = 1, 2, \dots, I$. In general, the number of outcomes I can be finite or unbounded. Then for $f_1, f_2, \dots, f_I$ to be a PDF it is required that (1) $f_i \geq 0$ and (2) $\sum_{i=1}^I f_i = 1$. Further, the discrete CDF

$$F_i \equiv \sum_{j=1}^i f_j \tag{2.24}$$

can be interpreted as the probability that one of the first i events occurs. This discrete CDF has the properties that (1) $F_1 = f_1$, (2) $F_I = 1$, and (3) $F_i \geq F_j$, if $i \geq j$.

In the discrete case, the population mean and variance take the forms

$$\langle x \rangle \equiv \mu(x) \equiv \sum_{i=1}^I x_i f_i \tag{2.25}$$

and

$$\sigma^2(x) \equiv \sum_{i=1}^I [x_i - \langle x \rangle]^2 f_i. \tag{2.26}$$

The discrete probability density function also can be written as a continuous PDF $f(x)$ with $x \in (-\infty, \infty)$ if the Dirac delta function $\delta(x - x_o)$ is used. This function vanishes at all real x values except at $x = x_o$, where it is infinite. Moreover, this function, with zero width and support at only one point x_o , has unit area. One can represent this function as a limiting case of the function $g_\epsilon(x)$, illustrated in Fig. 2.4 and defined as

$$g_\epsilon(x) = \begin{cases} \frac{1}{\epsilon}, & x_o - \frac{\epsilon}{2} \leq x \leq x_o + \frac{\epsilon}{2} \\ 0, & \text{otherwise.} \end{cases} \quad (2.27)$$

The Dirac delta function is then defined as

$$\delta(x - x_o) = \lim_{\epsilon \rightarrow 0} g_\epsilon(x). \quad (2.28)$$

The Dirac delta function has the following properties:

$$\delta(x) = \delta(-x), \quad (2.29)$$

$$\delta(ax) = \frac{1}{|a|} \delta(x), \quad (2.30)$$

and

$$x\delta(x) = 0. \quad (2.31)$$

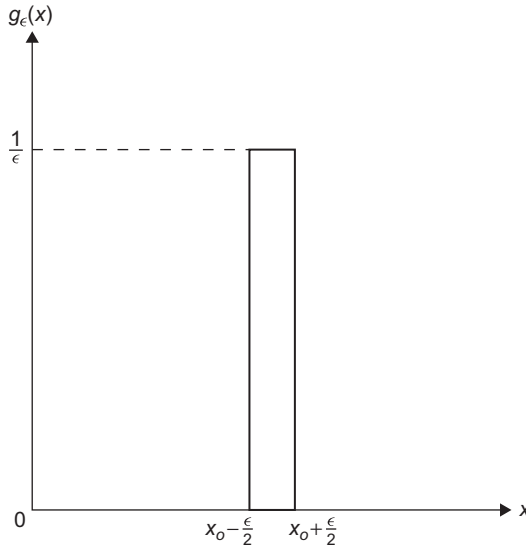


Figure 2.4 Function $g_\epsilon(x)$, which becomes a Dirac delta function as $\epsilon \rightarrow 0$.

These properties really have meaning only when the delta function is inside an integral. For example, the most widely used relation involving the delta function is

$$\int_a^b dx g(x) \delta(x - x_o) = \begin{cases} g(x_o), & \text{if } a \leq x_o \leq b \\ 0, & \text{otherwise,} \end{cases} \quad (2.32)$$

where $g(x)$ is any continuous function.

With this delta function, the discrete PDF can be written as a continuous PDF using the “Dirac comb”:

$$f(x) = \sum_{i=1}^I f_i \delta(x - x_i), \quad -\infty < x < \infty. \quad (2.33)$$

Substitution of this representation into Eqs. (2.12) and (2.13) immediately yields Eqs. (2.25) and (2.26). The CDF can also be written as a continuous function

$$F(x) = \int_{-\infty}^x f(x) dx = \sum_{x_i \leq x} f_i. \quad (2.34)$$

2.3 Multiple Random Variables

In most Monte Carlo applications multiple random variables are involved that often depend on each other. In this section some of the important properties of such probability functions are summarized.

2.3.1 Two Random Variables

The concept of a probability density function of a single random variable can be extended to probability density functions of more than one random variable. For two random variables, x and y , $f(x, y)$ is called the *joint probability density function* if it is defined and non-negative on the interval $x \in [a, b]$, $y \in [c, d]$ and if

$$\int_a^b \int_c^d f(x, y) dx dy = 1. \quad (2.35)$$

The functions

$$f_x(x) = \int_c^d f(x, y) dy \quad \text{and} \quad f_y(y) = \int_a^b f(x, y) dx \quad (2.36)$$

are called the *marginal* PDFs of x and y , respectively. The joint PDF can be written in terms of these marginal PDFs as

$$f(x, y) = f(x|y)f_y(y) = f(y|x)f_x(x), \quad (2.37)$$

where $f(x|y)$ is called the *conditional* PDF of x given y and $f(y|x)$ is called the conditional PDF of y given x . Equation (2.37) can also be used to define the conditional PDFs, namely,

$$f(x|y) = \frac{f(x, y)}{f_y(y)} \quad \text{and} \quad f(y|x) = \frac{f(x, y)}{f_x(x)}. \quad (2.38)$$

The mean and variance of x are given by

$$\langle x \rangle \equiv \int_a^b \int_c^d x f(x, y) dx dy = \int_a^b x f_x(x) dx \quad (2.39)$$

and

$$\sigma^2(x) \equiv \int_a^b \int_c^d (x - \langle x \rangle)^2 f(x, y) dx dy = \int_a^b (x - \langle x \rangle)^2 f_x(x) dx. \quad (2.40)$$

Similar expressions hold for $\langle y \rangle$ and $\sigma^2(y)$. A measure of how x and y are dependent on each other is given by the *covariance*, defined as

$$\text{covar} \equiv \int_a^b \int_c^d (x - \langle x \rangle)(y - \langle y \rangle) f(x, y) dx dy. \quad (2.41)$$

Finally, if $f(x, y) = g(x)h(y)$, where $g(x)$ and $h(y)$ are also PDFs, then x and y are *independent random variables*, and it is easily shown that $\text{covar}(x, y) = 0$.

Example 2.2 Isotropic Scattering

In simulating particle transport, one must keep track of a particle's direction. For instance, as neutrons migrate through matter, they scatter from nuclei. The probability of scattering from one direction into another direction is often expressed as a joint PDF of two angles, a polar angle θ (measured from the initial direction) and an azimuthal angle ψ . Generally, neutron scattering is *rotationally invariant*, meaning that the angle ψ is independent of the scatter angle θ and all azimuthal angles are equally likely. Thus, θ and ψ are independent and the joint PDF for neutron scattering can be expressed as

$$f(\theta, \psi) = g(\theta)h(\psi),$$

where

$$h(\psi) = \frac{1}{2\pi}, \quad \psi \in [0, 2\pi].$$

The form of the PDF $g(\theta)$ depends on such things as the type and energy of the radiation and the frame of reference in which the scatter angle is defined. These matters are more fully developed later, in Chapters 9 and 10, and in references such as [Shultis and Faw \[2000\]](#).

Bayes theorem

Bayes theorem, which follows from the axioms of probability, relates the conditional probabilities of two events, say x and y , with the joint probability density function $f(x, y)$ just discussed. For two random variables, this theorem states

$$f(x|y) = \frac{f_x(x)f(y|x)}{f_y(y)}. \quad (2.42)$$

This result is easily verified by using the definitions of Eq. (2.38). This theorem and its application to nonclassical statistical analysis is discussed at greater length in Chapter 6.

2.3.2 More Than Two Random Variables

The concept of a PDF for two random variables can be generalized to a collection of n random variables $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$. Note that each component random variable x_j , $j = 1, \dots, n$ can be either discrete or continuous. The function $f(\mathbf{x})$ is the joint PDF of $\mathbf{x}$ if it obeys both $f(\mathbf{x}) \geq 0$ for all values of $\mathbf{x} \in V$ and $\int_V f(\mathbf{x}) d\mathbf{x} = 1$, where V defines the “volume” over which $\mathbf{x}$ is defined. If $\mathbf{x}$ is decomposed such that $\mathbf{x} = \{\mathbf{x}_j, \mathbf{x}_k\}$, where $\mathbf{x}_k \in V_k$ and where $j + k = n$,² then

$$f_{\mathbf{x}_j}(\mathbf{x}_j) = \int_{V_k} f(\mathbf{x}) d\mathbf{x}_k, \quad (2.43)$$

is called the marginal PDF of $\mathbf{x}_j$ and

$$f(\mathbf{x}_k|\mathbf{x}_j) = \frac{f(\mathbf{x})}{f_{\mathbf{x}_j}(\mathbf{x}_j)} \quad (2.44)$$

is called the conditional PDF of $\mathbf{x}_k$, given $\mathbf{x}_j$. Also, in analogy to Eq. (2.40), each x_i has a variance and, in analogy to Eq. (2.41), each pair of random variables has a covariance. Finally, if

$$f(\mathbf{x}) = f_{x_1}(x_1)f_{x_2}(x_2) \dots f_{x_n}(x_n) = \prod_{i=1}^n f_{x_i}(x_i), \quad (2.45)$$

then all n random variables are independent and all covariances between any two pairs are zero.

Again, suppose z represents a stochastic process that is a function of the random variable $\mathbf{x}$, where $\mathbf{x}$ is governed by the joint PDF $f(\mathbf{x})$. Then $z(\mathbf{x})$ is also a random

² It is understood that j refers to the first j elements of $\mathbf{x}$ and k refers to the remaining components, i.e., those from $j + 1$ to n .

variable and one can define its expected value, by analogy with Eq. (2.16), as

$$\langle z \rangle \equiv \int_V z(\mathbf{x}) f(\mathbf{x}) d\mathbf{x}, \quad (2.46)$$

and its variance, by analogy with Eq. (2.17), as

$$\sigma^2(z) \equiv \langle [z(\mathbf{x}) - \langle z \rangle]^2 \rangle \equiv \int_V [z(\mathbf{x}) - \langle z \rangle]^2 f(\mathbf{x}) d\mathbf{x}. \quad (2.47)$$

This last result can be reduced to

$$\sigma^2(z) = \langle z^2 \rangle - \langle z \rangle^2. \quad (2.48)$$

The quantity $\langle z \rangle$ is properly called the *population mean*, while $\sigma^2(z)$ is the *population variance* and $\sigma(z) = \sqrt{\sigma^2(z)}$ is the *population standard deviation*. However, the formality of this nomenclature is often ignored and the terms *mean*, *variance*, and *standard deviation* are commonly used, respectively, for these quantities.

2.3.3 Sums of Random Variables

The purpose of a Monte Carlo calculation is to estimate some expected value $\langle z \rangle$ by the population mean or average given by Eq. (2.19). But to know how good this estimate is, the variance of $\bar{z}$, which is also a random variable, is also needed.

Begin by considering the random variable $\xi = \sum_{i=1}^N z_i$ where, in general, the z_i are distributed by a joint PDF $f(z_1, z_2, \dots, z_N)$. The expected value of a sum can be shown (see problem 6) to be the sum of expected values, i.e.,

$$\langle \xi \rangle = \left\langle \sum_{i=1}^N z_i \right\rangle = \sum_{i=1}^N \langle z_i \rangle = \sum_{i=1}^N \mu_i. \quad (2.49)$$

The variance of ξ is calculated as

$$\begin{aligned} \sigma^2(\xi) &= \sigma^2\left(\sum_{i=1}^N z_i\right) = \left\langle \left(\sum_{i=1}^N z_i - \left\langle \sum_{i=1}^N z_i \right\rangle\right)^2 \right\rangle \\ &= \left\langle \left(\sum_{i=1}^N (z_i - \langle z_i \rangle)\right)^2 \right\rangle = \left\langle \sum_{i=1}^N \sum_{j=1}^N (z_i - \langle z_i \rangle)(z_j - \langle z_j \rangle) \right\rangle \\ &= \sum_{i=1}^N \langle (z_i - \langle z_i \rangle)^2 \rangle + \sum_{i=1}^N \sum_{\substack{j=1 \\ j \neq i}}^N \langle (z_i - \langle z_i \rangle)(z_j - \langle z_j \rangle) \rangle \\ &= \sum_{i=1}^N \sigma^2(z_i) + 2 \sum_{i=1}^N \sum_{\substack{j=1 \\ j < i}}^N \text{covar}(z_i, z_j). \end{aligned} \quad (2.50)$$

Now if the z_i are all independent, then $f(z_1, z_2, \dots, z_N) = \prod_{i=1}^N f_i(z_i)$, and it is easy to show that $\text{covar}(z_i, z_j) = 0$ (see problem 7) so that

$$\sigma^2(\xi) = \sum_{i=1}^N \sigma^2(z_i). \quad (2.51)$$

One final property of the variance of a random number is needed, namely, for $\alpha =$ a constant,

$$\sigma^2(\alpha\xi) = \langle (\alpha\xi - \langle \alpha\xi \rangle)^2 \rangle = \alpha^2 \langle (\xi - \langle \xi \rangle)^2 \rangle = \alpha^2 \sigma^2(\xi). \quad (2.52)$$

Now apply these results to the Monte Carlo estimate of the sample mean $\bar{z} = (1/N) \sum_{i=1}^N z(x_i)$. Here, the $z(x_i) \equiv z_i$ are all identically distributed because the z_i are independently sampled from the same PDF $f(z)$. Thus, all the z_i have the same variance $\sigma^2(z)$ given by Eq. (2.18). Then the variance of the sample mean is, from Eqs. (2.51) and (2.52),

$$\begin{aligned} \sigma^2(\bar{z}) &= \sigma^2\left(\frac{1}{N}(z_1 + z_2 + \dots + z_N)\right) \\ &= \frac{1}{N^2} [\sigma^2(z) + \sigma^2(z) + \dots + \sigma^2(z)] = \frac{1}{N} \sigma^2(z). \end{aligned} \quad (2.53)$$

Although $\sigma^2(z)$ is not known, it can be approximated by the sample variance $s^2(z)$, as given by Eq. (2.23), so that

$$\sigma(\bar{z}) \simeq \frac{s(z)}{\sqrt{N}} = \frac{1}{\sqrt{N}} \sqrt{\frac{N}{N-1} (\bar{z}^2 - \bar{z}^2)} \simeq \frac{\sqrt{\bar{z}^2 - \bar{z}^2}}{\sqrt{N}} \text{ for large } N. \quad (2.54)$$

Hence, the Monte Carlo estimation of $\langle z \rangle \pm \sigma(z)$ is calculated as

$$\boxed{\langle z \rangle \pm \sigma(z) \equiv \int_V z(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \pm \sqrt{\int_V (z - \langle z \rangle)^2 f(\mathbf{x}) d\mathbf{x}} \simeq \bar{z} \pm \sqrt{\frac{\bar{z}^2 - \bar{z}^2}{N}}.} \quad (2.55)$$

Here, $\bar{z} = (1/N) \sum_{i=1}^N z(\mathbf{x}_i)$ and $\bar{z}^2 = (1/N) \sum_{i=1}^N z(\mathbf{x}_i)^2$, where the $\mathbf{x}_i$ are distributed according to $f(\mathbf{x})$ over the “volume” V .

2.4 The Law of Large Numbers

Monte Carlo is based on two fundamental statistical results: the *law of large numbers*³ and the *central limit theorem*. Many Monte Carlo books (e.g., [Carter and Cashwell](#)

³ The law of large numbers has strong and weak forms; this is discussed in Appendix B for those interested in the subtle details.

[1975], Ross [1970], Fishman [2003], and Landau and Binder [2002]) discuss the law of large numbers and the central limit theorem because of their importance to Monte Carlo.

The heart of a Monte Carlo analysis is to obtain an estimate of an expected value such as

$$\langle z \rangle = \int_a^b z(x) f(x) dx. \quad (2.56)$$

If one forms the estimate

$$\bar{z} = \frac{1}{N} \sum_{i=1}^N z(x_i), \quad (2.57)$$

where the x_i are suitably sampled from $f(x)$, the law of large numbers states that, as long as the mean exists and the variance is bounded,

$$\lim_{N \rightarrow \infty} \bar{z} = \langle z \rangle. \quad (2.58)$$

This law states that eventually the normalized summation of Eq. (2.57) approaches the expected value of Eq. (2.56). Here, the quadrature nodes x_i are “sampled” from the PDF $f(x)$ and the quadrature weights are equal to $1/(Nf(x_i))$. Details are given in Chapter 4 about how to obtain the samples $\{x_i, i = 1, 2, \dots, N\}$. A proof of the weak version of this law is given in Appendix B. The law of large numbers can be written in the alternative notation

$$\frac{1}{N} (z_1 + z_2 + \dots + z_N) \xrightarrow{N \rightarrow \infty} \langle z \rangle. \quad (2.59)$$

Note the conditions placed on Eq. (2.58) indicate that, in some pathological cases, Eq. (2.58) is not valid. For instance, for certain “fat-tailed” or “heavy-tailed” distributions no population mean can be defined. Brian Hayes [Hayes 2007] presented a discussion of the so-called “factoidal” function. The familiar factorial function of an integer n can be expressed as

$$n! = n(n-1)(n-2) \dots (2)(1), \quad (2.60)$$

the (1) being unnecessary, because it does not change the result, but is conventionally displayed. The factoidal function, identified as $n?$, can be expressed as

$$n? = (r_n)(r_n) \dots (1), \quad (2.61)$$

where r_n is a random integer sampled between 1 and n . Thus, the factoidal function begins by sampling an integer between 1 and n and recording this as $n_1 = r_n$.

If $n_1 = 1$, then sampling stops and $n? = 1$. Otherwise, a second integer is randomly sampled between 1 and n and recorded as n_2 . If $n_2 = 1$, then $n? = n_1$ and sampling stops; otherwise, the procedure is repeated. Eventually, on the k th sampling, a 1 is randomly sampled ($n_k = 1$) and $n? = n_1, n_2, \dots, n_{k-1}$, where $n_k = 1$. This function has no well-defined population mean. Hayes calculated, for $n = 10$, the sample mean after 1000 samples to be about 10^{50} , whereas after 100,000 samples he estimated it to be about 10^{80} ! Clearly, there is no convergence to a well-defined population mean. Hence, the law of large numbers does not apply, as the sample mean is unable to converge to a population mean that has the audacity not to exist.

The law of large numbers provides a prescription for determining the nodes and weights of a Monte Carlo quadrature scheme for estimating *well-defined* definite integrals, but it does not tell us how large N must be in practice. Do 10, 10^4 , 10^{20} , or $10^{10^{10}}$ samples suffice? The answer to this question is provided by the central limit theorem, discussed next.

2.5 The Central Limit Theorem

One of the important features of Monte Carlo is that not only can one obtain an estimate of an expected value (by the law of large numbers) but also one can obtain an estimate of the uncertainty in the estimate (by the central limit theorem).⁴ Thus, at the end of a Monte Carlo simulation, one can have an idea not only of what the answer is but also of how good the estimate of the answer is.

The central limit theorem (CLT) is a very general and powerful theorem. In one form [Kendall and Stuart 1977] it states that, for $\bar{z}$ obtained by samples from a distribution with mean $\langle z \rangle$ and standard deviation $\sigma(z)$,

$$\lim_{N \rightarrow \infty} \text{Prob} \left\{ \alpha \leq \frac{\bar{z} - \langle z \rangle}{\sigma(z)/\sqrt{N}} \leq \beta \right\} = \frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\beta} e^{-u^2/2} du. \quad (2.62)$$

In another form, the CLT can be written as

$$\lim_{N \rightarrow \infty} \text{Prob} \left\{ \frac{|\bar{z} - \langle z \rangle|}{\sigma(z)/\sqrt{N}} \leq \lambda \right\} = \frac{1}{\sqrt{2\pi}} \int_{-\lambda}^{\lambda} e^{-u^2/2} du. \quad (2.63)$$

Stated in words, the CLT provides the following insights:

1. The CLT tells us that the asymptotic distribution of $(\bar{z} - \langle z \rangle)/[\sigma(z)/\sqrt{N}]$ is a unit normal distribution or, equivalently, $\bar{z}$ is asymptotically distributed as a normal distribution with mean $\mu = \langle z \rangle$ and standard deviation $\sigma(z)/\sqrt{N}$.

⁴ Other advantageous features of Monte Carlo include that it can be applied to highly multidimensional and complex problems and that the user has a high degree of flexibility in how the sampling is done and the results are computed.

2. The astonishing feature about the CLT is that nothing is said about the distribution function used to generate the N samples of z , from which the random variable $\bar{z}$ is formed. No matter what the distribution is, provided it has a finite variance, the sample mean $\bar{z}$ has an approximately normal distribution for large samples. The restriction to distributions with finite variance is of little practical consequence because, in almost all practical situations, the variance is finite.
3. As $\lambda \rightarrow 0$, the right side of Eq. (2.63) approaches zero. Thus, the sample mean $\bar{z}$ approaches the true mean $\langle z \rangle$ as $N \rightarrow \infty$, a result that corroborates the law of large numbers.
4. Finally, the CLT provides a practical way to estimate the uncertainty in a Monte Carlo estimate of $\langle z \rangle$, because the sample standard deviation $s(z)$ of Eq. (2.23) can be used to estimate the population standard deviation $\sigma(z)$ in Eq. (2.63).

The CLT gives teeth to the Monte Carlo method because it provides an estimate of the uncertainty in the estimated expected value. Most important, it states that the uncertainty in the estimated expected value is proportional to $1/\sqrt{N}$, where N is the number of histories or samples of $f(x)$. If the number of histories is quadrupled, the uncertainty in the estimate of the sample mean is halved. A proof of the all-important CLT is given in Appendix C.

The uncertainty in the Monte Carlo estimation of the true mean can be obtained from Eq. (2.63) by approximating the unknown population standard deviation $\sigma(z)$ by the calculated sample standard deviation $s(z)$ given by Eq. (2.23). Thus, for large N , Eq. (2.63) can be written as

$$\text{Prob}\{\bar{z} - \lambda s(z)/\sqrt{N} \leq \langle z \rangle \leq \bar{z} + \lambda s(z)/\sqrt{N}\} \simeq \frac{1}{\sqrt{2\pi}} \int_{-\lambda}^{\lambda} e^{-u^2/2} du. \quad (2.64)$$

The right side of this result can be evaluated numerically and is called the *confidence coefficient*. The confidence coefficients, expressed as percentages and rounded or truncated, are called *confidence limits*. Results for various values of λ are shown in Table 2.1. Here, λ is the number of standard deviations, from the mean, over which the unit normal is integrated to obtain the confidence coefficient.

Thus, for a given confidence level, the Monte Carlo estimate of $\langle z \rangle$ is usually reported as $\bar{z} \pm \lambda s(z)/\sqrt{N}$. For example, the interval (for $\lambda = 1$) $[\bar{z} - s(z)/\sqrt{N}, \bar{z} + s(z)/\sqrt{N}]$ has a 68% chance of containing the true mean $\langle z \rangle$. At this point,

Table 2.1 Confidence limits for various values of λ .

λ	Confidence Coefficient	Nominal Confidence Limit
0.25	0.1974	20%
0.50	0.3829	38%
1.00	0.6827	68%
1.50	0.8664	87%
2.00	0.9545	95%
3.00	0.9973	99%
4.00	0.9999	99.99%

there is a somewhat sloppy use of language. Both the random interval $[\bar{z} - s(z)/\sqrt{N}, \bar{z} + s(z)/\sqrt{N}]$ and its specific numerical value for given values of $\bar{z}$ and $s(z)$ (e.g., [3.5, 5.2]) are both called *confidence intervals*. In the latter case, what is meant is that the interval [3.5, 5.2] has a 68% chance of containing the true value of the mean. By contrast, the interpretation that $[\bar{z} - s(z)/\sqrt{N}, \bar{z} + s(z)/\sqrt{N}]$ is a random variable means that, if samples of size N are repeatedly constructed and values of the uncertainty interval are calculated for each, then the relative frequency of those intervals that contain the true unknown mean would approach 68%.

In summary, the CLT guarantees that the deviation of the sample mean from the true mean approaches zero as $N \rightarrow \infty$. The quantity $\sigma/\sqrt{N}$ provides a measure of the deviation of the sample mean from the population mean after N samples. Use of the sample standard deviation $s(z)$, to approximate the population standard deviation $\sigma(z)$, allows the construction of a confidence interval about $\bar{z}$ that has a specified probability of containing the true unknown mean. As the sample size N increases, this confidence interval, the width of which is proportional to $1/\sqrt{N}$, becomes progressively smaller.

2.6 Monte Carlo Quadrature

The results of Eq. (1.17) for a one-dimensional integral can be extended to a multi-dimensional integral over a volume V of a function $g(\mathbf{x})$. An analysis, similar to that used for Eq. (2.55), yields

$$\langle g \rangle \pm \sigma(g) \equiv \int_V g(\mathbf{x}) d\mathbf{x} \pm \sqrt{\int_V (g - \langle g \rangle)^2 d\mathbf{x}} \simeq V \bar{g} \pm V \sqrt{\frac{\overline{g^2} - \bar{g}^2}{N}}, \quad (2.65)$$

where $\bar{g} = (1/N) \sum_{i=1}^N g(\mathbf{x}_i)$ and $(1/N) \overline{g^2} = \sum_{i=1}^N g(\mathbf{x}_i)^2$ with the $\mathbf{x}_i$ being uniformly distributed over V .

For V of one or two dimensions, deterministic numerical quadratures such as trapezoid, Simpson, or Gauss-Legendre quadrature are usually as efficient as Monte Carlo quadrature, which requires more evaluations of $g(\mathbf{x})$ to achieve the same accuracy. However, in many practical problems, the dimension of V may be very large. For example, in computing the distributions of stars in a galaxy or the energy of electrons in condensed matter, dimensions of 10^5 – 10^{30} are not uncommon. For these cases, Monte Carlo quadrature is the only practical way to evaluate the integrals. Even for dimensions of a few, Monte Carlo quadrature may be preferred, especially if evaluation of the integrand is computationally complex. [Example 2.3](#) illustrates the effectiveness of Monte Carlo quadrature for multidimensional integrals.

Example 2.3 A Multidimensional Integral

Consider the problem of calculating a five-fold integral

$$I = \int_0^1 dx_1 \int_0^1 dx_2 \int_0^1 dx_3 \int_0^1 dx_4 \int_0^1 dx_5 g(x_1, x_2, x_3, x_4, x_5). \quad (2.66)$$

One approach is to approximate each integral by a trapezoid at n equispaced ordinates between $x_1 = 0$ and $x_n = 1$, so that for a single integral,

$$\int_0^1 f(x) dx \simeq \delta \left\{ \frac{1}{2} [f(x_1) + f(x_n)] + \sum_{i=2}^{n-1} f(x_i) \right\},$$

where $\delta = 1/(n - 1)$ and $x_i = (i - 1)\delta$. For the five-dimensional integral of Eq. (2.66), the trapezoid approximation leads to the five-fold nested sum

$$I_{trap} = \delta^5 \sum_{i=1}^n w_i \sum_{j=1}^n w_j \sum_{k=1}^n w_k \sum_{l=1}^n w_l \sum_{m=1}^n w_m g(x_i, x_j, x_k, x_l, x_m),$$

where the weight $w = 1/2$ for the endpoints and equals 1 otherwise. A typical Monte Carlo approach would uniformly sample points in the five-dimensional unit hypercube, so that each coordinate is uniformly distributed between 0 and 1.

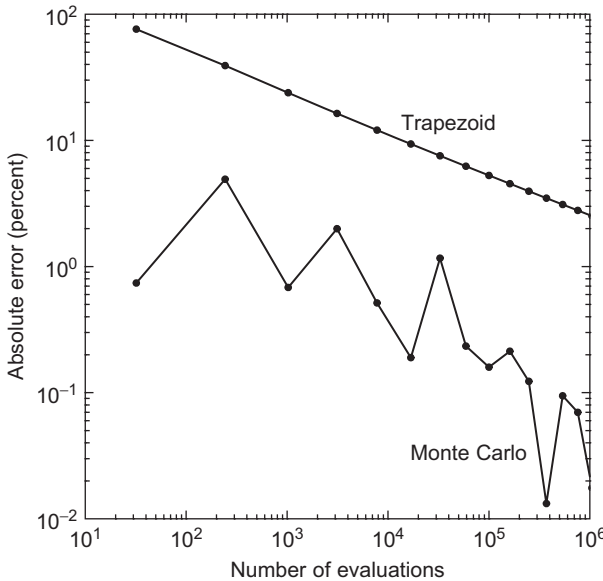


Figure 2.5 Trapezoid and Monte Carlo estimates of the five-dimensional integral of Eq. (2.66).

Consider Eq. (2.66) in which $g(\mathbf{x}) = \sqrt{x_1 x_2 x_3 x_4 x_5}$. This simple example can be easily integrated analytically to find $I_{\text{exact}} = (2/3)^5 \simeq 0.13169$. In Fig. 2.5, the absolute error in estimating this integral by both the trapezoid and Monte Carlo approaches is plotted as a function of the number of evaluations of the integrand $g(\mathbf{x}) = \sqrt{x_1 x_2 x_3 x_4 x_5}$. It is seen from these results that the Monte Carlo approach requires far fewer integrand evaluations, compared to the trapezoid approach, to achieve the same accuracy. This can result in considerable computational savings, especially if the integrand evaluation is time consuming. For multidimensional integrals of dimension ≥ 3 , the Monte Carlo approach is often, by far, the most efficient.

2.7 Monte Carlo Simulation

In many complex problems an estimate is sought of some mean value

$$\langle z \rangle = \int_a^b z(x) f(x) dx, \quad (2.67)$$

in which the underlying PDF $f(x)$ is not known a priori. Suppose, for example, that the expected energy deposited in a thyroid, $\langle E \rangle$, arising from gamma rays emitted by some radioactive source outside the body, is sought. The probability a source photon deposits energy in dE about E in the thyroid has a PDF $f(E)$. Then the expected energy deposition per source photon is simply $\langle E \rangle = \int E f(E) dE$. Unfortunately, $f(E)$ is unknown. But, as shown in Example 1.3, the unknown PDF for the Buffon's needle problem can be estimated numerically by performing a simulation of needle drops. Similarly, $f(E)$ can be constructed by simulating the movement of source photons as they migrate throughout the problem geometry and by recording the energy each simulated photon history deposits in the thyroid. With enough simulated photon histories, the PDF $f(E)$ can be approximated and its average value calculated $\bar{E} \simeq \langle E \rangle$. In such simulations one strives to mimic nature as accurately as possible and, hence, the simulation must be based on probability models describing the stochastic manner in which radiation interacts with ambient atoms as the radiation migrates through the problem geometry. Monte Carlo simulations that numerically mimic the actual physical processes involved are called *analog* Monte Carlo simulations.

However, in many problems an analog simulation is computationally impractical. For example, if the radiation source were far distant from the thyroid and there were many objects between the thyroid and source, very few simulated photons would ever reach the modeled person, let alone deposit any energy in the thyroid. Enormous numbers of simulated particle histories would have to be constructed to obtain even a few thyroid scores. Monte Carlo simulations, however, can still be used effectively by introducing biases in the simulation. For example, source photons could be forced to be emitted in directions toward the thyroid target and photon scattering could be biased so that scattered photons move preferentially toward the target. Such simulations no longer mimic nature and are called *nonanalog* Monte Carlo simulations. Of course, if the introduction of biases in the simulation were all that was done in nonanalog simulations, the estimated PDF $f(E)$ would be totally wrong. The key for such biased

simulations is to adjust the tally, or score, to undo the biases introduced into the physical models to obtain a correct score. The biased physics allows more photons to reach the thyroid target and the tally adjustments give correct tally values, thereby requiring far fewer particle histories to obtain statistically meaningful results compared to analog simulations. Some of the many tricks used in nonanalog Monte Carlo are discussed in Chapter 5.

Finally, there are many Monte Carlo simulations used to solve various mathematical problems that are not based on modeling natural processes. For example, in Chapter 6 it is shown that Markov chains can be used in a Monte Carlo method to solve integral equations and other interesting mathematical problems. Likewise, random walks of synthetic particles can be used to solve sets of linear algebraic equations, equations that lie at the heart of numerical solutions of heat conduction, fluid flow, diffusion problems, and many other important equations of science and engineering. We hint at such applications in [Example 2.4](#).

Example 2.4 Solving Linear Algebraic Equations

To demonstrate how Monte Carlo can be used to solve linear algebraic equations, consider the problem of heat conduction in a one-dimensional rod of length L . The temperature at the left end is $T(0)$ and at the right $T(L)$. There are no heat sources in the rod. The temperature profile is given by the heat conduction equation

$$k \frac{d^2 T(x)}{dx^2} = 0 \quad \text{of which the solution is} \quad T(x) = T(0) + \frac{T(L) - T(0)}{L}x.$$

To solve such equations numerically, often a discretization scheme is used to find the temperatures at a finite number of distances $\{x_i\}$, i.e., $T_i \equiv T(x_i)$. For example, assume an equimesh of n nodes with $x_1 = 0$ and $x_n = L$. Integration of the heat conduction equation over the interval $(x_{i-1/2}, x_{i+1/2})$, where $x_{i+1/2} \equiv (x_i + x_{i+1})/2$, and the use of finite differences yields

$$\int_{x_{i-1/2}}^{x_{i+1/2}} \frac{d^2 T(x)}{dx^2} = \left. \frac{dT}{dx} \right|_{x_{i+1/2}} - \left. \frac{dT}{dx} \right|_{x_{i-1/2}} \simeq \frac{T_{i+1} - T_i}{\Delta x} - \frac{T_i - T_{i-1}}{\Delta x} = 0.$$

Thus, the T_i , $i = 2, \dots, (n-1)$ are the solutions of the following $(n-2)$ linear algebraic equations (with $T_1 = T(0)$ and $T_n = T(L)$ known):

$$T_{i-1} - 2T_i + T_{i+1} = 0, \quad i = 2, \dots, (n-1),$$

or rearranging

$$T_i = T_{i-1}/2 + T_{i+1}/2, \quad i = 2, \dots, (n-1). \quad (2.68)$$

The solution of these equations can be estimated by the following random walk “game”:

1. Start a “particle” at x_2 with probability $P_2 = T(0)/(T(0) + T(L))$ or at x_{n-1} with probability $1 - P_2$.

2. At each step in the walk, move the particle to the left or right each with a 50% probability.
3. Each time that a particle lands on node i increase a counter $c_i = c_i + 1$, where c_i records the cumulative number of visits to node i .
4. If the particle reaches x_1 or x_n , kill it and start a new particle at x_2 or x_{n-1} , as in step 1.
5. Perform steps 1–4 for $(T(0) + T(L))/2$ particles. This constitutes one game.

After playing M complete games,

$$T_i = \langle c_i \rangle \simeq \bar{c}_i = \frac{1}{M} \sum_{j=1}^M c_i^j,$$

where c_i^j is the cumulative number of particle visits to node i in game j . Equivalently, one could play the game for a large number N of source particles continuously accumulating the c_i . The number of games represented by N is $N_g = 2N/(T(0) + T(L))$, so that

$$T_i = \langle c_i \rangle \simeq \frac{c_i}{N_g}.$$

Note that this equivalent approach does not require $(T(0) + T(L))/2$, the number of particles per game, to be an integer.

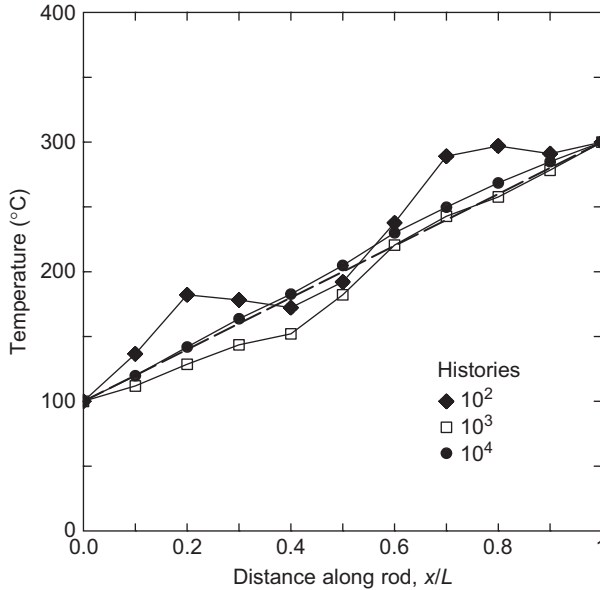


Figure 2.6 Monte Carlo results for the one-dimensional heat conduction problem of [Example 2.3](#). Here, $T(0) = 100^\circ\text{C}$ and $T(L) = 300^\circ\text{C}$. The exact profile is shown by the heavy dashed line. Results based on three different numbers of particle histories are shown.

That $\langle c_i \rangle = T_i$ can be seen from the fact that the expected number of visits to node i equals one-half of the expected number of visits to nodes $i - 1$ and $i + 1$, i.e.,

$$\langle c_i \rangle = \langle c_{i-1} \rangle / 2 + \langle c_{i+1} \rangle / 2, \quad i = 2, \dots, (n - 1).$$

These are the same equations as Eq. (2.68), and hence $\langle c_i \rangle = T_i$. Example results are shown in Fig. 2.6.

2.8 Summary

A probability density function of a single continuous variable is a non-negative function defined on an interval and normalized so that its integral over that interval is unity. The associated cumulative distribution function of x is interpreted as the probability that a random sample has a value less than or equal to x . These concepts extend naturally to discrete and multidimensional random variables.

Generally, probability distributions have well-defined population means and variances. These are explicit values that are characteristic of the distribution. Monte Carlo is a method that allows one to estimate the population mean and population variance by the sample mean and sample variance. These concepts apply to functions of a discrete variable, to functions of a continuous variable, and to functions of many variables, whether discrete or continuous.

The Monte Carlo estimates of the sample mean and sample variance almost surely approach the population mean and population variance as the number of samples (usually called “histories”) gets large. This important feature is a consequence of the law of large numbers, which can be stated as follows. If

$$\langle z \rangle = \int_a^b z(x) f(x) dx \quad (2.69)$$

and

$$\bar{z} = \frac{1}{N} \sum_{i=1}^N z(x_i), \quad (2.70)$$

with the x_i suitably sampled from $f(x)$, then, almost surely,

$$\lim_{N \rightarrow \infty} \bar{z} = \langle z \rangle. \quad (2.71)$$

The CLT then provides a prescription for estimating the uncertainty in the sample mean and sample variance.

It should be kept in mind that the term *standard deviation* is often used, rather loosely, for any of the following:

- the *population standard deviation* $\sigma(z)$;
- the *sample standard deviation* $s(z) = [N(\bar{z}^2 - \bar{z}^2)/(N - 1)]^{1/2}$;

- the *standard deviation of the sample mean* $\sigma(\bar{z}) = \sigma(z)/\sqrt{N}$; and
- the *estimation of the standard deviation of the sample mean* $s(\bar{z}) = s(z)/\sqrt{N} = [(\bar{z}^2 - \bar{z}^2)/(N-1)]^{1/2}$.

The context must be employed to indicate which interpretation is meant.

Finally, Monte Carlo can mean a wide variety of different calculational methods, but all are based on the statistical power of the central limit theorem that guarantees expected values can be approximated by averages and that allows confidence intervals to be constructed for the averages. Although many different applications of Monte Carlo appear to be quite different, they all depend on this critical theorem.

Bibliography

- CARTER, L.L. AND E.D. CASHWELL 1975, *Particle-Transport Simulation with the Monte Carlo Method*, TID-26607, National Technical Information Service, U.S. Department of Commerce, Springfield, VA.
- DUNN, P.F. 2005, *Measurement and Data Analysis for Engineering and Science*, McGraw-Hill, New York.
- FISHMAN, G.S. 2003, *Monte Carlo Concepts, Algorithms, and Applications*, Springer-Verlag, New York (corrected printing).
- HAYES, B. 2007, "Fat Tails," *Sci. Am.*, **95**, 200–204.
- KAHN, H. 1956, *Applications of Monte Carlo*, AECU-3259, United States Atomic Energy Commission.
- KENDALL, M. and A. SUART 1977, *The Advanced Theory of Statistics, Volume 1 Distribution Theory*, 4th ed., MacMillan Publishing Co., New York.
- LANDAU, D.P. AND K. BINDER 2002, *A Guide to Monte Carlo Simulations in Statistical Physics*, Cambridge University Press, Cambridge (reprinted).
- ROSS, S.M. 1970, *Applied Probability Models with Optimization Applications*, Holden-Day, San Francisco.
- SHULTIS, J.K. and R.E. FAW 2000, *Radiation Shielding*, American Nuclear Society, La Grange Park, IL.

Problems

1. What is the probability density function for a plane angle ϕ , expressed in radians, distributed uniformly between 0 and 2π ?
2. The discrete Poisson distribution has PDF

$$p(x) = \frac{\lambda^x}{x!} \exp(-\lambda),$$

where λ is a positive constant. (a) What is the CDF for this distribution? (b) What are the mean and variance of this distribution?

3. Radiation particles escaping into a vacuum from a plane surface are often observed to do so with an intensity proportional to the n th power of the cosine ω of the angle between the outward normal and the particle travel direction, i.e., proportional to ω^n , $0 \leq \omega \leq 1$. For $n = 1$, this is known as *Lambert's Law*. (a) What is the PDF $f(\omega)$ that describes this

phenomena, namely, that $f(\omega)d\omega$ is the probability that an escaping particle has a direction in $d\omega$ about ω ? (b) What is the associated cumulative distribution? (c) What is the mean and variance of this distribution? (d) What fraction of escaping particles escape at angles greater than 45° ?

4. Each die of a pair of loaded dice has twice the probability of giving a “six” than any other number all of which have equal probability. (a) Write a PDF that describes the probability of obtaining each possible outcome from a throw of this pair of dice. (b) What is the expected value of a throw? (c) What is the variance of the result of a throw? (d) Compare these results to those for a “fair” pair of dice.
5. A radioisotope decays exponentially with time according to

$$N(t) = N_0 e^{-\lambda t},$$

where $N(t)$ is the number of radioactive atoms remaining at time t , N_0 is the number of radioactive atoms at time $t = 0$, and λ is called the decay constant. For a radioisotope for which $\lambda = 0.0693 \text{ min}^{-1}$, what fraction of the atoms are expected to decay between the times $t = 15 \text{ min}$ and $t = 30 \text{ min}$?

6. Verify Eq. (2.49).
7. Show that the covariance terms in Eq. (2.50) vanish if the z_i are independent random variables.
8. Estimate the probability that $|\bar{z}_k - \langle z \rangle_k| \leq \lambda \sigma / \sqrt{N}$, for $\lambda = 0.6745, 1, 2, 3$, and 4 .
9. The weights of 20 football players are (in kg) 92, 94, 100, 86, 93, 91, 89, 86, 105, 96, 87, 90, 97, 87, 95, 93, 103, 88, 98, and 92. (a) Estimate the average weight of all the football players. (b) What is the 95% confidence interval for this estimate? (c) What is the probability that a player has a weight in excess of 120 kg? (d) What is the probability that a player has a negative weight? (e) What does this tell you about the CLT and real-world random variables?
10. Evaluate the integral $\int_0^5 x e^{-\alpha x} dx$ by Monte Carlo quadrature for $\alpha = 1, 2$, and 3 and compare these results to the analytical result

$$\frac{1 - e^{-5\alpha}(5\alpha + 1)}{\alpha^2}$$

for various sample sizes.

11. Some integrands are unbounded and yet their integrals exist. Evaluate by Monte Carlo quadrature the integrals

$$\int_0^1 \frac{dz}{\sqrt{x}} \quad \text{and} \quad \int_0^1 \frac{\sin x}{\sqrt{1-x}} dx.$$

Estimate the accuracy of your results for different sample sizes.

12. Show that the joint PDF of Example 2.2 for isotropic scattering obeys the appropriate normalization condition.

3 Pseudorandom Number Generators

*Monte Carlo needs numbers quite random
but practicality has mostly banned them.
So in order to score
we need do no more
than generate them pseudorandom!*

At the core of all Monte Carlo calculations is some mechanism to produce a long sequence of random numbers ρ_i that are uniformly distributed over the open interval (0,1). However, digital computers, by design, are incapable of producing random results. A true random sequence could, in principle, be obtained by coupling to our computer some external device that would produce a truly random signal. For example, we could use the time interval between two successive clicks of a Geiger counter placed near a radioactive source or use the “white noise” in some electronic circuit to provide truly random numbers.

However, use of such a random number generator would be disastrous! First, the feasibility of having to couple some external device that generates a random signal to every computer we want to use for Monte Carlo calculations would be impractical. But more important would be the impossibility of writing and debugging a Monte Carlo code if, on every run, a different sequence of random numbers were used. What is needed is a sequence of random numbers that is the same every time the program is run so that code errors can be found and so that the same results are produced when the same code is run on different computers.

One way to use true random numbers that would have the same sequence every time a code is run would be to create somehow a table of random numbers and then read them into the computer as they are needed. In fact, Marsaglia [1995] has produced a CD-ROM that contains 4.8×10^9 truly random numbers (generated by combining two pseudorandom number generators and three devices that use random physical processes). However, this is not a good idea. First, in modern Monte Carlo applications many more random numbers are needed, and second, and more importantly, the time to read a number from an external file is prohibitively long for modern Monte Carlo calculations.

An alternative to producing the same sequence of random numbers each time a program is run is to use a *pseudorandom* number generator. Such a generator is a deterministic algorithm that, given the previous numbers (usually just the last number) in the sequence, the next number can be efficiently calculated. Many pseudorandom number generators, some good but most poor, have been proposed and used over the years in a wide variety of Monte Carlo work. Designing better random number generators (the term “pseudo” is sometimes dropped from now on) is still an active

area of research. In this chapter, the basic properties of some important random number generators are discussed and several important random number generators are presented.

The term *quasirandom* numbers is sometimes encountered. These numbers are designed not so much to be random as to be extremely uniform over the interval $(0, 1)$ so as to allow very accurate integration by the Monte Carlo method. Although the generation of quasirandom numbers is not considered further here, the application of such numbers is considered in Section 5.11.2.

Although not reviewed in this chapter, there have been many highly deficient random number generators used in many studies. Some still continue to be used by researchers unaware of the biases being introduced into their calculations. Indeed, if all publications based on these questionable generators were to disappear, much valuable library shelf space would be recovered. When you embark on Monte Carlo investigations, be well aware of the pedigree of the random number generator you are using. Even more important, don't alter your random number generator unless you are very confident your changes are an improvement. The construction of random number generators is best left to the experts.

3.1 Linear Congruential Generators

Perhaps the most widely used random number generators are those based on the *linear congruential generator* proposed by [Lehmer 1951], which has the general form¹

$$x_{i+1} = (ax_i + c) \bmod m, \quad i \geq 0. \quad (3.1)$$

Here, integer $a > 1$ is called the *multiplier*, integer c the *increment*, and integer m the *modulus* of the generator. Starting with some initial integer *seed* x_0 , the sequence of integers x_i will lie in the range $0 \leq x_i < m$. Each x_i is then scaled to the interval $[0, 1)$ by dividing by m , i.e.,

$$\rho_i = \frac{x_i}{m}, \quad (3.2)$$

to obtain random numbers over the interval $[0, 1)$.

In the most frequently programmed congruential generator, the increment c is set to zero and Eq. (3.1) becomes the so-called *multiplicative congruential generator*

$$x_{i+1} = ax_i \bmod m, \quad i \geq 0. \quad (3.3)$$

¹ The notation $\bmod m$ means that the value of $ax_i + c$ is divided by m and the *remainder* resulting from the division is then kept as x_{i+1} . Two integers a and b are said to be *congruent modulo m* if $a \bmod m = b \bmod m$. Hence, the name of this random number generator. The actual generator Lehmer used was $x_{i+1} = 23x_i \bmod (10^8 + 1)$, which, while better than some successors, is not very good by today's standards.

Because of the modular arithmetic the values of the integers x_i cannot exceed $m - 1$ and, for a nonzero initial seed, x_i can never equal 0. Thus, the $\rho_i = x_i/m$ must lie in the open interval $(0, 1)$. For many Monte Carlo simulations it is important that the random number generator never returns a ρ_i that is exactly 1 or 0.

Just how many distinct integer values $1 \leq x_i \leq m - 1$ are obtained depends very critically on the relation between a and m . This is the crux of a good generator: pick a and m such that a vast number of different values x_i are obtained and that they and all pairs, triplets, quadruplets, and so on (i.e., all order k -tuples) “appear” to be distributed randomly.

If $k_i = [ax_{i-1}/m]$,² then Eq. (3.3) can be written in the alternate fashion:

$$\begin{aligned} x_1 &= ax_0 - k_1m \\ x_2 &= a^2x_0 - k_2m - ak_1m \\ x_3 &= a^3x_0 - k_3m - ak_2m - a^2k_1m \\ &\vdots \\ x_j &= a^jx_0 - m(k_j + ak_{j-1} + a^2k_{j-2} + \cdots + a^{j-1}k_1). \end{aligned}$$

Because $1 \leq x_i < m$, $1 \leq i \leq j$, the integer $(k_j + ak_{j-1} + \cdots + a^{j-1}k_1)$ must be the largest integer in a^jx_0/m , i.e., it must equal $[a^jx_0/m]$. Hence, an alternative representation of the generator of Eq. (3.3) is

$$x_j = a^jx_0 \bmod m, \quad j \geq 1. \quad (3.4)$$

The initial seed x_0 must not equal 0 or m because these generate a sequence of zeros. Because there are at most $m - 1$ integers x_i in $0 < x_i < m$, the maximum *period* or *cycle length* of the multiplicative consequential generator is $m - 1$. However, for a given m and a , the period may be much less than $m - 1$. Clearly, for a practical random number generator the modulus m must be very large and the multiplier a must be picked to produce a very long (if not full) period.

3.2 Structure of the Generated Random Numbers

All of the $\{\rho_i\}$ produced by a full-period generator are clearly uniformly distributed over the interval $(0, 1)$. However, a useful generator must produce *subsamples* that must also be distributed uniformly over $(0, 1)$. Moreover, there should be no correlation between successive numbers, i.e., each number should appear to be independent of the preceding number. Although such correlations can be minimized with proper choices for a and m , they are never completely eliminated. For a poor choice of a and m , the correlations become obvious.

² The notation $[x/y]$, where x and y are integers, means that only the integer portion of the quotient x/y is used, i.e., any fractional part is discarded.

For example, consider a full-period generator with $a = 2$ and a very large m . At some point in the cycle x_i will be returned as 1 (or some other small integer). The subsequent values of $x_{i+1}, x_{i+2}, \dots$ will then also be small integers. If $x_i = 1$, the subsequent sequence is 1, 2, 4, 8, 16, $\dots$. Clearly, this subsample is far from random.

Example 3.1 A Simple Multiplicative Congruential Generator

Consider the sequences produced by the simple generator $x_{i+1} = ax_i \bmod 13$ for $(1 \leq a < 13)$. All are started with $x_0 = 1$. The following results are obtained:

a	x_i												
	$i = 0$	1	2	3	4	5	6	7	8	9	10	11	12
1	1	1											
2	1	2	4	8	3	6	12	11	9	5	10	7	1
3	1	3	9	1									
4	1	4	3	12	9	10	1						
5	1	5	12	8	1								
6	1	6	10	8	9	2	12	7	3	5	4	11	1
7	1	7	10	5	9	11	12	6	3	8	4	2	1
8	1	8	12	5	1								
9	1	9	3	1									
10	1	10	9	12	3	4	1						
11	1	11	4	5	3	7	12	2	9	8	10	6	1
12	1	12	1										

Notice $a = 1$ produces only a sequence of ones, and hence a must always be > 1 . Only four multipliers—namely, $a = 2, 6, 7, 11$ —produce full periods. For other multipliers, cycles of lengths 6, 4, 3, and 2 are realized. Generally, the cycle length of the nonfull-period sequences depends on the initial seed. Note also that, except for $a = 12$, the nontrivial sequences occur in pairs, one being the reverse of the other.

As observed by Marsaglia [1968] the sequence of values x_i produced by any linear congruential generator has an extremely rigid structure. If k random numbers at a time (k -tuples) are used as coordinates of points in k -space, the points do not fill the space randomly, but rather they lie on a finite number—and possibly a small number—of parallel hyperplanes. There are at most $m^{1/k}$ such planes. Further, these points form a simple k -dimensional lattice.³

As a simple example, consider the generator

$$x_{i+1} = ax_i \bmod 29 \quad \text{with } x_0 = 1$$

³ A k -dimensional lattice is defined by k independent vectors $\mathbf{v}_j$ such that any point on the hyperplanes (lattice) is given by $\sum_{j=1}^k n_j \mathbf{v}_j$, where n_j are integers.

and plot the points $(x_0, x_1), (x_1, x_2), (x_2, x_3), \dots$. In this two-dimensional plot, the hyperplanes become straight lines. For this generator, $a = 3, 8, 10, 11, 14, 15, 18, 19, 21, 26$, and 27 produce full-period generators. However, the $m - 1 = 28$ points lie on different numbers of lines: two lines for $a = 14, 15$, and 27 ; three lines for $a = 3, 10, 19$, and 26 ; and six lines for $m = 8, 11, 18$, and 21 . Two examples of the resulting two-dimensional plots are shown in Fig. 3.1.

The left plot uses the generator $x_{n+1} = 3x_n \bmod 29$ to produce the sequence

1 3 9 27 23 11 4 12 7 21 5 15 16 19 28 26 20 2 6 18 25 17 22 8 24 14 13 10 1

The numbering of the points refers to the number pairs, i.e., 1 is the first pair (1, 3), 2 is the second pair (3, 9), and so on. This plot is somewhat alarming. All points in the lattice fall on just three lines with slope 3, and there are also ten lines of slope $-2/9$. The distances between adjacent points on these two sets of lines is very different (3.16 versus 9.22), and the full “volume” (area) of the square is far from being uniformly covered by the points, a property we would expect if we use a good random number generator.

The right plot in Fig. 3.1 uses the generator $x_{n+1} = 8x_n \bmod 29$ to produce the sequence

1 8 6 19 7 27 13 17 20 15 4 3 24 18 28 21 23 10 22 2 16 12 9 14 25 26 5 11 1

This is a much more reassuring picture. The points are more evenly distributed over the plot area and lie along six lines of slope $3/4$ or along seven lines of slope $-5/3$. Moreover, the distances between adjacent points of these two sets of lines are comparable (5.00 versus 5.83), indicating a nearly equidistribution of points over the area of the square.

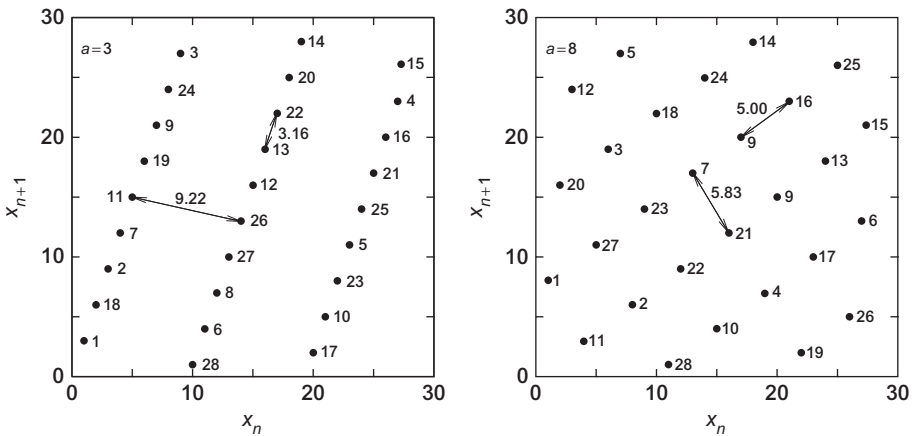


Figure 3.1 Two plots of successive numbers from $x_{n+1} = ax_n \bmod 29$. The left plot has a multiplier $a = 3$ and the right plot $a = 8$. Both are full-period generators, but the right generator fills the sample space much more uniformly than the left generator.

3.3 Characteristics of Good Random Number Generators

There are several properties that any random number generator must possess. First and foremost, the sequence of numbers must be distributed “randomly.” Of course, pseudorandom numbers are not truly random; but they should be able to pass the same statistical tests for randomness that a sequence of truly random numbers would pass. However, the number of possible tests for randomness is uncountably infinite. Thus, the finite number of statistical tests used to evaluate and find the “best” set of parameters for a particular generator should be those that emphasize known weaknesses of the generator. For example, tests that examine the effect of the lattice structure should be emphasized for congruential generators. Extensive methods for testing pseudorandom number generators are given by Knuth [1981] and Marsaglia [1985].

A second property a generator must possess is a long period. For 32-bit computers, periods of $2^{30} \simeq 10^9$ are easily obtained. However, for complex applications longer periods of 10^{18} up to 10^{170} are desirable. The very long-period generators, which are now available, are very useful for generating long independent subsequences so that many collaborators can independently work on subsimulations and later combine their results assured of statistical independence.

A third property a generator must have is that of *portability*. This means that the generator must be written in a high-level language like FORTRAN or C without capitalizing on tricks peculiar to a particular computer. In this way, the same results are obtained on different computers. Generators written in machine language are generally not portable.

A final concern is the efficiency of the generator, i.e., the time it takes to generate a random number. In the early days this was a very important criterion and many machine-dependent tricks were used, that make most early generators nonportable. With modern computers, the generation time is almost insignificant, with the function call time usually exceeding the actual calculation time. If generation time is of concern, two methods can be used to reduce it. First, the generator can be coded in-line to avoid a function call. Second, each function call to the generator should return a large vector of random numbers, thereby, greatly reducing the number of function calls.

3.4 Tests for Congruential Generators

Because of the lattice formed by congruential generators, the effect of the lattice granularity is of great concern. Not only should the k -dimensional hypercube be uniformly filled with lattice points, but the hyperplanes and points also should be close together to reduce the effect of the granularity. Here, several of the statistical tests that have been used to test various congruent generators are reviewed. More details and the theory behind these tests can be found in Fishman [1996].

3.4.1 Spectral Test

The maximum period for a congruential generator with prime m is $t(m) = m - 1$, while for a generation with $m = 2^\alpha$, $t(m) = m/4$. For a given multiplier a and modulus m ,

the distance between adjacent hyperplanes $d_k(a, m)$ is [Cassels 1959]

$$d_k(a, m) \geq d_k^{\min} \equiv \frac{1}{[t(m)]^{1/k}} \begin{cases} (3/4)^{1/4} & k = 2 \\ 2^{-1/6} & k = 3 \\ 2^{-1/4} & k = 4, \\ 2^{-3/10} & k = 5 \\ (3/64)^{1/12} & k = 6 \end{cases} \quad (3.5)$$

where $d_k^{\min}$ is the theoretical lower bound on the plane spacing. For a given modulus m the “best” multiplier a would produce $d_k(a, m)$ that is closest to the minimum possible hyperplane separation. The statistic

$$S_{1,k} = \frac{d_k^{\min}}{d_k(a, m)} \quad (3.6)$$

is such that $0 < S_{1,k} \leq 1$ for $k \geq 2$, and those values of a that have $S_{1,k}$ as close to unity as possible for $2 \leq k \leq 6$ are sought.

3.4.2 Number of Hyperplanes

The number of hyperplanes $N_k(a, m)$ formed by the lattice in the k -dimensional hypercube is bounded by [Marsaglia 1968]

$$N_k(a, m) \leq [k!t(m)]^{1/k}. \quad (3.7)$$

Then the statistic $S_{2,k} \equiv N_k(a, m)/[k!t(m)]^{1/k}$ has values between 0 and 1, and the best multiplier a from this test is the one that produces the largest values of $S_{2,k}$, i.e., the closest values to unity for $2 \leq k \leq 6$.

3.4.3 Distance between Points

The distance between the k -tuples in the k -dimensional hypercube should be as large as possible to ensure the best uniform coverage. Let $r_k(a, m)$ be the Euclidean distance between the nearest points. Then it can be shown that [Cassels 1959]

$$r_k(a, m) \leq \frac{1}{d_k^{\min}(a, m)[t(m)]^{2/k}}, \quad (3.8)$$

where $d_k^{\min}$ is given by Eq. (3.5). The statistic $S_{3,k} \equiv r_k(a, m)d_k^{\min}(a, m)[t(m)]^{2/k}$ has values between 0 and 1, and the best multiplier a from this test is, again, the one that produces the largest values of $S_{3,k}$, i.e., the closest values to unity for $2 \leq k \leq 6$.

3.4.4 Other Tests

Often a test for *discrepancy* is used to evaluate random number generators. This test quantifies how equidistributed the points are in the k -dimensional hypercube compared

to that which would occur with truly random numbers. It, thus, provides a benchmark that allows one to assess how closely the deterministic sequence $\{\rho_i\}$ looks random.

The lattice of k -tuples is defined by a set of k basis vectors $\mathbf{v}_i$. The closer the lengths $|\mathbf{v}_1|, \dots, |\mathbf{v}_k|$ are to each other, the more equidistant are the k -tuples to each other in the k -dimensional hypercube. The *Beyer quotient* [Beyer et al. 1971] is

$$q_k(a, t(m)) \equiv \min_{1 \leq i \leq k} |\mathbf{v}_i| / \max_{1 \leq i \leq k} |\mathbf{v}_i|, \quad k > 1. \quad (3.9)$$

Multipliers with a quotient close to unity are preferred.

Finally, various statistical tests of various hypotheses of samples of length n from the sequence $\{\rho_i\}$ should be made. These include hypotheses that (1) the sample is a sequence of independent identically distributed random variables, (2) the $\{\rho_i\}$, $1 \leq i \leq n$, are distributed uniformly over $(0, 1)$, (3) $\{\rho_{2i-1}, \rho_{2i}\}$, $1 \leq i \leq n/2$, are uniformly distributed over the unit square, (4) $\{\rho_{3i-2}, \rho_{3i-1}, \rho_{3i}\}$, $1 \leq i \leq (n-2)/2$, are distributed uniformly over the unit cube, and (5) all of the previous hypotheses are true simultaneously. Tests for these hypotheses are provided by Fishman [1996].

3.5 Practical Multiplicative Congruential Generators

The congruential random number generator is very fast, requiring very few operations to evaluate x_{i+1} , and, consequently, is now widely used. With some care this generator can be written in high-level computer languages and made portable. To produce very long periods and to reduce the granularity between the random numbers, the modulus m must be a very large integer, typically as large as the computer precision allows.

The generator of Eq. (3.3) can also be written in an alternate form by multiplying it by a^{-1} , the multiplicative inverse of a modulo m , defined as $a^{-1}a = 1 \bmod m$, to obtain

$$x_i = a^{-1}x_{i+1} \bmod m. \quad (3.10)$$

This means that for every sequence of random numbers there is another sequence that is the same but in reverse order provided $a \neq a^{-1} \bmod m$. This is what is observed in Example 3.1.

3.5.1 Generators with $m = 2^\alpha$

When digital computers were first introduced, computing efficiency was of major concern and often m was taken as some large power of 2, thereby enabling rapid calculation of the right side of Eq. (3.3). Just as multiplying (dividing) a decimal number by 10^n is easily done by moving the decimal point n places to the right (left), so is multiplying (dividing) a base 2 (or modulo 1) binary number by 2^n simply a matter of moving the binary point n places to the right (left). Shifting the binary point of a

number in a computer register is much faster than performing multiplication or division. However, such algorithms usually require machine language coding, and, thus, usually are not portable to different computer architectures. However, today the time needed to compute a random number is usually negligible. But $m = 2^\alpha$ is still of major interest, because with large values of α , very long periods can be obtained.

The maximum period of a generator with a modulus 2^α , ($\alpha \geq 3$) is $m/4$ or $2^{\alpha-2}$. This maximum cycle length is realized for any multiplier if $a \bmod 8 = 3$ or 5 and the initial seed is an odd integer [Knuth 1981]. It is found that multipliers with $a = 5 \bmod 8$ produce numbers that are more uniformly distributed than those produced with $a = 3 \bmod 8$. That the seed must be odd is a nuance that a user must remember when using this type of generator.

What multipliers should be used for a given modulus? This is a topic that has received considerable attention with many early proposals having subsequently been shown to be somehow deficient. Fishman [1990] has exhaustively investigated congruential generators with $m = 2^{32}$ and with $m = 2^{48}$ by using the tests of Section 3.4 to find the “best” multipliers. A summary of his findings is given in Table 3.1. The first five multipliers for both $m = 2^{32}$ and $m = 2^{48}$ performed the best in all the statistical tests. For $m = 2^{32}$, the widely used multiplier 69,069 had significantly poorer test scores. Details and test scores are given by Fishman [1996].

A very practical example of a multiplicative congruential generator with modulus $m = 2^{48} \simeq 2.81 \times 10^{14}$ is that employed by the widely used particle transport code

Table 3.1 Recommended multipliers for generators of the form $x_{i+1} = ax_i \bmod 2^\alpha$. The second multiplier in the pairs shown in the first column is $a^{-1} \bmod 2^{32}$, and hence produces the same sequence as a , but in reverse order (after Fishman [1996]).

$m = 2^{32}$		$m = 2^{48}$	
$a = 5^n \bmod 2^{32}$	n	$a = 5^n \bmod 2^{48}$	n
1099087573 ^a	9649599	68909602460261 ^a	528329
4028795517	93795525		
2396548189 ^a	126371437	33952834046453 ^a	8369237
3203713013	245509143		
2824527309 ^a	6634497	43272750451645 ^a	99279091
1732073221	96810627		
3934873077 ^a	181002903	127107890972165 ^a	55442561
1749966429	190877677		
392314069 ^a	160181311	55151000561141 ^a	27179349
2304580733	211699269		
410092949 ^b	—	44485709377909 ^c	66290390456821
69069 ^d	—	19073486328125 ^e	19

^aFishman [1990], ^bBorosh and Niederreiter [1983], ^cDurst as reported by Fishman [1996],

^dMarsaglia [1972], ^eBeyer as reported by Fishman [1996].

MCNP [2003]. This generator has a multiplier $a = 5^{19} = 19,073,486,328,125$. This generator, which has a period of $2^{46} \simeq 7.04 \times 10^{13}$, also uses a stride (see Section 3.7) of 152,917 for each new particle history.⁴

3.5.2 Prime Modulus Generators

Many congruential random number generators take m as a large prime number (or a number with a large prime number as a factor). With a prime number modulus, a full-period generator is obtained for a multiplier a that is a *primitive root* modulo m [Fuller 1976].⁵

To find a primitive root for a given prime modulus, one can test successive values of $a = 2, 3, \dots$ to find a relatively small primitive root a_1 . However, this root is a poor choice for a prime modulus generator because a small value of x_i results in an ascending sequence of small values, hardly what we would expect from random numbers. To avoid such low-order correlations, a primitive root multiplier a should be chosen such that $a \lesssim \sqrt{m}$. To find larger primitive roots from a small primitive root a_1 it can be shown that [Fishman and Moore 1986]

$$a = a_1^n \bmod m \quad (3.11)$$

is also a primitive root if integer n is relative prime to $m - 1$, i.e., if n and $m - 1$ have no factors in common (other than 1). An illustration of finding the primitive roots is given in Example 3.2.

One widely used prime modulus is the Mersenne prime $2^{31} - 1 = 2,147,483,647$.⁶ For this modulus, the smallest primitive root is $k_1 = 7$. The factors of $2^{31} - 1$ are 2, 3, 7, 11, 31, 151, and 331. Thus, $7^5 = 16,807$ is also a primitive root and produces a full cycle length. First, introduced by Lewis, Goodman, and Miller [1969], the random number generator

$$x_{i+1} = 7^5 x_i \bmod (2^{31} - 1) \quad (3.12)$$

has become well accepted. The modulus $m = 2^{31} - 1$ is an obvious choice for the 32-bit architecture of today's personal computers. This generator has passed many empirical tests for randomness, its theoretical properties have been well studied, and, more importantly, it has a vast history of successful applications.

⁴ Beginning with MCNP5, an alternative, optional 63-bit generator is available that has a period of 9.2×10^{18} .

⁵ If the smallest integer k that satisfies $a^k = 1 \bmod m$ is $k = m - 1$, then a is called a primitive root, modulo m .

⁶ Mersenne primes have the form $2^p - 1$, where p is a prime number. Numbers of this form are prime for primes $p \leq 31$ (except for $p = 11, 23$, and 29). However, most large p do not yield primes. One exception is for $p = 32,582,657$, which produces a Mersenne prime number with 9,808,358 digits.

Example 3.2 Primitive Roots

Consider the simple generator $x_{i+1} = ax_i \bmod 13$. What values of a produce a full-cycle generator? For such a prime modulus generator all primitive roots produce full cycles. Thus, first find a small primitive root, i.e., find an a such that the smallest integer k that satisfies $a^k \bmod 13 = 1$ is $k = m - 1 = 12$. It is easily verified that $2^k \bmod 13 = 2, 4, 8, 3, 6, 12, 11, 9, 5, 10, 7, 1$ for $k = 1, 2, \dots, 12$. Hence, $a = 2$ is the smallest primitive root.

The factors of $m - 1 = 12$ are 3, 2, and 2. Then the values of $n < m - 1$ that have no factors (except 1) in common with 3 and 2 are $n = 5, 7$, and 11. Hence, the primitive roots for $m = 13$ are $2, 2^5 \bmod 13 = 6, 2^7 \bmod 13 = 11$, and $2^{11} \bmod 13 = 7$. Thus, $a = 2, 6, 7$, and 11 produce full-period generators, a result previously seen in [Example 3.1](#).

3.5.3 A Minimal Standard Congruential Generator

[Park and Miller \[1988\]](#) have proposed [Eq. \(3.3\)](#) with $m = 2^{31} - 1$ and $a = 16,807$ as a “minimal standard” generator because it passes three important tests: (1) it is a full-period generator, (2) its sequence $\dots x_1, x_2, \dots, x_{m-1}, x_1, \dots$ is “random,” and (3) $ax_i \bmod m$ can be efficiently implemented (with some clever coding discussed in [Section 3.5.4](#)) on a computer using 32-bit integer arithmetic. This is not to say that this particular generator is “ideal.” In fact, it does have some flaws (see [Section 3.5.5](#)). But it is widely used and, as such, is a standard against which other random number generators can be compared.

Is the multiplier $m = 16807$ the optimum multiplier for use with $m = 2^{31} - 1$? Probably not. There are 534,600,000 primitive roots of $m = 2^{31} - 1$, which produce a full period. Hence, test 1 eliminates about 75% of the possible multipliers. [Fishman and Moore \[1986\]](#) examined the remaining 25% for passing test 2. To pass this test, the Euclidean distance between adjacent hyperplanes has to be no more than 25% of the theoretical minimum for dimensions $2 \leq k \leq 6$ (the *lattice test*). Only 410 multipliers met this criterion. The multiplier 16,807 is not among them (nor are other commonly used multipliers). However, *none* of these 410 optimal test-2 multipliers lend themselves to the clever programming trick of test 3 that permits a coding of the generator on machines using 32-bit integers. [Park and Miller \[1988\]](#) found that only 23,093 multipliers pass both tests 1 and 3. By replacing the original 25% criterion for test 2 by a 30% criterion, several good 32-bit multipliers were found, the two best being $a = 48,271$ and $a = 69,621$. Until further testing is done and confidence is gained with these two new multipliers, the original $a = 16,807$ will continue to be used widely.

3.5.4 Coding the Minimal Standard

Coding Park and Miller’s minimal standard random number generator,

$$x_{i+1} = ax_i \bmod m \text{ with } a = 16807 \text{ and } m = 2^{31} - 1 = 2147483647, \quad (3.13)$$

for a computer based on 32-bit integers is not straightforward because, for $1 \leq x_i \leq m-1$, the product ax_i often exceeds the maximum value expressible as a 32-bit integer. One could avoid this problem by using the extended-precision option available in most computer languages. For example, in FORTRAN one could use

```
DOUBLE PRECISION FUNCTION randm(dseed)
DOUBLE PRECISION dseed
  dseed = DMOD(16807.d0*dseed,2147483647.d0)
  randm = dseed/2147483647.d0
RETURN
END
```

However, using double-precision floating-point arithmetic is considerably less efficient than integer arithmetic, especially because this function can be called billions of times in a typical Monte Carlo calculation.

It is much better to use 32-bit integer arithmetic and a trick proposed by [Schrage \[1979\]](#) and improved in 1983 [[Bratley et al. 1983](#)] to avoid the integer overflow problem. In this method, the modulus m is approximately factored as

$$m = aq + r, \quad \text{where } q = [m/a] \quad \text{and} \quad r = m \bmod a. \quad (3.14)$$

With this factorization, the generator of [Eq. \(3.13\)](#) can be evaluated as [[Press et al. 1996](#)]

$$x_{i+1} = ax_i \bmod m = \begin{cases} a(x_i \bmod q) - rk & \text{if } \geq 0 \\ a(x_i \bmod q) - rk + m & \text{if above is } < 0 \end{cases}, \quad (3.15)$$

where $k = [x_i/q]$. How does this reformulation avoid integer overflows? Recall that $1 \leq x_i < m-1$. If $r < q$, then it can be shown [[Park and Miller 1988](#)] that both $a(x_i \bmod q)$ and rk are in the range $0, \dots, m-1$. This factorization with $r < q$ is the requirement for a generator to pass Park and Miller's test 3. For the minimal standard generator ($m = 2^{31} - 1$ and $a = 16,807$), the Schrage's factorization uses $q = 127,773$ and $r = 2836$.

Here is a very efficient FORTRAN routine that implements this algorithm for the minimal standard generator. The value of x_i is input as `iseed` and the value of x_{i+1} is returned in `iseed`. The returned value of `ran` is $\rho_{i+1} = x_{i+1}/m$.

```
FUNCTION ran(iseed)
INTEGER iseed,ia,im,iq,it,k
REAL ran,am
PARAMETER (ia=16807,im=2147483647,am=1./im,
&          iq=127773,it=2836)
k = iseed/iq
iseed = ia*(iseed-k*iq)-it*k
IF(iseed.LT.0) iseed=iseed+im
ran = am*iseed
RETURN
END
```

It is easy to modify this program to use the two optimal multipliers found by [Park and Miller \[1988\]](#), namely, $a = 48,271$ (with $q = 44,488$ and $r = 3399$) and $a = 69,621$ (with $q = 30,845$ and $r = 23,902$).

One final note about programming the minimal standard generator in other computer languages. It is easy to make mistakes, and to check the proper operation of the code, one should start with an initial seed of $x_1 = 1$ and ensure that the generator returns $x_{10001} = 1,043,618,065$. Implementation of this generator in other programming languages is considered in Appendix E.

3.5.5 Deficiencies of the Minimal Standard Generator

The minimal standard generator of [Eq. \(3.13\)](#) is not perfect. Indeed, a small value of ρ_i , say less than 10^{-6} , is produced one time in 10^6 . This value is followed by another small number less than 0.017. For very improbable events, this serial correlation of small ρ_i values can lead to incorrect simulation results. Even more subtle low-order serial correlations exist in this generator [[Press et al. 1996](#)]. With any small multiplier a (needed to pass Park and Miller's test 3), all congruential generators have low-order serial correlations.

To minimize low-order correlations, the multiplier of a multiplicative congruential generator should be approximately equal to the square root of the modulus, which is more restrictive than the requirement that $r < q$ in order to pass Park and Miller's test 3.

3.5.6 Optimum Multipliers for Prime Modulus Generators

As [Park and Miller \[1988\]](#) suggest, the minimal standard's multiplier $a = 16,807$ is probably not the best multiplier for $m = 2^{31} - 1$. In a massive effort, [Fishman and Moore \[1986\]](#) exhaustively studied all possible primitive roots for this modulus, including the values 48,271 and 69,621 suggested by [Park and Miller \[1988\]](#) as being superior to the minimal standard multiplier. The results of the "best" multipliers are summarized in [Table 3.2](#). The multiplier $a = 62,089,911$ was overall the best, with $a = 69,621$ the best for multipliers with $r < q$. Although 16,807 is not unreasonable,

Table 3.2 Recommended multipliers for prime modulus congruential random number generators (after [Fishman \[1996\]](#)).

$m = 2^{31} - 1$	multipliers a with $r < q$			
multiplier a	modulus m	multiplier a	$q = \lfloor m/a \rfloor$	$r = m \bmod a$
742938285 ^a	$2^{31} - 1$	48271 ^b	44488	3399
950706376 ^a	$2^{31} - 1$	69621 ^b	30845	23902
1226874159 ^a	$2^{31} - 1$	16807 ^c	127773	2836
62089911 ^a	$2^{31} - 1$	39373 ^d	54542	1481
1343714438 ^a	$2^{31} - 85$	40014 ^d	53668	12211
630360016 ^e	$2^{31} - 249$	40692 ^d	52774	3791

^aFishman [1996], ^bPark and Miller [1988], ^cLewis et al. 1969, ^dL'Ecuyer [1988],

^ePayne et al. [1969].

it did significantly worse in the tests for randomness than did all the other multipliers shown in this table.

3.6 Shuffling a Generator's Output

The simplicity and ease of implementation of congruential random number generators have resulted in their widespread use despite known limitations such as low-order correlations and potentially imperfect equidistributedness. Generators with carefully chosen multipliers for a given modulus can produce k -tuples that are somewhat equidistributed. Is there some minor change that could be made in the generation process to make the k -tuples (or lattice) more equidistributed and random?

One simple idea is to shuffle the numbers produced by a generator so that all generated numbers appear in the output stream, but in an order different from the order in which they were generated. To accomplish this, one could use the output of a second (and perhaps simpler) generator to shuffle or permute the output sequence from the main generator. Such shuffling can increase the period because no longer does the same value follow a given value every time it occurs. Moreover, the shuffling breaks up the original lattice structure replacing it with a different one with a different number of hyperplanes. Another benefit of shuffling is that it removes low-order correlations.

Bays and Durham [1976] have shown that a single generator can be used to first fill a table (vector) with x_i values and then use subsequent values from the random number generator to select and replace particular table entries. In this way, the output from the random number generator are shuffled and low-ordered correlations are removed. This shuffling can also increase the period of the generator.

The Bays and Durham shuffling algorithm is described below. The generator produces random values x_i and the output stream from the shuffling is denoted by y_i . The shuffling table (vector) has length k .

1. (Initialization) Fill the table vector T_j with x_j , $j = 1, \dots, k$; place $y_1 = x_{k+1}$ as the first entry in the output stream; and set $i = 1$.
2. Compute the table pointer $j = 1 + [ky_i/(m - 1)]$.
3. Set $i = i + 1$.
4. Place $y_i = T_j$ into output stream.
5. Get the next number x_{k+i} and place it in T_j .
6. Loop back to step 2 to place another number into the output stream.

As an example of the effectiveness of this shuffling consider the example shown in the left plot of Fig. 3.1 where the full-cycle generator $x_{i+1} = 3x_i \bmod 29$ produces far from uniform distributions of 2-tuples over the square. With the Bays and Durham shuffling algorithm applied to this generator, the results shown in Fig. 3.2 are obtained. The cycle length is at least several thousand, and the plot area is much more uniformly covered.

Shuffling programs for the minimal standard random number generator are given by Press et al. [1996]. With such shuffling, even more statistical tests for randomness are met.

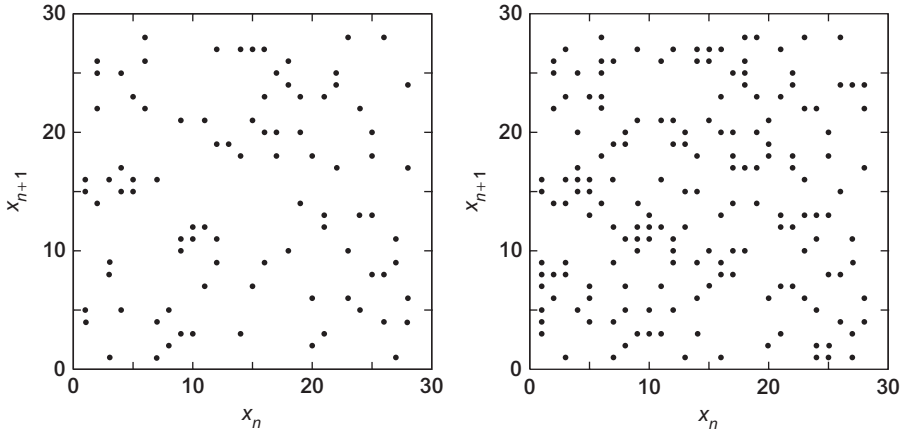


Figure 3.2 The effect of using the Bays and Durham shuffling scheme on the random number generator $x_{i+1} = 3x_i \bmod 29$ (used to generate the pair distribution of the left plots in Fig. 3.1). The left plot is for the first 101 pairs and the right plot is for the first 201 pairs. Besides a much longer cycle length, a more uniform distribution is achieved.

3.7 Skipping Ahead

Often it is useful to generate separate independent sequences of random numbers from the same congruential generator. For example, with a very long period generator, independent substrings could be obtained so that different machines could work simultaneously on the same simulation and the independent results from each machine can, therefore, be combined.

It is not a good idea to simply use a different seed for each sequence, because with bad choices the resulting sequences may overlap significantly. The best way to generate separate independent sequences is to skip ahead a specified distance in the sequence to start a new subsequence.

This skipping ahead is easily done for a congruential generator

$$x_{i+1} = ax_i \bmod m. \quad (3.16)$$

By recursive application of this relation it is easily shown (see problem 2) that

$$x_{i+k} = a^k x_i \bmod m. \quad (3.17)$$

With this relation, it is easy to start a new sequence some number of integers ahead of the start of the previous sequence.

Suppose the n th sequence started with x_n . Then the next sequence must begin with x_{n+s} , where s is the jump or *stride* between the two sequences. From Eq. (3.17) the $(n+1)$ th sequence is begun with the integer

$$x_{n+s} = a^s x_n \bmod m = bx_n \bmod m, \quad \text{where } b = a^s \bmod m. \quad (3.18)$$

The remainder of the new sequence is then obtained from Eq. (3.16).

3.8 Combining Generators

An obvious way to increase the period and randomness of a random number generator is to use two or more generators and merge their outputs. For example, Collins [1987] suggested using many generators each producing an output stream of random numbers. At each step, the output from one generator is used to select randomly one of the other generators of which the number is selected as the random number for that step. Such a procedure would vastly increase the cycle length. However, it is rather inefficient in that it requires many wasted calculations.

A note of caution. When one combines generators of which the output is in the open interval $(0, 1)$, it is possible, because of the finite machine precision, to obtain a 0 or a 1 (see problem 10). Special care must be exercised when coding such generators to prevent an endpoint being produced.

3.8.1 Bit Mixing

A very common way of improving randomness and increasing cycle length is to combine the integers x_i and y_i produced by two different generators to generate a new integer $z_i = x_i \odot y_i$, where $\odot$ is some binary operator, typically addition or subtraction modulo one, to produce a mixing of the bits in the base 2 representations of x_i and y_i . This is very efficient and easily implemented.

3.8.2 The Wichmann-Hill Generator

Wichmann and Hill [1982] proposed the following generator:

$$\begin{aligned}x_{i+1} &= 171x_i \bmod 30269 \\y_{i+1} &= 172y_i \bmod 30307 \\z_{i+1} &= 170z_i \bmod 30323\end{aligned}$$

with the i th random number in $(0, 1)$ calculated as

$$\rho_i = \left(\frac{x_i}{30269} + \frac{y_i}{30307} + \frac{z_i}{30323} \right) \bmod 1. \quad (3.19)$$

This generator requires three seeds (x_0, y_0, z_0) and because the moduli are relatively prime, this generator cannot yield an exact zero (although finite machine precision requires careful programming to avoid zero). The period of this generator is about 10^{12} . This generator has been investigated for its long-range correlations and found to be comparable to other good generators [De Matteis and Pagnutti 1993]. Moreover, it is easily programmed for a 16-bit machine.

3.8.3 The L'Ecuyer Generator

L'Ecuyer [1988] has suggested a way to combine several different sequences with different periods so as to generate a sequence of which the period is the least common

multiple of the two. Such a combination greatly helps reduce the serial correlations present in each generator. In particular, L'Ecuyer proposed the following generator [Gentle 1998]:

$$x_i = 40001x_{i-1} \bmod 2147483563 \quad (3.20)$$

$$y_i = 40692y_{i-1} \bmod 2147483399 \quad (3.21)$$

$$z_i = x_i - y_i, \quad (3.22)$$

where, if $z_i < 1$, then $z_i = z_i + 2,147,483,562$. The random number uniformly distributed in $(0, 1)$ is given by

$$\rho_i = \frac{z_i}{2147483589} = 4.656613z_i \times 10^{-10}. \quad (3.23)$$

Because x_i and y_i are uniform distributions over almost the same interval, the difference is also uniform. Moreover, a value of $\rho_i = 1$ cannot be obtained because the normalizing constant is slightly greater than 2,147,483,536. The period of this generator is of order 10^{18} , and L'Ecuyer claims this generator produces a good sequence of random numbers. It can also be programmed on a 32-bit machine (see James [1990] for a portable FORTRAN version). It has one minor limitation. It is not easy to generate separate independent sequences because it is not easy to skip ahead with this algorithm.

Press et al. [1996] offer the RAN2 program that is based on this L'Ecuyer generator with a final Bays-Durham shuffle added. This generator produces a sequence with a period of $> 2.3 \times 10^{18}$. The authors claim this generator is “perfect.”

3.9 Other Random Number Generators

Variations to the congruential generators and some other random number generators have been proposed. The basic idea is that by making the generator more complex better randomness and longer periods can be achieved. The properties of many of these generators are often not well established, and the problem of finding optimum generators (in some sense) is still an area of active research. In the following, a few of these generator variants are briefly outlined. For a more in-depth discussion and for summaries of even more generators, the reader is referred to Gentle [1998] and Fishman [1996].

3.9.1 Multiple Recursive Generators

Instead of using only x_i to generate the next random number x_{i+1} , one could use several earlier numbers by using

$$x_i = (a_1x_{i-1} + a_2x_{i-2} + \cdots + a_nx_{i-n}) \bmod m. \quad (3.24)$$

Some of these *multiple recursive* multiplicative congruential generators can have much longer cycles than a simple multiplicative congruential generator. Such generators

have been studied by L'Ecuyer et al. [1993] who make recommendations for various moduli (including $m = 10^{31} - 1$), give a portable code for $n = 5$ that does not yield a 0 or 1, and show how to skip ahead in the sequence.

3.9.2 Lagged Fibonacci Generators

Reducing the famous Fibonacci sequence $x_i = x_{i-1} + x_{i-2}$ modulo m has been proposed as the basis for a random number generator. However, it has poor random properties. Much better results are obtained if two earlier results some distance apart are combined. Thus, the *lagged Fibonacci congruential generator* is

$$x_i = (x_{i-j} + x_{i-k}) \bmod m. \quad (3.25)$$

If the lags j and k and the modulus m are chosen properly good random sequences can be obtained [Altman 1989] that can have cycle lengths as large as $m^k - 1$ [Gentle 1998]. Knuth [1981] gives suggested values for the lags j and k that produce periods for $j > k$ of $(2^j - 1)(2^{m-1})$.

Marsaglia et al. [1990] have proposed a lagged Fibonacci generator with lags of 97 and 33 that has a period of $2^{144} \simeq 10^{43}$. Moreover, it is portable and gives bit-identical results on all 32-bit machines and has passed extensive statistical tests despite limited theoretical understanding of this type of generator. James [1990] gives a very clever FORTRAN program RANMAR for this generator. A very useful feature of RANMAR is the ease with which independent subsequences can be obtained. The initialization of this generator requires a 32-bit integer. Each different integer produces an independent (nonoverlapping) sequence of average length $\simeq 10^{30}$, enough for very complex simulations.

3.9.3 Add-with-Carry Generators

A variant of the lagged Fibonacci generator, introduced by Marsaglia and Zaman [1991], is the *add-with-carry* generator

$$x_i = (x_{i-j} + x_{i-k} + c_i) \bmod m, \quad (3.26)$$

where the “carry” $c_1 = 0$ and c_{i+1} is given by

$$c_{i+1} = \begin{cases} 0, & \text{if } x_{i-j} + x_{i-k} + c_i < m \\ 1, & \text{otherwise} \end{cases}.$$

With proper choices for j , k , and m , very long period generators can be obtained. Marsaglia and Zaman give values that produce periods of the order of 10^{43} . Moreover, this generator can be implemented in base-2 arithmetic. It has since been shown that the add-with-carry generator is equivalent to a linear congruential generator with a very large prime modulus [Tezuka and L'Ecuyer 1992], which gives one some confidence in this generator and also shows how to program congruential generators with large prime moduli.

Other variations of this generator are the subtract-with-carry and the multiply-with-carry, also introduced by Marsaglia and Zaman [1991]. Much work still needs to be done to determine the random properties and optimal parameter values for these generators.

James [1990] gives a FORTRAN program for 32-bit machines for the subtract-with-carry variant. He uses $j = 24$, $k = 10$, and $m = 2^{24}$ to produce a generator with a period of $\simeq 2^{570} \simeq 10^{171}$!

3.9.4 Inversive Congruential Generators

The multiplicative inverse x^{-} of a number $x \bmod m$ is defined by $1 = x^{-}x \bmod m$. Eichenauer and Lehn [1986] proposed a random number generator based on the multiplicative inverse, namely,

$$x_i = (ax_{i-1}^{-} + c) \bmod m. \quad (3.27)$$

This generator does not yield regular lattice hyperplanes as does a linear congruential generator, and it appears to have better uniformity and fewer serial correlation problems. However, calculation of multiplicative inverses is computationally expensive.

3.9.5 Nonlinear Congruential Generators

Knuth [1981] suggested the generator

$$x_i = (dx_{i-1}^2 + ax_{i-1} + c) \bmod m \quad \text{with} \quad 0 \leq x_i < m. \quad (3.28)$$

Other polynomials could be used as well. But the properties of such generators are not well established. Eichenauer et al. [1988] consider a more general variant, namely,

$$x_i = f(x_{i-1}) \bmod m, \quad (3.29)$$

with a wide variety of functions $f(x)$. Although such generators can produce very good uniformity properties, not enough is known yet for serious application of these generators. There needs to be much more study about nonlinear generators if they are to become useful.

3.10 Summary

Although Monte Carlo methods depend critically on procedures to generate efficiently a sequence of pseudorandom numbers that exhibit many of the qualities of true random numbers, a vast number of studies have relied on random number generators that are now known to be highly deficient. Many computer languages and code packages have built-in random number generators and, often, they are far from optimal. Indeed, it would be very interesting to see how many studies based on these inferior generators would have produced different results if they were based on modern proven random number generators. *Know thy random number generator!*

Bibliography

- ALTMAN, N.S. 1989, "Bit-wise Behavior of Random Number Generators," *SIAM J. Sci. Stat. Comp.*, **9**, 839–847.
- BAYS, C. AND S.D. DURHAM 1976, "Improving a Poor Random Number Generator," *ACM Trans. Mathematical Software*, **2**, 59–64.
- BEYER, W.A., R.B. ROOF, AND D. WILLIAMSON 1971, "The Lattice Structure of Multiplicative Congruential Pseudorandom Vectors," *Math. Comput.*, **25**, 345–363.
- BOROSH, I. AND H. NIEDERREITER 1983, "Optimal Multipliers for Pseudorandom Number Generation by the Linear Congruential Method," *BIT*, **23**, 65–74.
- BRATLEY, P., B.L. FOX, AND E.L. SCHRAGE 1983, *A Guide to Simulation*, Springer-Verlag, New York.
- CASELS, J.W.S. 1959, *An Introduction to the Geometry of Numbers*, Springer-Verlag, Berlin.
- COLLINS, B.J. 1987, "Compound Random Number Generators," *J. Am. Stat.*, **82**, 525–527.
- DE MATTEIS, A. AND S. PAGNUTTI 1993, "Long-Range Correlation Analysis of the Wichmann-Hill Random Number Generator," *Stat. Comput.*, **3**, 67–70.
- EICHENAUER, J., H. GROTHE, AND J. LEHN 1988, "Marsaglia's Lattice Test and Non-Linear Congruential Pseudo Random Number Generators," *Metrika*, **35**, 241–250.
- EICHENAUER, J. AND J. LEHN 1986, "A Nonlinear Congruential Pseudo Random Number Generator," *Statistische Hefte*, **27**, 315–255.
- FISHMAN, G.S. 1990, "Multiplicative Congruential Random Number Generators with Modulus 2^β : an Exhaustive Analysis for $\beta = 32$ and a Partial Analysis for $\beta = 48$," *Math. Comput.*, **54**, 331–334.
- FISHMAN, G.S. 1996, *Monte Carlo: Concepts, Algorithms, and Applications*, Springer-Verlag, New York.
- FISHMAN, G.S. AND L.R. MOORE 1986, "An Exhaustive Analysis of Multiplicative Congruential Random Number Generators with Modulus $2^{31} - 1$," *SIAM J. Sci. Stat. Comp.*, **7**, 24–25.
- FULLER, A.T. 1976, "The Period of Pseudo-Random Numbers Generated by Lehmer's Congruential Method," *Comput. J.*, **19**, 173–177.
- GENTLE, J.E. 1998, *Random Number Generation and Monte Carlo Methods*, Springer-Verlag, New York.
- JAMES, F. 1990, "A Review of Pseudorandom Number Generators," *Comput. Phys. Comm.*, **60**, 329–344.
- KNUTH, D.E. 1981, Semi-Numerical Algorithms, Vol. 2, in *The Art of Computer Programming*, 2nd ed., Addison-Wesley, Reading, MA.
- L'ECUYER, P. 1988, "Efficient and Portable Combined Random Number Generators," *Commun. ACM*, **31**, 742–749.
- L'ECUYER, P., F. BLOUIN, AND R. COUTURE 1993, "A Search for Good Multiplicative Recursive Random Number Generators," *ACM T. Model. Comput. S.*, **3**, 87–98.
- LEHMER, D.H. 1951, "Mathematical Methods in Large-Scale Computing Units," *Annu. Comput. Lab. Harvard Univ.*, **26**, 141–146.
- LEWIS, P.A.W., A.S. GOODMAN, AND J.M. MILLER 1969, "A Pseudo-Random Number Generator for the System/360," *IBM Syst. J.*, **8**, 136–146.
- MARSAGLIA, G. 1968, "Random Numbers Fall Mainly in the Plane," *P. Natl. Acad. Sci. USA*, **61**, 25–28.
- MARSAGLIA, G. 1972, "The Structure of Linear Congruential Sequences," in *Applications of Number Theory to Numerical Analysis*, S.K. ZAREMBA, ed., Academic Press, New York.
- MARSAGLIA, G. 1985, "A Current View of Random Number Generators," in *Computer Science and Statistics: The Interface*, L. BILLARD, ed., Elsevier, Amsterdam.

- MARSAGLIA, G. 1995, *The Marsaglia Random Number CDROM, Including the DIEHARD Battery of Tests of Randomness*, Dept. Statistics, Florida State University, Tallahassee.
- MARSAGLIA, G., A. ZAMAN, AND W.-W. TSANG 1990, *Stat. Probabil. Lett.*, **9**, 35.
- MARSAGLIA, G. AND A. ZAMAN 1991, "A New Class of Random Number Generators," *Ann. Appl. Probab.*, **1**, 462–480.
- MCNP 2003, *A General Monte Carlo N-Particle Transport Code, Version 5*, Vol. I *Overview and Theory*, LA-UR-03-1987, Los Alamos National Lab., Los Alamos, NM.
- PARK, S.K. AND K.W. MILLER 1988, "Random Number Generators: Good Ones Are Hard to Find," *Commun. ACM*, **31**, 1192–1201.
- PAYNE, W.H., J.R. RABUNG, AND T.P. BOGYO 1969, "Coding the Lehmer Pseudorandom Number Generator," *Commun. ACM*, **12**, 85–86.
- PRESS, W.H., S.A. TEUKOLSKY, W.T. VETTERLING, AND B.P. FLANNERY 1996, *Numerical Recipes*, 2nd ed., Cambridge University Press, Cambridge.
- SCHRAGE, L. 1979, "A More Portable FORTRAN Random Number Generator," *ACM T. Math. Software*, **5**, 132–138.
- TEZUKA, SHU AND P. L'ECUYER 1992, "Analysis of Add-with-Carry and Subtract-with-Borrow Generators," *Proc. 1992 Winter Simulation Conference*, 443–447, Assoc. Comp. Mach., New York.
- WICHMANN B.A. AND I.D. HILL 1982, "An Efficient and Portable Pseudorandom Number Generator," *Applied Statistics*, **31**, 188–190 (corrections, 1984 *ibid.* **33**, 123).

Problems

1. Repeat the analysis for [Example 3.1](#) using different seeds.
2. All variables in this problem are integers. The congruence $a \equiv b \pmod m$ means that $(b - a)$ is divisible by m . (a) From this observation, show that

$$ia \equiv ib \pmod m \quad \text{and} \quad a^j \equiv b^j \pmod m,$$

and (b) if $x_{n+1} \equiv ax_n \pmod m$ $n = 0, 1, \dots$ show that

$$x_n \equiv a^n x_0 \pmod m.$$

3. Show that the congruential generator $x_{i+1} \equiv ax_i \pmod m$ implies $x_{i+k} \equiv a^k x_i \pmod m$. Illustrate with a simple example.
4. In [Fig. 3.1](#) overlapping pairs of numbers (x_1, x_2) , (x_2, x_3) , $\dots$ were used to define the points on the plane. What would the plots look like if adjacent pairs (x_1, x_2) , (x_3, x_4) , $\dots$ were used?
5. What is the minimum distance between hyperplanes in the hypercube of dimensions $k = 2, \dots, 6$ for a congruential generator with modulus (a) $m = 2^{31} - 1$ and (b) $m = 2^{48}$? Comment on the granularity of the lattice points.
6. What is the maximum number of hyperplanes in the hypercube of dimensions $k = 2, \dots, 6$ for a congruential generator with modulus (a) $m = 2^{31} - 1$ and (b) $m = 2^{48}$? What do these results tell you about the equidistributedness of the lattice points?
7. Using C or FORTRAN, write a multiplicative congruential generator based on the Mersenne prime $m = 2^8 - 1 = 31$. (a) What multipliers produce full periods? (b) Make plots of overlapping pairs of numbers similar to those in [Fig. 3.1](#) and determine which multiplier(s) give the most uniform covering of the plane.

8. Determine the primitive roots for the generator $x_{i+1} = ax_i \bmod 17$ that produce full-cycle generators.
9. For many years the program RANDU was widely used. This generator is defined by

$$x_{i+1} = 65539x_i \bmod 2^{31}.$$

Although this generator has good random properties for the two-dimensional lattice, it has a very poor three-dimensional lattice structure because all triplets (x_i, x_{i+1}, x_{i+2}) lie on only 15 planes. Clearly, if such a triplet were used to define a starting location for a source particle in a transport calculation, the source would be poorly modeled.

Write a FORTRAN or C program to implement this generator and generate a sequence of length 30,002. For all triplets in this sequence, (x_i, x_{i+1}, x_{i+2}) , in which $0.5 \leq x_{i+1} \leq 0.51$, plot x_i versus x_{i+2} . What does this plot tell you about this generator [Gentle 1998]?

10. Consider the combined generator of Eq. (3.19). Because of the relative primeness of the moduli, an exact 0 cannot be obtained. But could a code using integer arithmetic produce a 0? First, consider a computer with a very small binary integer, say, 2 bits; then extend to larger integer precisions. For a 32-bit computer, what is the probability a 0 is obtained?

4 Sampling, Scoring, and Precision

*All the girls thought Tom was a bore,
and with them he had little rapport.
But studying at nights
Monte Carlo's delights,
he finally learned how to score.*

The Monte Carlo method relies inherently on the concepts of *sampling* and *scoring*. Sampling is the process of picking a particular value of a random variable from its governing distribution. For example, suppose a random variable is governed by the exponential distribution. Then, sampling that variable generates numerical values of the variable such that a large number of samples reproduces the exponential distribution (at least approximately). Scoring is the process whereby the expected value of some random variable, such as $z(x)$, is estimated. In this case, the score of a Monte Carlo history is the calculation of $z(x_i)$ that results from the i th sampling. In the simplest case, the score is zero or one (failure or success). The overall score, $\bar{z}$, is the cumulative average of N history scores. Some of these concepts and methods for implementing them are explored in this chapter. Another formal method of sampling that is sometimes used when the PDF is ill-defined or otherwise difficult to sample by straightforward means is based on Markov chains. This method has evolved into its own procedure, called Markov chain Monte Carlo (MCMC), which is treated separately in Chapter 6. However, MCMC is in essence just another way of sampling in order to estimate expectations from sample averages.

4.1 Sampling

A variety of techniques can be used to generate values of a random variable that has a specific distribution. Perhaps the most frequently used method to sample from a discrete distribution that has two equally probable values is to flip a coin. Tossing a die can be used to generate values of a random variable that has six equally probable outcomes. Many lotteries select a winning number by randomly picking numbered tiles or balls from a cage in which the contents are mixed. Picking numbered pieces of paper from a hat can also be used to generate random numbers. Throwing darts at the stock pages of the *Wall Street Journal* can be used to randomly select stocks for your portfolio. Dropping matchsticks onto a hardwood floor can be used as a physical simulation of Buffon's needle problem.

Many random physical processes can also be used to generate random numbers with various distributions. “White noise” in electronic circuits can be used to generate random values from a uniform distribution. The time interval between successive decays in a radioactive source can be shown to have an exponential distribution. The hiss from the cosmic microwave background radiation pervading our universe likewise can be used to generate sequences of random numbers.

However, computer-generated random numbers are almost always used in Monte Carlo calculations, in which thousands, millions, or even billions of random numbers are routinely needed. Such computer-generated numbers are not truly random, but proper algorithms can generate extremely long sequences of numbers that pass many statistical tests for randomness. These so-called *pseudorandom* number generators were discussed in Chapter 3.

All the sampling techniques discussed in this chapter for generating values of a random variable x with a specific distribution are based on having a pseudorandom number generator that can produce long sequences of pseudorandom numbers ρ_i that are uniformly distributed on the interval $[0, 1]$.¹ How this is done computationally was discussed in Chapter 3. The symbol ρ_i is used in this book to refer to a random (or pseudorandom) number sampled from the uniform distribution on the unit interval (also called the unit rectangular distribution).

4.1.1 Inverse CDF Method for Continuous Variables

There is a one-to-one correspondence between any two CDFs, because CDFs are monotone increasing functions between zero and unity. This result indicates that the probability that x , of which the PDF is $f(x)$ and CDF is $F(x)$, has a value less than x_i is equal to the probability that y , of which the PDF is $g(y)$ and CDF is $G(y)$, has a value less than y_i , i.e.,

$$F(x_i) = G(y_i). \quad (4.1)$$

Now, the unit rectangular PDF $\mathcal{U}(0, 1)$ (see Appendix A) has the CDF $U(\rho) = \int_0^\rho d\rho' = \rho$. Replace $G(y_i)$ with $U(\rho_i) = \rho_i$ in Eq. (4.1) to obtain

$$F(x_i) = \rho_i. \quad (4.2)$$

The inverse CDF method begins by selecting a number ρ_i uniformly between 0 and 1. Then the desired value of x_i , distributed as $f(x)$, is obtained by solving

$$x_i = F^{-1}(\rho_i), \quad (4.3)$$

¹ The notation $\mathcal{U}(a, b)$ is used to represent a uniform distribution over the interval $[a, b]$. That ρ_i are random numbers uniformly distributed on the interval $[0, 1]$ can be written more compactly as $\rho_i \leftarrow \mathcal{U}(0, 1)$.

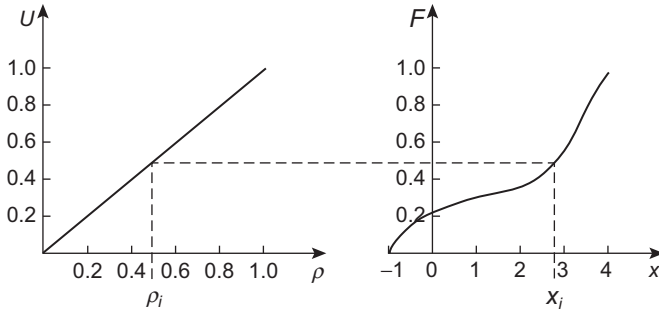


Figure 4.1 A pseudorandom number ρ_i selected from the unit rectangular distribution corresponds to one and only one value x_i of another random variable of which the CDF is given by F .

where F^{-1} is the function obtained by solving Eq. (4.2) for x_i . Although this is a very efficient method for generating samples of random variables (or variates) with a given distribution $f(x)$, it is limited to those distributions for which Eq. (4.2) can be solved for x_i , i.e., to distributions of which the inverse F^{-1} is known. Inverse CDF straightforward sampling is illustrated schematically in Fig. 4.1.

Example 4.1 Inverse CDF Method for a Uniform Distribution

Consider how to obtain a sample from a uniform distribution defined on the interval $[a, b]$. From Eq. (4.2) one can write

$$\rho_i = F(x_i) = \frac{1}{b-a} \int_a^{x_i} du = \frac{x_i - a}{b-a}.$$

Solving for x_i then yields

$$x_i = F^{-1}(\rho_i) = a + \rho_i(b-a).$$

Example 4.2 Inverse CDF for an Exponential Distribution

Consider sampling from an exponential distribution $f(x) = \alpha e^{-\alpha x}$ with $x \in [0, \infty)$ and $\alpha > 0$. The CDF for this distribution with parameter α can be written as

$$F(x) = \int_0^x \alpha e^{-\alpha x'} dx' = 1 - e^{-\alpha x}.$$

The random variable x_i is then found by solving Eq. (4.2), i.e.,

$$\rho_i = 1 - e^{-\alpha x_i},$$

to obtain

$$x_i = -\frac{1}{\alpha} \ln(1 - \rho_i).$$

Because the random quantity $1 - \rho_i$ is also distributed according to $U(0, 1)$, a more computationally efficient result is

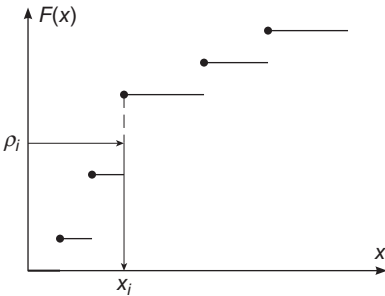
$$x_i = -\frac{1}{\alpha} \ln(\rho_i),$$

thereby saving one subtraction operation each time a sample is generated.

The validity of the inverse CDF method, or any other sampling procedure, can be established by showing that the probability of obtaining a value of x within the interval between some fixed value x and $x + dx$ is $f(x) dx$. For the inverse CDF method, one notes that the value of x_i corresponds to the random number ρ_i , and the probability of x_i being between x and $x + dx$ is the same as

$$\begin{aligned} \text{Prob}\{x < F^{-1}(\rho_i) < x + dx\} &= \text{Prob}\{F(x) < \rho_i < F(x + dx)\} \\ &= F(x + dx) - F(x) \\ &= \int_x^{x+dx} f(x') dx' = f(x) dx. \end{aligned} \quad (4.4)$$

4.1.2 Inverse CDF Method for Discrete Variables



The inverse CDF method also can be used for discrete variables as illustrated in Fig. 4.2. Suppose that the random variable x has only discrete values x_i with probabilities f_i . In other words, the PDF is

$$f(x) = \sum_{i=1}^N f_i \delta(x - x_i), \quad (4.5)$$

so that the CDF is given by Eq. (2.34)

$$F(x) = \sum_{x_i \leq x} f_i. \quad (4.6)$$

To generate values of x that have this distribution, first select a random number ρ uniformly from $\mathcal{U}(0, 1)$, and then find the number x_i that satisfies

$$f_1 + \cdots + f_{i-1} < \rho \leq f_1 + \cdots + f_i, \quad (4.7)$$

or, equivalently,

$$F(x_{i-1}) < \rho \leq F(x_i). \quad (4.8)$$

To find x_i , the following simple methodical search algorithm can be used:

1. Pick ρ from $\mathcal{U}(0, 1)$.
2. Set $i = 1$.
3. If $\rho \leq F(x_i)$, return x_i .
4. Otherwise, set $i = i + 1$ and go to step 3.

For many discrete distributions, such as the Poisson, binomial, and geometric distributions (see Appendix A for details), the mass points x_i , i.e., the values of x for which $f(x) \neq 0$, are the integers $0, 1, 2, \dots$. Moreover, for these distributions, the ratio of successive PDF values $R(x) = f(x+1)/f(x)$ is often a simple function of x that can be used to construct values of $F(x)$ while searching for the x that satisfies

$$F(x-1) < \rho \leq F(x). \quad (4.9)$$

The above search algorithm is thus modified as follows:

1. Pick ρ from $\mathcal{U}(0, 1)$.
2. Set $x = 1$, $\Delta F = f(0)$, and $F = \Delta F$.
3. If $\rho \leq F(x)$, return x .
4. Otherwise, set $x = x + 1$, $\Delta F = R(x)\Delta F$, $F = F + \Delta F$, and go to step 3.

Using the inverse CDF method for a discrete distribution is essentially a table lookup procedure; the search can be performed sequentially, as in the above algorithms, or by some tree traversal. Faster sampling times can be achieved if nodes with high probability are tested before nodes with low probability. For example, in unimodal discrete distributions with a large number of nodes, it is better to start at the node with the greatest probability. Various other methods have been proposed for speeding a table search (see [Marsaglia \[1963\]](#); [Norman and Cannon \[1972\]](#); and [Chen and Asau \[1974\]](#)).

Other Monte Carlo applications of table lookups

Often a continuous random variable may have a table of cumulative values but an unknown function for the CDF over all or part of the range. For example, an analyst may be able to assign probabilities in certain regions of the distribution but be unable to supply an underlying functional form. Or the continuous distribution may be produced by a previous Monte Carlo calculation and is known only in a “binned” or histogram form. Sometimes a very computationally expensive function must be evaluated billions of times, and it is faster to tabulate the functions and use the faster table lookup method. To sample from such tabulated distributions, some approximate interpolating function is fit to histogram data (a piecewise linear function is easiest, although quadratic and cubic functions can be used) and the inverse CDF method is then used to obtain random variates.

Example 4.3 Sampling from a Discrete Distribution

Consider the flipping of a coin. Let x be the random variable that describes whether a head ($x_1 = 1$) or tail ($x_2 = 0$) is obtained on a flip. The head outcome has probability $f_1 = 0.5$ (assuming a “fair” coin) and the tail outcome has probability $f_2 = 0.5$. To generate each sample x_i , a pseudorandom number ρ_i , uniformly distributed over $[0, 1]$, is generated. If $\rho_i < 0.5$, then set $x_i = 0$; otherwise, set $x_i = 1$. This procedure is an example of the generation of a random sample from a Bernoulli distribution (see Appendix A).

4.1.3 Rejection Method

The inverse CDF method presents difficulties if the PDF $f(x)$ cannot be integrated analytically to obtain $F(x)$ or, even if $F(x)$ can be obtained easily, $F^{-1}(\rho)$ is difficult to obtain analytically. An alternate approach in such cases, first proposed by von Neumann [1951], is the *rejection method*. Consider Fig. 4.3, which illustrates such a PDF $f(x)$, defined analytically within the range $[a, b]$, and known to be less than or equal to M . Two steps are required to select x_i from the $f(x)$ distribution:

1. Select a random value of $x \in [a, b]$ as $x_i = a + \rho_i(b - a)$.
2. Select another random number ρ_j . If $\rho_j M \leq f(x_i)$, then accept x_i . Otherwise, return to step 1.

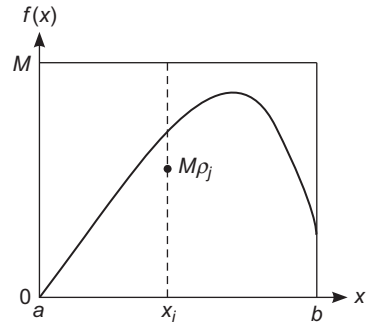


Figure 4.3 Basic rejection method.

This process of selecting values of x is simple. The efficiency, or probability of not rejecting x_i , is given by the ratio of the areas $\int_a^b dx f(x)$ and $M(b - a)$, namely, $1/[M(b - a)]$. This efficiency may be very low if M is chosen inordinately large or if $f(x)$ is highly peaked. Improvements in efficiency may be achieved as follows.

Consider Fig. 4.4, which illustrates a PDF $f(x)$ along with a similar PDF $g(x)$ defined within the same limits and suitable for the CDF method, i.e., to have an easily obtainable CDF $G(x)$, which can in turn be easily inverted to give $G^{-1}(\rho)$. Note that

$$F(x) = \int_a^x dx' h(x')g(x'), \quad \text{where} \quad h(x) \equiv \frac{f(x)}{g(x)}. \quad (4.10)$$

If $g(x)$ is a close approximation to $f(x)$, then $h(x)$ does not deviate markedly from unity. At any rate, $g(x)$ must be chosen so that $h(x)$ does not exceed some maximum bounding value, $h_{\max}$, in the interval a to b . It is more efficient, but not mandatory, to make $h_{\max}$ the actual maximum value of h between the limits a and b . The rejection technique proceeds as follows:

1. Select a random number ρ_1 .
2. Obtain a tentative value of x from the equation $x_t = G^{-1}(\rho_1)$.
3. Select another random number ρ_2 .

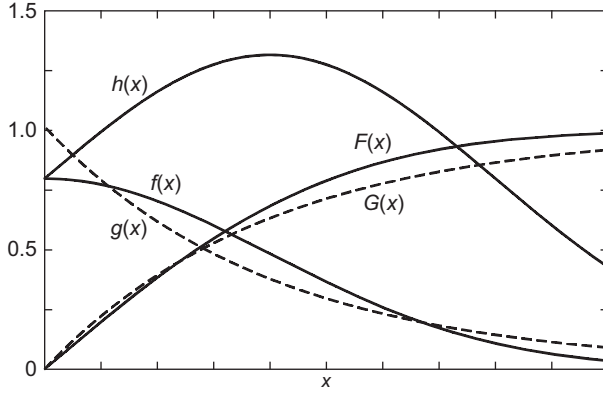


Figure 4.4 Graphs of the $f(x)$ (solid line), the more tractable PDF $g(x)$ (dashed line), the ratio $h(x) = f(x)/g(x)$, and the CDFs $F(x)$ and $G(x)$. The value of $h_{\max}$ is any finite bound of $h(x)$, but in practice should not greatly exceed the maximum value of $f(x)$.

4. If $\rho_2 > h(x_t)/h_{\max}$, reject the value x_t and return to step 1.²
5. If $\rho_2 < h(x_t)/h_{\max}$, accept x_t as a valid sample of x .

To verify that this procedure does produce values of x_t that are distributed according to $f(x_t) = h(x_t)g(x_t)$, first consider the probability that ρ_2 is less than $h(x)/h_{\max}$, namely,

$$P\{\rho_2 < h(x_t)/h_{\max}\} = \int_0^{h(x_t)/h_{\max}} d\rho_2 = h(x_t)/h_{\max}. \quad (4.11)$$

Let $u(x_t, \rho_2)$ be the joint probability distribution of the two random variables x_t and ρ_2 . The conditional probability of x_t given that $\rho_2 < h(x_t)/h_{\max}$ is, from Bayes theorem of Eq. (2.42) and from Eq. (4.11),

$$\begin{aligned} P\{x_t | \rho_2 < h(x_t)/h_{\max}\} &= \frac{u(x_t, \rho_2)P\{\rho_2 < h(x_t)/h_{\max}\}}{\int_a^b u(x_t, \rho_2)P\{\rho_2 < h(x_t)/h_{\max}\}dx_t} \\ &= \frac{u(x_t, \rho_2)[h(x_t)/h_{\max}]}{\int_a^b u(x_t, \rho_2)[h(x_t)/h_{\max}]dx_t}. \end{aligned} \quad (4.12)$$

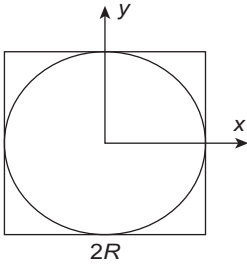
² The probability that ρ_2 exactly equals $h(x_t)/h_{\max}$ is ignored here and throughout this section. This probability is negligible, in principle; but in practical calculations using pseudorandom numbers, the situation may arise on very rare occasions. In such cases, it is arbitrary whether to group the “equal” case with the “less than” or “greater than” case.

But x_t and ρ_2 are independent random variables, so $u(x_t, \rho_2) = g(x_t)U(\rho_2) = g(x_t)$ because the density function of the unit distribution is simply unity. Substitution of this result into Eq. (4.12) gives

$$\begin{aligned} P\{x_t | \rho_2 < h(x_t)/h_{\max}\} &= \frac{g(x_t)[h(x_t)/h_{\max}]}{\int_a^b g(x_t)[h(x_t)/h_{\max}]dx_t} \\ &= \frac{g(x_t)h(x_t)}{\int_a^b f(x_t)dx_t} = g(x_t)h(x_t), \end{aligned} \quad (4.13)$$

the desired result.

Example 4.4 Sampling Uniformly Over a Disk



Consider the problem of sampling (x_i, y_j) points uniformly over the disk of radius R shown to the left. One method to generate such coordinate pairs is to uniformly sample from the bounding square containing the disk and to use the rejection technique to obtain samples inside the disk. Specifically, the algorithm is:

1. Select ρ_i and ρ_j from $U(0, 1)$.
2. Set $x_i = R(1 - 2\rho_i)$ and $y_j = R(1 - \rho_j)$.
3. Accept this pair if $x_i^2 + y_j^2 < R^2$.

Clearly, the sampling efficiency of this approach is the disk-area/square-area = $\pi/4$.

To obtain a better efficiency, the inverse CDF method could be used. The probability a coordinate pair is in dr at a distance r from the center is $f(r) = 2\pi r dr / \pi R^2 = 2r dr / R^2$. The associated CDF is $F(r) = r^2 / R^2$. Then the following procedure is used to generate (x_i, y_j) pairs:

1. Set $\rho_i = F(r_i) = r_i^2 / R^2$.
2. Solve for $r_i = R\sqrt{\rho_i}$.
3. Sample a polar angle from $U(0, 2\pi)$, i.e., choose $\phi_j = 2\pi\rho_j$.
4. Set $x_i = r_i \cos \phi_j$ and $y_j = r_i \sin \phi_j$.

Although this second approach has 100% efficiency, the calculation of sines and cosines is very computationally expensive compared to the arithmetic involved with the rejection method approach, and so the rejection method, although less efficient, may be preferred.

4.1.4 Composition Method

This method is also known as the alternative PDF or *probability mixing* method. It sometimes happens that the PDF for a process is cumbersome to use but that it can be

split into separate functions, each of which is easier to sample. Suppose, for example, that

$$f(x) \equiv A_1 f_1(x) + A_2 f_2(x), \quad a \leq x \leq b, \quad (4.14)$$

where the functions $f(x)$, $f_1(x)$, and $f_2(x)$ are all proper PDFs. Because $f(x)$ is a PDF, then the sum of the constants A_1 and A_2 must be unity.

A value of x may be sampled from $f(x)$ by the following procedure:

1. Select a random number ρ_i .
2. If $\rho_i < A_1$, use $f_1(x)$ as the PDF for selecting a value of x .
3. If $\rho_i > A_1$, use $f_2(x)$ as the PDF for selecting a value of x .

The extension to an arbitrary number of separate PDFs is straightforward. The constants A_i are just the probabilities that $f_i(x)$ is used to generate a random sample. This extension is considered in the next section.

The great advantage of this method is that sampling from a very complex distribution can be achieved by sampling from much simpler distributions using the inverse CDF or acceptance-rejection methods. Another advantage of this method is that it is often possible to give high sampling probabilities A_i to the PDFs that are inexpensively sampled and lower probabilities to those functions that are expensive to sample (see [Example 4.5](#)).

Example 4.5 Use of the Composition Method

Generate random variates from the PDF

$$f(x) = \frac{7}{8}[1 + (x-1)^6], \quad 0 \leq x \leq 1.$$

This PDF can be decomposed into

$$f(x) = \frac{7}{8} f_1(x) + \frac{1}{8} f_2(x),$$

where

$$f_1(x) = 1, \quad [\mathcal{U}(0, 1)] \quad f_2(x) = 7(x-1)^6, \quad 0 \leq x \leq 1.$$

Thus, with ρ_1 and ρ_2 from $\mathcal{U}(0, 1)$, the sampling scheme is

$$x = \begin{cases} \rho_2, & \text{if } \rho_1 \leq 7/8 \\ 1 + \sqrt[7]{\rho_2 - 1}, & \text{if } \rho_1 > 7/8 \end{cases}.$$

Notice that the difficult sampling scheme from $f_2(x)$ is used only one out of eight times, on the average.

4.1.5 Rectangle-Wedge-Tail Decomposition Method

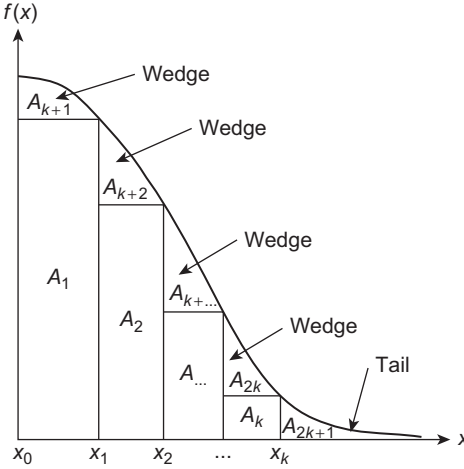


Figure 4.5 Decomposition of a PDF into a series of rectangles, wedges, and a tail.

and, as shown in the next section, there is a fast method for sampling from a nearly linear PDF. The tail region takes the longest time to sample, so the decomposition tries to minimize the probability of taking a sample from the tail, i.e., area $A_{2k+1} \ll 1$.

The PDF $f(x)$ is, thus, decomposed in a generalization of Eq. (4.14) as

$$f(x) = \sum_{i=1}^k A_i f_i(x) + \sum_{i=1}^k A_{k+i} f_{k+i}(x) + A_{2k+1} f_{2k+1}(x). \quad (4.15)$$

The probabilities of sampling from the various regions are given by

$$\begin{aligned} A_i &= f(x_i)(x_i - x_{i-1}), \quad i = 1, 2, \dots, k, \\ A_{k+i} &= \int_{x_{i-1}}^{x_i} f(x) dx - A_i, \quad i = 1, 2, \dots, k, \\ A_{2k+1} &= \int_{x_k}^{\infty} f(x) dx. \end{aligned} \quad (4.16)$$

Because the sum of all the A_i is the area under the PDF, $\sum_{i=1}^{2k+1} A_i = 1$, as required for the composite method. The corresponding PDFs for each region are

³ Here, it is assumed the PDF is on an infinite interval (x_0, ∞) . This method is easily extended to the interval $(-\infty, \infty)$ by including a second tail region.

One of the fastest methods for generating random variables from a continuous PDF is based on the rectangle-wedge-tail method [MacLaren et al. 1964]. This method, which is a special application of the composition method, decomposes the area under the PDF into a contiguous series of k rectangles, k wedges, and one tail region, as shown in the figure to the left.³ By using many narrow rectangles, most of the area under the PDF is contained in the rectangles, so that by using the composite method of the previous section, there is a high probability of sampling from a rectangular PDF, a very easy and fast procedure. Also, the PDFs for the wedges become almost linear,

$$\begin{aligned}
f_i(x) &= 1/(x_i - x_{i-1}), & x_{i-1} \leq x < x_i, & \quad i = 1, 2, \dots, k, \\
f_{k+i}(x) &= [f(x) - f(x_i)]/A_{k+i}, & x_{i-1} \leq x < x_i, & \quad i = 1, 2, \dots, k, \\
f_{2k+1}(x) &= f(x)/A_{2k+1}, & x \geq x_k. &
\end{aligned} \tag{4.17}$$

4.1.6 Sampling from a Nearly Linear PDF

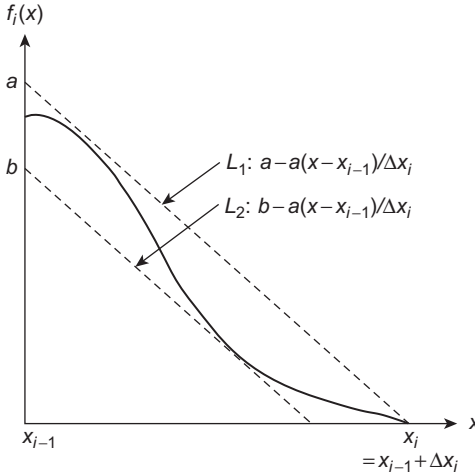


Figure 4.6 A nearly linear PDF.

By decomposing a PDF into many rectangles, wedges, and a tail, the sampling in a wedge region is often from a PDF $f_i(x)$, $x \in [x_{i-1}, x_i]$, that varies almost linearly from some value at one end of the interval to zero at the other end of the interval, as illustrated in Fig. 4.6. The interval width is denoted by $\Delta x_i \equiv x_i - x_{i-1}$. Also shown on this figure are two straight parallel lines that “squeeze” the PDF. Rather than use the L_1 as a limiting function for the rejection method (see Section 4.1.3), the two parallel lines are used to limit the sampling volume and give a much higher chance of success. Knuth [1998], gives the follow-

ing simple algorithm to sample from a nearly linear PDF such as that shown in Fig. 4.6.

1. Generate ρ_1 and ρ_2 from $\mathcal{U}(0, 1)$; set $u = \min(\rho_1, \rho_2)$, $v = \max(\rho_1, \rho_2)$, and $x = x_{i-1} + u\Delta x_i$.
2. If $v \leq b/a$, then return x as an acceptable sample.
3. Else, if $v \leq u + f_i(x)/a$, then return x as an acceptable sample.
4. Otherwise, no acceptable value is found and return to step 1 to try again.

4.1.7 Composition-Rejection Method

A generalization of the rejection method is the composition-rejection method of sampling. Suppose that the PDF $f(x)$ can be written as

$$f(x) = \sum_{j=1}^n A_j f_j(x) g_j(x), \tag{4.18}$$

in which all A_j are positive, all $g_j(x)$ are PDFs with analytical invertible CDFs, and all $f_j(x)$ have the property $0 \leq f_j(x) \leq 1$. A value of the random variate x may be selected as follows:

1. Select a random number and a corresponding value of i , where the probability of selecting i is proportional to A_i , i.e., with probability $A_i / \sum_{j=1}^n A_j$.

2. Select a test variate x_t from the PDF $g_i(x)$ using the inverse cumulative distribution method, i.e., $x_t = G_i^{-1}(\rho_i)$.
3. If $\rho_{i+1} \leq f_j(x_t)$, accept $x = x_t$; otherwise, reject x_t and return to step 1.

The efficiency of this composition-rejection method is $1/\sum_i A_i$.

4.1.8 Ratio of Uniforms Method

Suppose $h(z)$ is some integrable non-negative function on the interval $(-\infty, \infty)$ and one wants to sample from its PDF

$$f(z) = h(z) / \int_{-\infty}^{\infty} h(z) dz. \quad (4.19)$$

A rather remarkable result by [Kinderman and Monohan \[1977\]](#) often allows fast generation of samples from this distribution.

Consider the region V in the x - y plane defined by the values of x and y that satisfy

$$V: 0 \leq x^2 \leq h(y/x). \quad (4.20)$$

If random points x and y are uniformly sampled over V , then it can be shown [[Fishman 2003](#)] that the ratio $z \equiv y/x$ is a random variable of which the PDF is $f(z)$.

The region V may be complex; but if $h(x)$ and $x^2 h(x)$ are bounded in V , the region V can be embedded in a simple rectangle region R and the rejection technique used to sample from V . Such a rectangular region R is the region containing those x and y values that satisfy

$$R: 0 \leq x \leq b \quad \text{and} \quad c \leq y \leq d, \quad (4.21)$$

where

$$b = \max \sqrt{h(x)}, \quad c = \min x\sqrt{h(x)}, \quad \text{and} \quad d = \max x\sqrt{h(x)}. \quad (4.22)$$

The following algorithm implements this ratio of uniforms sampling scheme:

1. Generate ρ_1 and ρ_2 from $\mathcal{U}(0, 1)$.
2. Set $x = b\rho_1$, $y = c + (d - c)\rho_2$, and $z = y/x$.
3. If $x^2 \leq h(x)$, accept z as a sample; otherwise, go to step 1 and try again.

This very simple algorithm is easy to program and generally very fast, provided $h(x)$ is easy to evaluate.

4.1.9 Sampling from a Joint Distribution

To sample from a joint PDF (see Section 2.3), one usually samples first from the marginal PDF for a single variable, then works through the remaining variables by

sampling from the appropriate conditional PDFs. For example, the joint PDF $f(x, y, z)$ can be decomposed as

$$f(x, y, z) = f(x|y, z)f(y|z)f_z(z). \quad (4.23)$$

A sample of z is first picked from $f_z(z)$, then a sample of y is selected from the conditional PDF $f(y|z)$, and finally, a sample of x is obtained from the conditional PDF $f(x|y, z)$ given the other two values. In each case, standard sampling techniques are used. This sampling technique is illustrated in [Example 4.6](#). An alternative sampling technique for joint PDFs is Markov chain Monte Carlo, which is covered in Chapter 6.

4.1.10 Sampling from Specific Distributions

Over the last half century there have been literally many hundreds of papers on how to compute random values from specific distributions, each claiming some new wrinkle or advantage over previous methods. Many of these papers present little new, a consequence of the large diversity of disciplines that use Monte Carlo techniques, so that users in one area are unaware of techniques used in another. Likewise, there is little intellectual investment needed to make some subtle modification in one step of an existing algorithm and claim some improvement. Almost all algorithms depend on using random numbers uniformly distributed on $[0, 1]$ and some combination of the methods discussed earlier in this section. Both the algorithm efficiency and speed, which often depend on the computing platform, are important considerations. In some cases, the most efficient algorithm in one range of the distribution is not the most efficient in another range.

Example 4.6 Sampling from a Joint Distribution

Consider the joint PDF

$$f(\theta, \phi) = \frac{1}{K} \sin \theta, \quad \frac{\pi}{4} \leq \theta \leq \frac{\pi}{2}, \quad 0 \leq \phi \leq \theta,$$

where $K = 1 + \sqrt{2}(\pi - 4)/8 \simeq 0.8483$ to ensure proper normalization. PDFs of this general type arise naturally when sampling from a solid angle subtended by a surface at a point. A procedure to sample from this PDF is developed. Begin by constructing the marginal PDFs

$$\begin{aligned} f_\theta(\theta) &= \int_0^\theta f(\theta, \phi) d\phi = \frac{\theta \sin(\theta)}{K} \\ f_\phi(\phi) &= \frac{1}{K} \int_{\pi/4}^{\pi/2} \sin(\theta) d\theta = \frac{-\cos(\pi/2) + \cos(\pi/4)}{K} = \frac{1}{\sqrt{2}K} \simeq 0.8336. \end{aligned}$$

The conditional PDFs are

$$f(\theta|\phi) = \frac{f(\theta, \phi)}{f_\phi(\phi)} = \sqrt{2} \sin \theta \quad \text{and} \quad f(\phi|\theta) = \frac{f(\theta, \phi)}{f_\theta(\theta)} = \frac{1}{\theta}.$$

The joint PDF can then be sampled in two ways; both are considered here, although it is seen that one is preferable.

One way is to sample θ from $f_\theta(\theta)$ and then ϕ from $f(\phi|\theta)$. With Eq. (4.2) a random value θ_i is found from

$$\rho_i = \frac{1}{K} \int_{\pi/4}^{\theta_i} \theta \sin \theta \, d\theta = \frac{\sin \theta_i - \theta_i \cos \theta_i - \sqrt{2}(4 - \pi)8}{K}.$$

This is a transcendental equation that first must be solved for θ_i (e.g., by trial and error) before sampling for ϕ . Clearly, this approach is computationally expensive.

Alternatively, sample ϕ from $f_\phi(\phi)$ and then θ from $f(\theta|\phi)$. With Eq. (4.2) a value of ϕ_i is found from

$$\rho_{1i} = f_\phi(\phi_i) = \frac{1}{\sqrt{2}K} \int_0^{\phi_i} d\phi = \frac{\phi_i}{\sqrt{2}K} \simeq 0.8335\phi_i.$$

This result can be easily inverted to give $\phi_i = \rho_{1i}/0.8335$. Then a value θ_i is picked from the conditional PDF $f(\theta|\phi_i)$. Again, Eq. (4.2) yields

$$\rho_{2i} = \int_{\pi/4}^{\theta_i} f(\theta|\phi) \, d\theta = \sqrt{2} \int_{\pi/4}^{\theta_i} \sin \theta \, d\theta = 1 - \sqrt{2} \cos \theta_i,$$

of which the solution, after replacing $(1 - \rho_{2i})$ by ρ_{2i} , is found to be $\theta_i = \cos^{-1}(\rho_{2i}/\sqrt{2})$.

Finding the “best” algorithm for a given distribution is not easy. Simple algorithms can be programmed easily, but often the best algorithms require many lines of code and have multiple variate subranges, break points, empirical constants, and different sets of mixing functions. In Appendix A, some simple methods are provided for calculating random variates from commonly encountered distributions. The reader is referred to McGrath et al. [1975], Devroye [1986], Gentle [1998], and Fishman [2003] for more thorough treatments of optimum algorithms for sampling from particular distributions.

4.2 Scoring

In Monte Carlo calculations, the results, or *scores*, of a large number N of histories are used to obtain an estimate of some random variable $z(x)$, which may or may not be known explicitly a priori. By generating values x_i of the random variable x from its governing PDF $f(x)$, the estimate of $\langle z \rangle$ is found from

$$\langle z \rangle \equiv \int_a^b z(x)f(x)dx \simeq \bar{z} \equiv \frac{1}{N} \sum_{i=1}^N z(x_i). \quad (4.24)$$

Before any histories are simulated, various *accumulators* are set to zero. Then, after each history, the accumulators are incremented by the value produced by each history for the quantity being accumulated. For example, to find $\bar{z}$, an accumulator S_1 is used

to sum the values of $z(x_i)$ produced by each history, i.e., after each history the value of $z(x_i)$ is added to $\mathcal{S}_1$. After N histories, the estimate of $\langle z \rangle$ is given by $\bar{z} = \mathcal{S}_1/N$.

Besides computing the sample mean $\bar{z}$ from the accumulated value of $\mathcal{S}_1$, other statistics of the histories are also accumulated throughout the Monte Carlo calculation. For example, an accumulator $\mathcal{S}_2$ sums the $z^2(x_i)$ for each history so that $\bar{z}^2 \equiv (1/N) \sum_{i=1}^N z^2(x_i) = \mathcal{S}_2/N$. With this result and Eq. (2.56), an estimate of the standard deviation of the sample mean is obtained, namely,

$$s(\bar{z}) \simeq \frac{s(z)}{\sqrt{N}} = \frac{1}{\sqrt{N}} \sqrt{\frac{N}{N-1} (\bar{z}^2 - \bar{z}^2)} = \frac{1}{N-1} \sqrt{\bar{z}^2 - \bar{z}^2} \quad (4.25)$$

$$= \frac{1}{N\sqrt{N-1}} \sqrt{N\mathcal{S}_2 - \mathcal{S}_1^2}. \quad (4.26)$$

Because $s(\bar{z})$ is proportional to $1/\sqrt{N-1} \simeq 1/\sqrt{N}$ for large N , four times the number of histories must be used to halve the standard deviation on the estimate of $\bar{z}$. This is an inherent weakness of the Monte Carlo method for, in order to obtain an accurate value of $\bar{z}$, one must reduce its standard deviation to be acceptably small. Another approach to reduce $s(\bar{z}) = s(z)/\sqrt{N}$ is to make $s(z)$ smaller, thereby reducing the spread of the tally results. Such reduction of $s(z)$ can be accomplished by using various *variance reduction techniques* that modify the Monte Carlo simulation. Some of these techniques are discussed in Chapter 5.

4.2.1 Statistical Tests to Assess Results

To assess how good an estimate the mean $\bar{z}$ is of the expected value $\langle z \rangle$ several statistical measures should also be calculated to provide confidence in the result. For example, the standard deviation of the mean $s(\bar{z})$ is one such statistic. Ideally, a good result for $\bar{z}$ should have a small $s(\bar{z})$. However, how small is small? To provide confidence in the precision of a Monte Carlo result, several other statistics generated from the histories should be computed and examined carefully. The analysis of such statistical measures of a Monte Carlo calculation is very important to ensure that a precise estimate of the expected value has been obtained. Sometimes, through bad luck (i.e., an incomplete sampling of all possible histories), a very imprecise result is obtained. The statistical analysis of a Monte Carlo calculation is a very critical aspect of the analysis. Several important Monte Carlo statistics are discussed below.

Relative error and figure of merit

The *relative error* is defined as $R \equiv s(\bar{z})/\bar{z}$ and is a useful number because it gives a measure of the statistical precision of $\bar{z}$ as a fractional result with respect to the estimated mean. From Eqs. (4.24) and (4.25),

$$R = \frac{1}{\sqrt{N-1}} \frac{\sqrt{\bar{z}^2 - \bar{z}^2}}{\bar{z}} = \frac{1}{\sqrt{N-1}} \sqrt{\frac{N}{N-1} \frac{\mathcal{S}_2}{\mathcal{S}_1^2} - \frac{1}{N-1}}. \quad (4.27)$$

For large N , $\sqrt{N-1} \simeq \sqrt{N}$ and the above simplifies to

$$R = \sqrt{\frac{S_2}{S_1^2} - \frac{1}{N}}. \quad (4.28)$$

Some important observations. First, if all the $z_i \equiv z(x_i)$ are equal, then $R = 0$, i.e., low-variance solutions minimize the spread of the z_i . Second, if all the $z_i = 0$, R is defined as 0. If there is only one nonzero z_i , then $R \rightarrow 1$ as $N \rightarrow \infty$. Finally, if the z_i all have the same sign, then $\bar{z}^2 > \overline{z^2}$ so that $s(\bar{z}) < \bar{z}$ and $R < 1$. Thus, R , for z_i of the same sign, is always between zero and unity. For cases with z_i of both signs, this is not true. The smaller R , the better is the precision of $\bar{z}$. Typically, a good result should have $R \leq 0.05$ [MCNP 2003].

Another useful statistic for a Monte Carlo calculation is the *figure of merit* defined as $FOM \equiv 1/(R^2T)$, where T is the runtime needed to produce N histories. Because T is proportional to N and R^2 is proportional to $1/N$ (see Eq. (4.27)), the FOM should be approximately constant throughout a Monte Carlo run, save for statistical fluctuations early in the run. This statistic has several important uses. First, if the FOM is not roughly constant (except at the beginning of a run), the estimated confidence limits for $\bar{z}$ may not overlap the expected value $\langle z \rangle$. Second, short runs with different methods for estimating $\bar{z}$, for example, with different variance reduction techniques, allow a quick determination of which is the best method, namely, the one that produces, for the same computing time, the largest FOM (smallest R). Finally, the computing time needed to achieve a specified relative error can be estimated as $T \simeq 1/(R^2FOM)$.

Variance of the variance

Higher moments $z^n(x_i)$ are also often accumulated in order to find other statistical measures of precision of Monte Carlo results. One such statistic is the variance of the variance (VOV) [Estes and Cashwell 1978; Pederson 1991]. The relative VOV is like the relative error R of the mean, except that it is a measure of the variance of R . Specifically, it is defined as [Pederson 1991]

$$VOV = \frac{s^4(\bar{z})}{s^2(s^2(\bar{z}))} = \frac{\sum_{i=1}^N (z_i - \bar{z})^4}{\left[\sum_{i=1}^N (z_i - \bar{z})^2 \right]^2} - \frac{1}{N}. \quad (4.29)$$

To compute this statistic, four accumulators are needed: $S_m = \sum_{i=1}^N z_i^m$, $m = 1, 2, 3, 4$. Then expansion of Eq. (4.29) in terms of the sums of powers of z_i gives

$$\begin{aligned} VOV &= \frac{S_4 - 4S_1S_3/N + 6S_2S_1^2/N^2 - 3S_1^4/N^3}{[S_2 - S_1^2/N]^2} - \frac{1}{N} \\ &= \frac{S_4 - 4S_1S_3/N + 8S_2S_1^2/N^2 - 4S_1^4/N^3 - S_2^2/N}{[S_2 - S_1^2/N]^2}. \end{aligned} \quad (4.30)$$

From many statistical tests [MCNP 2003], it has been found that *VOV* should be less than 0.1 to increase the chance that reliable confidence intervals are obtained. Moreover, multiplication of the numerator and denominator of Eq. (4.30) by $1/N$ shows that the *VOV* should decrease monotonically as $1/N$.

Estimating the underlying PDF

In addition to calculating various statistics from the histories, the scores of each history are usually binned according to their values in order to estimate the (usually unknown) underlying PDF $f(z)$. From this estimate of $f(z)$ an important statistical test can be made about the reliability of the estimated PDF to quantify how well the N histories have sampled very rare but very important high-scoring events. The central limit theorem requires the first two moments of $f(z)$ to exist, so that if proper sampling of high-scoring events has occurred, the largest values of z_i should reach an upper bound (if one exists) or should decrease faster than $1/|z|^3$ so that $\langle z^2 \rangle = \int_{-\infty}^{\infty} z^2 f(z) dz$ exists. Thus, if a Monte Carlo analysis has properly sampled high- z histories, the high- z portion of the empirical distribution $f(z)$ should have a well-defined upper bound or decrease at least as rapidly as $1/z^3$. For many difficult simulations this requirement on $f(z)$ is the hardest to satisfy.

4.2.2 Scoring for “Successes-Over-Trials” Simulation

The simplest scoring is for problems in which each history results in a “success,” $z(x_i) = 1$, or a “failure,” $z(x_i) = 0$. In such a Monte Carlo problem the accumulator S_1 is the number of successes and the number of histories N is the number of trials. Thus, an estimate of the probability of success is S_1/N . Buffon’s needle problem, used in many examples in this book, is such a *success-over-trials* problem. Another example of a success-over-trials problem is a simulation of the popular game of craps (see Example 4.7).

Example 4.7 Craps as a Success-over-Trials Game

Consider the roll of a pair of dice in the game of craps. What is the probability of obtaining a winning “seven”? If $\mathbf{x} \equiv (J, K)$, $J, K = 1, \dots, 6$ are the number of dots on each die, then a winning seven is obtained from the six combinations (1,6), (2,5), (3,4), (4,3), (5,2), (6,1). Of the possible $6 \times 6 = 36$ possible outcomes, the probability of a seven is $P_7 = 6/36 \simeq 0.1666667$.

In a Monte Carlo analysis, the outcomes J and K are sampled from the discrete PDF $f_i = 1/6$, $i = 1, \dots, 6$. From Section 4.1.2 the sampled value of J is set to

$$J = n \quad \text{if} \quad (n-1)(1/6) \leq \rho_i < n(1/6), \quad n = 1, \dots, 6.$$

A similar procedure is used to select a value of K . Then, the *score* of a simulated roll (*history*) is the success of the roll, i.e.,

$$z(\mathbf{x}_i) = \begin{cases} 1 & \text{if } J + K = 7 \\ 0 & \text{otherwise} \end{cases}.$$

After N simulated dice rolls, form the composite score of successes $\mathcal{S}_1 = \sum_{i=1}^N z(\mathbf{x}_i)$. From Eq. (4.26), the estimate of the probability of rolling a seven is $\bar{P}_7 \simeq \mathcal{S}_1/N$ and the estimate of the standard deviation of this probability is, in the limit of large N ,

$$\sigma(\bar{z}) \simeq s(\bar{P}_7) \equiv \frac{1}{\sqrt{N}} \sqrt{\frac{S_2}{N} - \bar{P}_7^2}.$$

From such a Monte Carlo analysis, one can loosely say that the probability of rolling a seven, on any individual roll of the dice, is $\bar{P}_7 \pm s(\bar{P}_7)$, whereas, of course, what is really meant is that the estimate of the true probability of rolling a seven is $\bar{P}_7$ and that the true value has about a 68% probability of being within $\pm s(\bar{P}_7)$ of the estimated value. Results of such a simulation are shown in Table 4.1.

Table 4.1 Sample mean and standard deviation for throwing a seven using two dice, for various values of N .

N	$\bar{P}_7$	$\sigma(\bar{P}_7)$
10	0.2	0.1
100	0.21	0.04
1,000	0.156	0.011
10,000	0.1626	0.0037
100,000	0.1652	0.0012

4.2.3 Use of Weights in Scoring

In many problems most histories contribute little, if any, to $\bar{z}$. For example, in the Buffon's needle drop problem, if the separation distance D between lines is much greater than the needle length L , a needle of which either end is a distance more than L from the nearest line has no chance to fall on the line. To continue the history by calculating the coordinates of the other end and then to determine if this end is on the other side of a line (as was done in Example 1.3) is a waste of computer time. Similarly, if the integral $I = \int_a^b g(x)dx$ has an integrand with highly peaked but narrow features, then uniform sampling over the interval (a, b) produces many integrand evaluations that contribute little to the value of I .

The great power of Monte Carlo is that many methods have been developed over the past half-century to greatly enhance the number of histories that produce a high score and to minimize histories that contribute little to $\bar{z}$. These methods, therefore, *bias* the Monte Carlo calculation to produce histories that yield a large score or contribution to $\bar{z}$. Of course, the resulting value of $\bar{z}$ would be totally different from a value of $\bar{z}$ that would have been obtained without bias, and so would be a very bad estimate of $\langle z \rangle$. To correct the score or *tally* of such biased histories to give an unbiased value of $\bar{z}$, each history is assigned a *weight* W , which may change throughout the history to compensate for the biases introduced at various steps of the history.

For example, if, in estimating $\langle z \rangle = \int_a^b z(x)f(x)dx$ by Monte Carlo, some subinterval of $[a, b]$, in which $z(x)$ is highly peaked, were sampled ten times as frequently as

would be done if the x_i were truly sampled from $f(x)$, then the weight of such biased values of x_i should be $W_i = 1/10$. The weights for x_i outside this biased subinterval are $W_i = 1$. In this way, the score for the i th history is $\zeta_i = W_i z(x_i)$ and an unbiased tally is

$$\bar{z} = \frac{1}{N} \sum_{i=1}^N W_i z(x_i) = \frac{1}{N} \sum_{i=1}^N \zeta_i. \quad (4.31)$$

Similarly, in Buffon's needle problem with widely spaced lines, one could drop the end of needles only within a distance L of a line, but then reduce the weight of such needles to account for the fact that needle ends should be uniformly dropped over $0 < x < D$. This biased approach to Buffon's problem is illustrated in [Example 4.8](#).

4.2.4 Scoring for Multidimensional Integrals

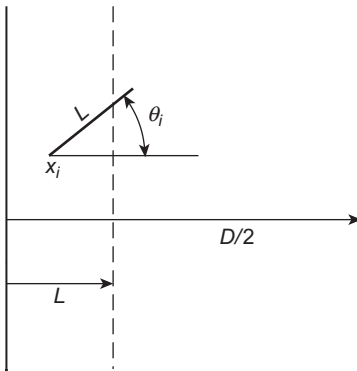
An interesting question arises in relation to multiple integrals. Consider a multidimensional integral of the form

$$\langle z \rangle = \int_V z(\mathbf{x}) f(\mathbf{x}) d\mathbf{x}, \quad (4.32)$$

where f is a joint PDF of, say, n stochastic variables $\mathbf{x} = \{x_j, j = 1, 2, \dots, n\}$. In principle, it is possible to construct a Monte Carlo estimate of the form

$$\bar{z}_a = \frac{1}{N_1} \sum_{i_1=1}^{N_1} \frac{1}{N_2} \sum_{i_2=1}^{N_2} \dots \frac{1}{N_n} \sum_{i_n=1}^{N_n} z(x_{1,i_1}, x_{2,i_2}, \dots, x_{n,i_n}), \quad (4.33)$$

Example 4.8 A Biased Buffon's Needle Simulation



In this example, a biased Buffon's needle simulation is considered in which the grid separation distance D is much greater than the length L of the needle. In an analog simulation, the needle end is dropped uniformly over the interval $(0, D)$. However, needle ends falling in $(L, D-L)$ cannot overlap a grid line, resulting in many wasted simulated needle drops. Problem symmetry also can be used to simplify the simulation (see figure to left). Clearly, the probability P_{cut} that a needle, of which one end lands in $(0, D/2)$, lands on a grid line is the same as that for a needle of which the end lands in $(D/2, D)$. Moreover, needle ends dropped in $(L, D/2)$ cannot cut the illustrated grid line at $x = 0$. Thus, in this biased simulation, needle ends are dropped uniformly over

the subinterval $(0, L)$. However, the probability an end would actually land in $x_i \in (0, L)$ in an analog simulation is $L/(D/2)$. So the weight of these dropped needles is set to $W_i = L/(D/2)$ to remove the bias in this sampling. In the next segment of a needle's analog history, an angle θ_i of the needle with respect to the x -axis is uniformly sampled over $(0, 2\pi)$. But, only a needle of which the θ_i angle is in the range $(\pi/2, 3\pi/2)$ can possibly intersect the grid line at $x = 0$. To eliminate needle drops that do not intersect the grid line, the selection of θ_i is biased to occur only in $(\pi/2, 3\pi/2)$, thereby sampling this range twice as often as would be done in an analog simulation. Accordingly, to keep the score unbiased, the weight of the needle W_i is halved. In the next segment of a needle's history, the x -coordinate of the needle's other end is calculated as $x_i^o = x_i + L \cos \theta_i$, and if $x_i^o < 0$, the simulated needle drop has crossed a grid line. Its contribution to the tally is recorded as W_i (here, $z_i = 1$ if the grid line is intersected and 0 if not). In Fig. 4.7, results for the analog and biased (nonanalog) simulations are shown. Clearly, biased simulations, in which the targeted outcome is frequent, require much less computational effort.

where $x_{j,i}$ is the i th sample of the j th component of $\mathbf{x}$, and these are sampled from the appropriate marginal and conditional PDFs, as discussed in Section 4.1.9. This prescription samples $\prod_{i=1}^n N_i$ histories, but only N_1 values of x_1 and for each of these N_2 values of x_2 , and so forth. A simpler and more efficient Monte Carlo estimator takes the form

$$\bar{z}_b = \frac{1}{N} \sum_{i=1}^N z(\mathbf{x}_i), \quad (4.34)$$

which involves N histories each using n samples to form each $\mathbf{x}_i$. This latter approach is more efficient because each component of $\mathbf{x}$ is sampled in each history, providing better coverage of the sample space.

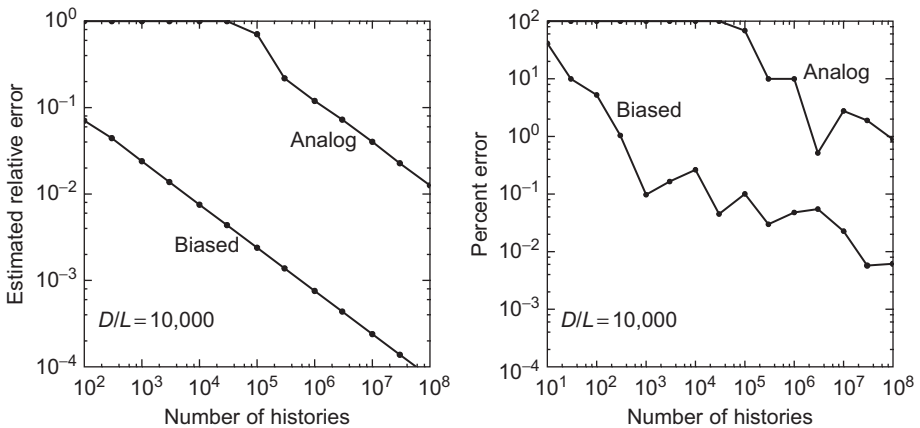


Figure 4.7 The left plot shows the estimated relative error $R = s(\bar{z})/\bar{z}$ for the biased and analog simulations. For $D/L = 10,000$ used in this example, Eq. (1.1) gives $P_{cut} = 2L/(\pi D) \simeq 6.36620 \times 10^{-5}$. The right figure shows the actual percent error for the two Monte Carlo estimates.

4.3 Accuracy and Precision

In common parlance the terms *precision* and *accuracy* are often used rather loosely. However, there is an important distinction between the two. In Monte Carlo, the *precision* of the calculation is limited by the uncertainty in $\bar{z}$ caused by the statistical fluctuations in the z_i and is measured by the standard deviation $\sigma(\bar{z})$. As more histories are run, $\sigma(\bar{z}) \simeq s/\sqrt{N}$ becomes smaller and a more precise estimate is obtained for $\bar{z}$. By contrast, the *accuracy* of $\langle z \rangle \simeq \bar{z}$ in Monte Carlo is how far $\bar{z}$ is from the physical quantity being estimated. The difference between the true physical value $\langle z \rangle$ and the estimate $\bar{z}$, in the limit as $N \rightarrow \infty$, is the *systematic error*, an error that is rarely known. It is quite possible to obtain a very precise Monte Carlo result (by running many histories) that is not very accurate, perhaps because approximations or errors were made in the models of the physical phenomenon being studied.

The same concepts apply to measurement theory and analysis of physical systems. If a given physical quantity is measured several times, by multiple people or even by the same person employing the same instrumentation, a range of values is usually obtained. For instance, a class of third graders each measuring the height of a fence with rulers are liable to get similar but different numbers. The precision of their measurements is measured by how far the individual measurements differ from the average. The accuracy of their measurement is a measure of how far their average is from the true height of the fence and is determined by how good their rulers are and how good their measurement technique is. Similarly, a good marksman is capable of both high accuracy (the average of the bullet pattern is close to the bull's eye) and high precision (the spread in the pattern is small). However, if someone misaligns the gun's sights without the marksman's knowledge (and the marksman does not check his result after each firing), the result may still be precise but is less accurate.

Because Monte Carlo is, in essence, an experiment with numbers used to simulate a real experiment, it is subject to the imprecision of actual experiments. Just as scientists trying to understand nature have used crude instruments and made grossly inaccurate measurements, which are usually improved as methods evolve, Monte Carlo using poor models can lead to inaccurate results. In these cases, it is customary to interpret Monte Carlo results by their precision, as measured by the sample standard deviation, and say these results are accurate to within the assumptions made in the simulation model. These concepts are explored in more detail as follows.

4.3.1 Factors Affecting Accuracy

The accuracy of a Monte Carlo analysis can be compromised by two main factors: (1) code errors and (2) user errors. The computer code may have large uncertainties or errors in the physical constants used in the physical models. Likewise, the physical models used in the analysis may be too simplistic or even wrong. A poor random number generator may also introduce errors. Finally, there may be coding errors (bugs) that, for instance, can occasionally produce an incorrect high score for an unusual history. These are factors that can only be remedied by benchmarking results against measurements, theory, or other independent calculations to find and identify any source of disagreement.

User errors

Another factor affecting accuracy is user error. For instance, a user may misuse a previously developed Monte Carlo code. An incomplete understanding of the typically many code features and options often allows users, particularly beginners, to misuse them, resulting in meaningless output. Likewise, simple mistakes in the input file, for example, a mistyped density or an incorrectly placed object in the geometry, can produce results that, while apparently reasonable, are extremely inaccurate. To avoid this source of inaccuracies, careful review of the input and output is required. For complex problems it is often useful to develop the complex model by first using simpler models upon which to build the complex model. At each stage in the model development, the changes in the output should be plausible from the changes made to the model.

Random number generators

Many early Monte Carlo calculations used pseudorandom number generators that, in hindsight, were far from ideal. The resulting inaccuracies introduced into the results by such generators are largely unknown. However, with modern random number generators the compromise of accuracy due to the inadequacies of the generators has largely been eliminated. But one potential danger still exists. The Achilles heel of all pseudorandom number generators is that they have a finite cycle length, i.e., the same sequence of random numbers eventually repeats. If a Monte Carlo calculation uses many more random numbers than the cycle length of the generator, then inaccuracies are introduced by using the same sequence of random numbers multiple times.

The repeated use of the same subsequence of random numbers can lead to *false convergence*. In Fig. 4.8, results of the Buffon's needle simulation used in Example 1.4 are shown for the case $D = 2L$. However, in this simulation a great many random numbers were discarded between needle drops so that after about 500 simulated needle drops, the cycle length of the random number generator was exceeded. Even though the estimate of P_{cut} appears to converge, it converges to a value about 5% below the true value of $P_{cut} \simeq 0.318310$. This false convergence is easily identified by the same repeated pattern (with decreasing amplitude) produced by each cycle of the random number generator. Fortunately, it is easy to avoid such false convergences by ensuring that the cycle length of your random number generator is never exceeded. A good rule of thumb is never use more than 10–20% of the random numbers in a cycle.

4.3.2 Factors Affecting Precision

The factors affecting the precision of a Monte Carlo result mostly arise from user-made choices for the calculation. Different ways of scoring a tally can affect the precision of the tally. Similarly, the use of variance reduction techniques (see Chapter 5) can profoundly affect the precision of a tally as does use of problem symmetry to simplify the Monte Carlo model. Finally, the precision of a tally is proportional to $1/\sqrt{N}$. Quadrupling the number of histories increases the precision by a factor of two. However, relying on N to increase precision is usually the last resort, because reducing $s(\bar{z})$, i.e., the spread among the scores, is generally preferred.

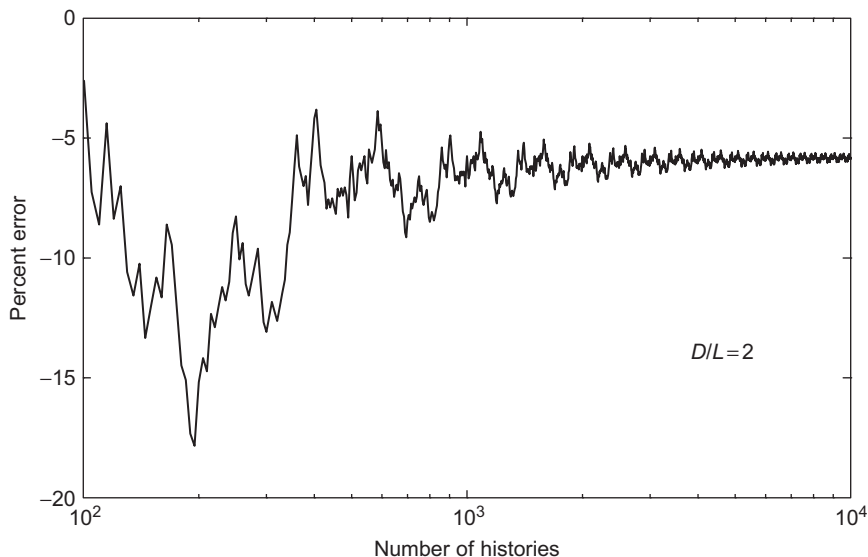


Figure 4.8 A false convergence in a simulation on Buffon's needle problem. Between each simulated needle drop using the method described in Example 1.4, a large number of random numbers were discarded so that the cycle length of the random number generator was exceeded.

Round-off errors

In principle, an arbitrary precision can be obtained in a score if a sufficient number of histories is run. However, it turns out that round-off errors associated with representing real numbers on digital computers, among other things, limits the precision. In a perfect world, all calculations would be perfect. However, digital computers use finite collections of bits to represent numbers. A number like $2/3$ must be represented in a digital computer by a finite number of digits, thus introducing a small round-off error. No matter how carefully a Monte Carlo simulation is conducted, it can never reconstruct the value $2/3$ exactly.

In fact, the actual situation is worse. In all digital computers, no matter what finite precision is used, there is always some small number ϵ such that $1 + \epsilon \neq 1$. In many Monte Carlo calculations, a history's score is added to a summation accumulator $\mathcal{S}$. Eventually, after a large number of histories, $\mathcal{S}$ becomes so large that adding a small score no longer changes the value of $\mathcal{S}$. As examples, consider the two summations $\mathcal{S}_1 = \sum_{i=1}^N (1/N)$ and $\mathcal{S}_2 = (1/N) \sum_{i=1}^N 2\rho_i$, where ρ_i is uniformly sampled on $(0, 1)$. Clearly, $\mathcal{S}_1 = 1$ no matter the value of N and, because $\langle \rho_i \rangle = 1/2$, $\mathcal{S}_2$ should converge to unity as $N \rightarrow \infty$. Results for these two summations are shown in Table 4.2 for single precision values of $\mathcal{S}_1$ and $\mathcal{S}_2$ on a 32-bit machine. In these examples, the effects of round-off errors become increasingly severe as N exceeds 10^7 .

Because every history score is limited by the computer's ability to represent real numbers, round-off errors accumulate. In many Monte Carlo calculations, in which

Table 4.2 Results of using single precision on a 32-bit machine to estimate $\mathcal{S}_1$ and $\mathcal{S}_2$, both expected values of which are unity. After 10^7 histories round-off error clearly becomes evident.

N	$\mathcal{S}_1 = \sum_{i=1}^N (1/N)$	$\mathcal{S}_2 = (1/N) \sum_{i=1}^N 2\rho_i$
10	1.0000000	1.1705571
30	1.0000000	0.9968364
1×10^2	1.0000000	0.9874166
3×10^2	1.0000000	1.0246525
1×10^3	1.0000000	0.9853277
3×10^3	1.0000000	1.0023125
1×10^4	1.0000000	0.9994981
3×10^4	1.0000000	1.0018269
1×10^5	1.0000000	0.9994974
3×10^5	1.0000000	1.0000912
1×10^6	1.0000000	1.0002490
3×10^6	1.0000000	0.9999786
1×10^7	1.0000000	1.0001853
3×10^7	0.5592405	0.9998386
1×10^8	0.1677722	0.3355443
3×10^8	0.0559241	0.1118481
1×10^9	0.0167772	0.0335544

summation accumulators are used, the estimate $\bar{z}$ may appear to be converging to $\langle z \rangle$ as N increases. But often, as N becomes very large, $s(\bar{z})$ stops decreasing and begins to increase! This is a sure sign that round-off errors have become important and that using more histories will not produce increasingly precise results. If such large numbers of histories are to be used, extended-precision arithmetic must be used for the summation accumulators, and even then there is a limit to the attainable precision. The use of double precision is always preferable to using single precision, unless it inordinately increases the computation time. Whether or not double precision increases the compute time depends on the architecture of the computer.

Another way to mitigate round-off errors, in situations in which tally scores can vary widely in magnitude, is to accumulate the largest scores in separate accumulation bins and to accumulate the sum of only small scores in a single bin. Then $\bar{z}$ is computed by adding all the scores, from smallest to largest, so that small scores are rounded correctly when the largest scores are added. The binning of the largest tallies has an additional benefit. From these binned tallies, a histogram is obtained of the high-end tail of underlying PDF $f(z)$. Several statistical tests can then be performed to ensure that the high-end tail has been sufficiently sampled.

Alternatively, the Monte Carlo calculation can be performed in M independent batches (each using a different substring of random numbers and a value of N that

avoids round-off errors). The values of all the $\bar{z}_i$, $i = 1, \dots, M$ are then averaged. This approach also has the advantage that $s(\bar{z})$ can be directly estimated from the variation among the $\bar{z}_i$.

In summary, whereas the strong law and the central limit theorem indicate that, in principle, a result can be obtained to any desired precision, in fact, the precision is limited by the number of bits used to represent numbers in the computer. For this reason, it is generally advisable to employ extended-precision arithmetic in Monte Carlo calculations on digital computers whenever feasible. Although Monte Carlo cannot produce results to arbitrary precision, the good news is that it can often produce results with several significant figures of precision with modest effort.

4.4 Summary

Monte Carlo simulations are performed by repeatedly sampling from probability distributions and scoring the results to form average quantities. The sampling process generally begins by generating pseudorandom numbers uniformly on the unit interval (i.e., between 0 and 1). A sample ρ_i from the unit interval is then transformed to a random sample over the range of interest and distributed according to the relevant probability distribution. When possible, this is usually achieved by what is called inverse CDF sampling, which involves inverting the appropriate CDF, namely,

$$x_i = G^{-1}(\rho_i). \quad (4.35)$$

When this inversion is difficult or impossible, other methods, such as rejection sampling or the composition method, must be used.

For a joint probability distribution, the PDF is decomposed into a product of marginal and conditional PDFs, e.g.,

$$f(x, y, z) = f(x|y, z)f(y|z)f_z(z). \quad (4.36)$$

First, the marginal PDF is sampled, in this case $f_z(z)$, and then the other variables are sampled from their conditional distributions, using the previously sampled values. For example, y from $f(y|z)$ given the sample for z and then sample x from $f(x|y, z)$ given the samples for y and z .

The scoring process is simply bookkeeping. To account for nonanalog or biased sampling, each sample is assigned a weight W_i that corrects for any bias. The quantity

$$\zeta_i = W_i z(x_i) \quad (4.37)$$

is scored and accumulators for ζ_i and ζ_i^2 are incremented. After all histories have been completed, the averages

$$\bar{z} = \frac{1}{N} \sum_{i=1}^N \zeta_i \quad \text{and} \quad \overline{z^2} = \frac{1}{N} \sum_{i=1}^N \zeta_i^2 \quad (4.38)$$

are calculated. The quantity $\bar{z}$ is the sample mean, the estimate of $\langle z \rangle$. The precision of this result is given by the standard deviation of the sample mean

$$s(\bar{z}) = \frac{s(z)}{\sqrt{N}} = \frac{1}{\sqrt{N-1}} \sqrt{z^2 - \bar{z}^2}. \quad (4.39)$$

Finally, other statistics, in addition to $s(\bar{z})$, are often used to give further assessment about the reliability of the estimate $\bar{z}$. Of course, most actual Monte Carlo simulations involve more complex integrands and PDFs of several random variables. But the general procedure just outlined is followed, with appropriate modifications to account for all of the variables.

The difference between accuracy and precision is discussed. Precision is a measure of deviation from the sample mean and is often quantified as a standard deviation. Accuracy is a measure of the (often unknown) difference between the true value or population mean and the sample mean. In Monte Carlo calculations, increasing the number of histories is an obvious way to improve the precision, up to a point. Because of round-off errors in digital computers, there is a limit to the precision that can be obtained in a Monte Carlo calculation.

Bibliography

- CARTER, L.L. AND E.D. CASHWELL 1975, *Particle-Transport Simulation with the Monte Carlo Method*, U.S. Energy Research and Development Administration, Available through National Technical Information Service, Springfield, VA.
- CHEN, H.C. AND Y. ASAU 1974, "On Generating Random Variates from an Empirical Distribution," *AIIE T.*, **6**, 163–166.
- DEVROYE, L. 1986, *Non-Uniform Random Variate Generation*, Springer-Verlag, New York.
- ESTES, G. AND E. CASHWELL 1978, *MCNP1B Variance Error Estimator*, TD-6-27-78, Los Alamos Nat. Lab., NM.
- FISHMAN, G.S. 2003, *Monte Carlo Concepts, Algorithms, and Applications*, Springer, New York (corrected printing).
- KAHN, H. 1956, *Applications of Monte Carlo*, AECU-3259, U.S. Atomic Energy Commission, available through Technical Information Service Extension, Oak Ridge, TN.
- KALOS, M.H., F.R. NAKACHE, AND J. CELNIK 1968, "Monte Carlo Methods in Reactor Computations," in *Computing Methods in Reactor Physics*, H. Greenspan, C.N. Kelber, and D. Okrent, eds., Gordon and Breach, New York.
- KINDERMAN, A.J. AND J.F. MONAHAN 1977, "Computer Generation of Random Variables Using the Ratio of Uniform Deviates," *ACM T Math. Software*, **3**, 257–260.
- KNUTH, 1998, *The Art of Computer Programming, Volume 2, Seminumerical Algorithms*, 3rd ed., Addison-Wesley Publishing Company, Reading, MA.
- MACLAREN, M.D., G. MARSAGLIA, AND T.A. BRAY 1964, "A Fast Procedure for Generating Exponential Random Variables," *Commun. ACM*, **7**, 298–300.
- MARSAGLIA, G. 1963, "Generating Discrete Random Variable in a Computer," *Commun. ACM*, **6**, 37–38.
- MCGRATH, E.J., D.C. IRVING, S.L. BASIN, R.W. BURTON, S.J. JAQUETTE, W.R. KELTER, S.C., AND C.A. SMITH 1975, *Techniques for Efficient Monte Carlo Simulation*, Vol. II: "Random

- Number Generation for Selected Probability Distributions,” ORNL-RSIC-38, Radiation Shielding Information Center, Oak Ridge Nat. Lab., Oak Ridge, TN.
- MCNP 2003, *A General Monte Carlo N-Particle Transport Code, Version 5*, Vol. I *Overview and Theory*, LA-UR-03-1987, Los Alamos Nat. Lab., Los Alamos, NM.
- NORMAN, J.E. AND L.E. CANNON 1972, “A Computer Program for the Generation of Random Variables from any Discrete Distribution,” *J. Stat. Comput. Sim.*, **1**, 331–348.
- PEDERSON, S.P. 1991, *Mean Estimation in Highly Skewed Samples*, Report LA-12114-MS, Los Alamos Nat. Lab., NM.
- VON NEUMANN, J. 1951, “Various Techniques Used in Connection with Random Digits,” *Bur. Stand. Math. Ser.*, No. 12.

Problems

1. In [Example 4.3](#), a tail outcome was selected if $x_i < 0.5$. What would be the consequences if you had selected $x_i \leq 0.5$ as the criterion for obtaining a tail? Explain.
2. Show that for the discrete distribution with M mass points, all with equal probability, the inverse CDF method to generating random variates x reduces to $x = \lceil \rho M \rceil$.⁴
3. In principle, one could solve [Eq. \(4.2\)](#) numerically for x_i for distributions $F(x)$ of which the inverse cannot be obtained analytically. Why is this not a practical method for generating random variates?
4. You wish to obtain a random sample η_i from a distribution of which the PDF is given by $g(y) = 2y$. What value of η_i corresponds to a pseudorandom number, $\rho_i = 0.25$, selected from the unit interval?
5. Consider a spherical shell with inner and outer radii a and b and the problem of how to find the radius r of points uniformly distributed within the shell. Show that the PDF is

$$f(r) = \frac{3r^2}{b^3 - a^3}, \quad a \leq r \leq b,$$

and the CDF

$$F(r) = \frac{r^3 - a^3}{b^3 - a^3},$$

and that the random radii are given by $r_i = [a^3 + \rho_i(b^3 - a^3)]^{1/3}$.

6. Devise a computer program to select radii for points uniformly distributed within a sphere of unit radius using the following two methods. (a) Show that the probability the radius is in dr about r is given by $p(r) dr = 3r^2 dr$ and then use the simple rejection technique with $M = 3$. (b) Embed the sphere in a cube with sides of length 2 and sample points uniformly in the cube and accept the point if it is inside the sphere. Which method is more efficient?
7. It is desired to generate the sine and cosine of a random angle ψ in the interval $(0, 2\pi)$ using a method attributed to [von Neumann \[1951\]](#). Show that this can be accomplished as follows: Select two random numbers, ρ_i and ρ_j , and set $\xi_i = 2\rho_i - 1$ and $\xi_j = 2\rho_j - 1$. Then, in those cases for which $\xi_i^2 + \xi_j^2 < 1$,

$$\cos \psi = \frac{x_i^2 - x_j^2}{x_i^2 + x_j^2} \quad \text{and} \quad \sin \psi = \frac{2x_i x_j}{x_i^2 + x_j^2}.$$

Show that the efficiency of this rejection method is given by $\pi/4$.

⁴ The ceiling function $\lceil y \rceil$ gives the smallest integer $\geq y$.

8. Verify by calculation the following rejection technique for sampling values of x distributed according to the Gaussian PDF [Kahn 1956]:

$$f(x) = \sqrt{2/\pi} \exp[-x^2/2], \quad \text{for all } x \geq 0.$$

- (a) Select random numbers ρ_1 and ρ_2 and let $x = -\ln \rho_1$ and $y = -\ln \rho_2$. (b) Accept x if $0.5(x-1)^2 \leq y$; otherwise, return to step a. (c) Select random numbers ρ_3 and assign to x the negative sign if $\rho_3 \leq 0.5$. Verify that the efficiency is given by $\sqrt{\pi/2} \exp[-1/2] = 0.76$.
9. Verify, by calculation, the following technique for sampling energies of prompt fission neutrons according to the PDF [Carter and Cashwell 1975]:

$$f(E) = \sqrt{4E/\pi T^3} \exp[-E/T],$$

where E is the neutron energy and T is a characteristic energy. Select three random numbers ρ_1 through ρ_3 and then set

$$E = T \left[-\ln \rho_1 - (\ln \rho_2) \cos^2(\pi \rho_3/2) \right].$$

10. Apply the rectangle-wedge-tail method to the problem of sampling from a unit exponential distribution $f(x) = e^{-x}$, $x \geq 0$. Use an equimesh such that $x_i - x_{i-1} = \delta$, so that $x_i = i\delta$, $i = 1, 2, \dots, k$. Show that

$$A_i = \delta e^{-i\delta}, \quad i = 1, 2, \dots, k,$$

$$A_{k+i} = e^{-i\delta} [e^\delta - 1 - \delta], \quad i = 1, 2, \dots, k,$$

$$A_{2k+1} = e^{-k\delta},$$

and that corresponding PDFs for each region are

$$f_i(x) = 1/\delta, \quad x_{i-1} \leq x < x_i, \quad i = 1, 2, \dots, k,$$

$$f_{k+i}(x) = \frac{e^{-x+i\delta} - 1}{[e^\delta - 1 - \delta]}, \quad x_{i-1} \leq x < x_i, \quad i = 1, 2, \dots, k,$$

$$f_{2k+1}(x) = e^{-x+k\delta}, \quad x \geq x_k.$$

Write a program to implement sampling using this rectangle-wedge-tail method, and compare the speed of this sampling scheme to that based on the inverse CDF method (see Example 4.2).

11. Verify by calculation the following rejection technique for sampling energies of prompt fission neutrons according to the Cranberg formula, with PDF

$$f(E) = 0.453 \exp[-E/0.965] \sinh \sqrt{2.29E},$$

where E is the neutron energy in MeV. (a) Select random numbers ρ_1 and ρ_2 and let $x = -\ln \rho_1$ and $y = -\ln \rho_2$. (b) If $-4.57257x + [y - 1.06918(x+1)]^2 \leq 0$, set $E = 1.99657x$ MeV; otherwise, return to step a. According to Kalos et al. [1968], the efficiency of this method is 0.74.

5 Variance Reduction Techniques

*There once was a student who'd fret,
if ever an answer she'd get.
The events were so rare,
few histories got there;
but she hadn't tried Russian roulette.*

The law of large numbers and the central limit theorem provide a powerful quadrature prescription for calculating expected values, namely, one samples from the governing probability distributions and forms appropriate averages to estimate the expected values. In the last chapter, methods were discussed for performing sampling from various distributions.

So what's left? Is that all there is to Monte Carlo calculations? The answer is a resounding NO. One of the features of Monte Carlo is that it is so flexible and adaptable that it can be used to solve a very complex problem for which one can't even begin to formulate the underlying distribution of which the expected value is sought. However, the real power of Monte Carlo is that the sampling procedure can be intentionally biased toward the region where the integrand is large or to produce simulated histories that have a better chance of creating a rare event, such as Buffon's needle falling on widely spaced lines. Of course, with such biasing, the scoring then must be corrected by assigning weights to each history in order to produce a corrected, *unbiased* estimate of the expected value. In such *nonanalog* Monte Carlo analyses, the sample variance $s^2(\bar{z})$ of the estimated expectation value $\bar{z}$ is reduced compared to that obtained by an unbiased or *analog* analysis. In Example 4.8, biasing a simulation of Buffon's problem with widely spaced lines to force all dropped needles to have one end within a needle's length of a grid line was seen to reduce the relative error by two orders of magnitude over that of a purely analog simulation.

A Monte Carlo calculation seeks an estimator $\bar{z}$ to some expected value $\langle z \rangle$. The goal of variance reduction methods is to produce a more precise estimate of some expected value than could be obtained in a purely analog calculation using the same computational effort. For "successes over trials" problems in which a success is very unlikely (e.g., flipping ten heads in a row), variance reduction tries to bias the calculation so that more successes are obtained. For problems in which a continuum of outcomes is possible, variance reduction strives to reduce the spread among the history results and bring them closer to the mean.

In this chapter, several methods are considered to reduce the variance of the estimator $\bar{z}$ for

$$\langle z \rangle \equiv \int_V z(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \simeq \bar{z} = \frac{1}{N} \sum_{i=1}^N z(\mathbf{x}_i), \quad (5.1)$$

where the $\mathbf{x}_i$ are sampled from the joint PDF $f(\mathbf{x})$. The standard deviation of the sample mean is given by Eq. (2.54), namely,

$$s(\bar{z}) = \frac{1}{\sqrt{N-1}} \sqrt{\overline{z^2} - \bar{z}^2}. \quad (5.2)$$

Because both $\overline{z^2}$ and $\bar{z}^2$ must always be positive, the sample variance s^2 can be reduced by reducing their difference. Thus, the various variance reduction techniques discussed in this chapter are directed, ultimately, to minimizing the quantity $\overline{z^2} - \bar{z}^2$. Note that, in principle, it is possible to attain zero variance, if $\overline{z^2} = \bar{z}^2$, which occurs if every history yields the sample mean. This is not, unfortunately, very likely. However, it is possible to reduce substantially the variance among histories by introducing various biases into a Monte Carlo calculation.

Over the years, many clever variance reduction techniques have been developed for performing biased Monte Carlo calculations. The introduction of variance reduction methods into Monte Carlo calculations can make otherwise impossible Monte Carlo problems solvable. However, use of these variance reduction techniques requires skill and experience. Nonanalog Monte Carlo, despite having a rigorous statistical basis, is, in many ways, an “art” form and cannot be used blindly.

Before looking at specific techniques for reducing the variance of an estimator, several approaches are summarized. These range from very basic to very esoteric.

Transform the Problem: In some calculations, problem symmetry or other features of the problem can be used to create an equivalent problem that has the same expected value but that can be treated by Monte Carlo much more efficiently. Often such a transformation produces far better variance reduction than any other method. However, it takes considerable understanding of the fundamental processes governing the problem to find such a transformation. A good example for a radiation shielding problem is given in the next section. Unfortunately, transformations that increase the calculational efficiency are not readily evident for many problems.

Truncation Methods: This is the simplest of all variance reduction techniques. In this approach, the underlying physical models from which random samples are taken are simplified so that each history or sample takes, on average, less computer time. Thus,

for a given computational time more histories or samples can be taken, and because the $s(\bar{z})$ varies as $1/\sqrt{N}$,¹ the variance of the estimator is reduced. For example, some detail in the geometry far from the scoring region generally has little effect on $\langle z \rangle$ and so needs not be modeled. Equivalently, a detailed model of the low probability tail of the PDF sampling distribution is not needed if $z(\mathbf{x})$ is small in the tail region.

Model simplification is a “brute force” approach, because quadrupling the number of histories only halves the standard deviation. Moreover, physical insight into a particular problem is required to introduce simplifications that reduce the calculational complexity for parts of sampling space that have little effect on the expected value. Also, the only way to verify that some simplification has negligible effect on the estimate of $\langle z \rangle$ is to perform the calculation with and without the simplification.

Modified Sampling Methods: By far, most variance reduction techniques modify the underlying sampling PDFs to bias the sampling. Such biasing is used to increase the likelihood that a history produces a nonzero score, i.e., a nonzero $z(\mathbf{x}_i)$. Biased sampling can still produce an unbiased score provided each history is assigned a “weight” that is adjusted to compensate for any biases introduced at the various steps in the history. By recording a history’s weight in the tally, an unbiased value of $\bar{z}$ is achieved.

History Control Methods: Two very powerful variance reduction techniques are to use *Russian roulette* and/or *splitting* to alter, on-the-fly, the number of histories that can potentially produce a nonzero tally. As a history moves through phase space, it is possibly killed if it enters a subregion from which it is unlikely to produce a nonzero score, or it is possibly split into multiple histories if a history leaving this subregion is likely to yield a nonzero score. Similarly, in biased sampling, Russian roulette and splitting can be used to eliminate histories that have too small a weight or split if the history obtains too large a weight. This technique is often used in combinations with other variance reduction methods.

Modify the Tally: In some problems, the function $z(\mathbf{x})$ can be modified to allow more nonzero values of $z(\mathbf{x}_i)$ without significantly changing $\bar{z}$. For example, several different $z(\mathbf{x})$ may have nearly the same expected value, so the average of all such $z(\mathbf{x})$ should be used to increase the likelihood that a history produces a nonzero score. Problem symmetry often suggests ways to modify the tally so that more histories give nonzero values.

Incorporate Partially Deterministic Methods: These sophisticated techniques circumvent normal random sampling at certain stages of a history by substituting deterministic or analytical results. For example, if a history reaches a certain subregion of the problem phase space, an analytical estimate can be made of the eventual value of the tally. These techniques include next-event estimators or they control the random number sequence.

Several of the fundamental methods for variance reduction are reviewed and several examples are given in the following sections.

¹ From Eq. (4.25) it is seen that $s(\bar{z})$ varies as $1/\sqrt{N-1}$, but from here on it is assumed that N is large, and the distinction between N and $N-1$ can be ignored.

5.1 Use of Transformations

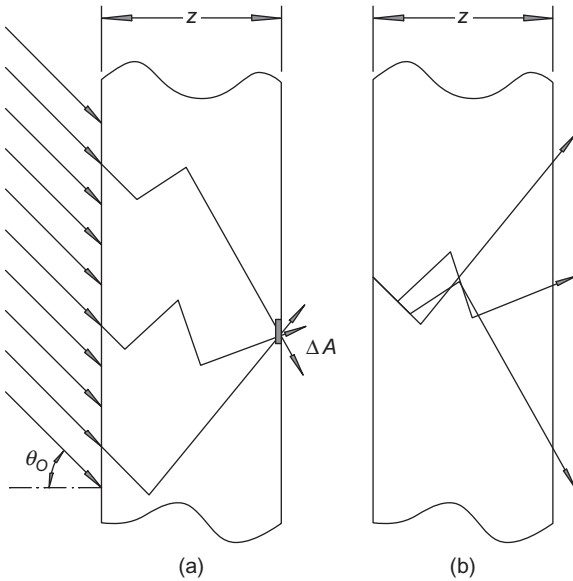


Figure 5.1 Equivalent problems. After Shultis and Faw [2000].

Although not a traditional variance reduction technique, use of problem symmetries or transformations can often be very effective in reducing the variance of some estimator. Consider, for example, the problem shown in Fig. 5.1(a). Here, monodirectional radiation particles are incident uniformly over the left surface of a radiation shield. The Monte Carlo simulation problem is to estimate the number of particles transmitted through a small unit area ΔA on the right surface. If this problem were simulated by sampling incident particles uniformly over the whole left

surface, hardly any particles would ever reach the scoring area ΔA .

A much more efficient, but equivalent, simulation can be realized if it is recognized that the probability a particle eventually reaches the right surface is independent of where on the left surface the particle is incident. Thus, a far superior simulation is to start all incident particles at the same point on the left surface and then use the entire right surface as the scoring surface, as shown in Fig. 5.1(b). The ratio of the number of transmitted particles to the number of incident particles in Fig. 5.1(b) is the same as the ratio of the number of transmitted particles, per unit surface area, to the number of incident particles, per unit surface area, in Fig. 5.1(a).

Similarly, when using Monte Carlo to evaluate an integral $\int_a^b f(x)dx$ of which the integrand $f(x)$ varies widely over the range of integration, the use of a change of variables can often produce an equivalent integral $\int_c^d g(s)ds$ with an integrand that has much less variation. This technique is closely related to importance sampling, discussed in section 5.2. Example 5.1 provides an illustration of the effectiveness of using an integral transformation.

Example 5.1 Transforming an Integral

Consider the evaluation of the integral

$$I = \int_0^1 f(x)e^{\alpha x} dx, \quad \alpha > 0,$$

where $f(x)$ is some slowly varying function. As α increases, the integrand becomes more and more peaked near the upper limit, and sampling uniformly over $[0, 1]$ produces many values of the integrand that contribute little to the value of $\bar{I}$. To avoid this, use the change of variables $s = e^{\alpha x}/\alpha$. Then $ds = e^{\alpha x} dx$ and $x = (1/\alpha) \ln \alpha s$. The transformed integral is thus

$$I = \int_{1/\alpha}^{e^\alpha/\alpha} f\left(\frac{1}{\alpha} \ln \alpha s\right) ds.$$

For the case $f(x) = x^2$ and $\alpha = 20$, both integrands, normalized by dividing by their maximum value at the upper limit, are shown in Fig. 5.2. The transformed integrand is seen to be much more uniform over the integration range and, thus, is more amenable to efficient evaluation by Monte Carlo. For this example, the first form of the integral can readily be evaluated analytically to give $I = e^\alpha [(1/\alpha) - (2/\alpha^2) + (2/\alpha^3)] - (2/\alpha^3) \simeq 2.195372 \times 10^7$ for $\alpha = 20$. The percentage error as a function of the number of histories is shown in Fig. 5.3 for both the analog case (sampling uniformly over the interval $[0, 1]$) and the transformed case. It is rare in Monte Carlo calculations to know the exact or expected value, but in this example it is known. Unlike relative errors, which decrease monotonically, actual errors exhibit a stochastic decrease with the number of histories.

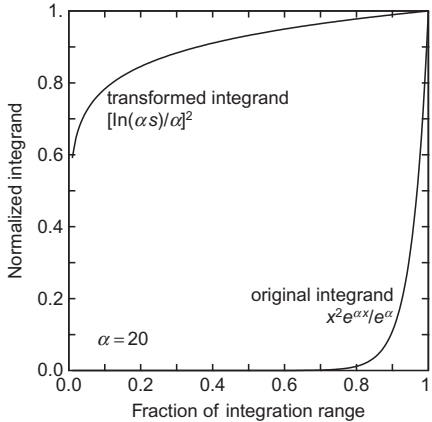


Figure 5.2 The normalized original and transformed integrands. The integrands are divided by their maximum value.

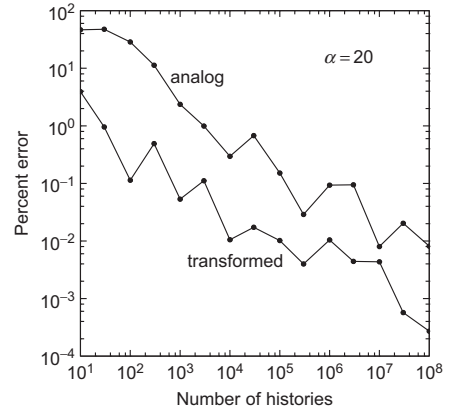


Figure 5.3 Percentage of absolute error as a function of the number of Monte Carlo histories used.

5.2 Importance Sampling

Suppose the expected value of a multidimensional function $z(\mathbf{x})$ is sought, i.e.,

$$\langle z \rangle = \int_V z(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \simeq \bar{z} = \frac{1}{N} \sum_{i=1}^N z(\mathbf{x}_i), \quad (5.3)$$

where the $\mathbf{x}_i$ are sampled from the PDF $f(\mathbf{x})$. Often, however, the function $z(\mathbf{x})$ is extremely small over most of V and almost all of the value of $\langle z \rangle$ comes from a small subregion of V where the PDF samples $\mathbf{x}$ infrequently. Many samples contribute little and, thus, represent a waste of computer time. It seems reasonable that better results could be obtained, with the same number of histories, if the sampling were preferentially made in regions where the integrand $|z(\mathbf{x})f(\mathbf{x})|$ is large in magnitude. This idea was first proposed by Marshall [1956]. To effect such a biased sampling, the integral can be transformed by introducing an arbitrary PDF $f^*(\mathbf{x})$ as

$$\langle z \rangle = \int_V \frac{z(\mathbf{x})f(\mathbf{x})}{f^*(\mathbf{x})} f^*(\mathbf{x}) d\mathbf{x} \equiv \int_V z^*(\mathbf{x}) f^*(\mathbf{x}) d\mathbf{x} = \langle z^* \rangle, \quad (5.4)$$

where $z^*(\mathbf{x}) = z(\mathbf{x})W(\mathbf{x})$ and the *weight* or *importance* function is $W(\mathbf{x}) \equiv f(\mathbf{x})/f^*(\mathbf{x})$. Thus, z and z^* have the same expectation, i.e.,

$$\langle z \rangle = \langle z^* \rangle \simeq \overline{z^*} = \frac{1}{N} \sum_{i=1}^N z(\mathbf{x}_i) W(\mathbf{x}_i), \quad (5.5)$$

where the $\mathbf{x}_i$ are sampled from $f^*(\mathbf{x})$.

Now consider the variance of the two expected values $\langle z \rangle$ and $\langle z^* \rangle$, namely,

$$\sigma^2(z) = \langle z^2 \rangle - \langle z \rangle^2 \quad \text{and} \quad \sigma^2(z^*) = \langle z^{*2} \rangle - \langle z^* \rangle^2 = \langle z^{*2} \rangle - \langle z \rangle^2, \quad (5.6)$$

where

$$\begin{aligned} \langle z^{*2} \rangle &= \int_V z^{*2}(\mathbf{x}) f^*(\mathbf{x}) d\mathbf{x} = \int_V z^2(\mathbf{x}) W^2(\mathbf{x}) f^*(\mathbf{x}) d\mathbf{x} \\ &= \int_V z^2(\mathbf{x}) W(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \neq \int_V z^2(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} = \langle z^2 \rangle. \end{aligned} \quad (5.7)$$

Hence, the two variances are, in general, different. If the variance for importance sampling, $\sigma^2(z^*)$, is smaller than that for straightforward sampling, $\sigma^2(z)$, then importance sampling gives a better estimate of the sample mean for the same number of histories.

When does this improvement occur? If $f^*(\mathbf{x})$ is chosen such that the weight function $W(\mathbf{x}) < 1$ over regions in which $z(\mathbf{x})$ makes a large contribution to the expected value, then from Eq. (5.7),

$$\langle z^{*2} \rangle = \int_V z^2(\mathbf{x}) W(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} < \int_V z^2(\mathbf{x}) f(\mathbf{x}) d\mathbf{x}. \quad (5.8)$$

For this choice of $W(\mathbf{x})$ it follows that $\sigma^2(z^*) < \sigma^2(z)$. In fact, if $f^*(\mathbf{x})$ is chosen as

$$f^*(\mathbf{x}) = f(\mathbf{x})z(\mathbf{x})/\langle z \rangle, \quad (5.9)$$

then $W(\mathbf{x}) = f(\mathbf{x})/f^*(\mathbf{x}) = \langle z \rangle / z(\mathbf{x})$, and Eq. (5.8) reduces to

$$\langle z^{*2} \rangle = \int_V z^2(\mathbf{x}) W(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} = \langle z \rangle \int_V z(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} = \langle z \rangle^2, \quad (5.10)$$

which, from Eq. (5.6), yields $\sigma(z^*) = 0$!

This result indicates that, if the expected value is known, a Monte Carlo procedure can be developed that has zero variance. Of course, $\langle z \rangle$ is not known or there would be no need to perform a Monte Carlo calculation. But if a PDF f^* can be formed that behaves more like the product $z(x)f(x)$ than does $f(x)$, $\langle z \rangle$ can be estimated with a smaller sample variance.

Thus, it is seen that importance sampling can reduce the sample variance. In practice, this can be achieved by using knowledge of the underlying processes to choose an appropriate f^* that favors histories that lead to successes (or to results close to the mean). Although sampling from f^* introduces a bias, this is accounted for by the weight factor $W = f/f^*$ in Eq. (5.5), so that the estimate is an unbiased estimator for $\langle z \rangle$.

Example 5.2 Buffon's Needle Problem with Importance Sampling

As an application of importance sampling with a discontinuous PDF, consider Buffon's needle problem in which the distance D between the grid lines is much larger than the length L of the needles. The probability that a randomly dropped needle lands on a grid line is given by Eq. (1.6), namely,

$$P_{cut} = \langle z \rangle = \int_0^D z(x) f(x) dx = \int_0^D \left(\frac{s_1(x) + s_2(x)}{2\pi L} \right) \left(\frac{1}{D} \right) dx,$$

where the functions $s_1(x)$ and $s_2(x)$ are given by Eqs. (1.4) and (1.5). This result is of the form of Eq. (1.8) and thus could be evaluated by the Monte Carlo quadrature method described in Section 1.3 in which random values of x_i that are uniformly distributed on the interval $[0, D]$ are used to calculate

$$P_{cut} = \langle z \rangle \simeq \bar{z} = \frac{1}{N} \sum_{i=1}^N z(x_i),$$

as was done in Example 1.2.

However, when $D \gg L$ this is very inefficient because most x_i will fall in the interval $[L, D - L]$, for which $z(x_i) = 0$ and, thus, make no contribution to the score $\bar{z}$. Rather than sample from the uniform PDF $f(x) = 1/D$, $x \in [0, D]$, it is better to sample from the discontinuous PDF

$$f^*(x) = \begin{cases} \frac{1}{2L}, & 0 < x < L \quad \text{or} \quad D - L < x < D, \\ 0, & \text{otherwise} \end{cases}.$$

Then P_{cut} is given by

$$P_{cut} = \int_0^D \frac{z(x)f(x)}{f^*(x)} f^*(x) dx = \int_0^D z(x)W(x)f^*(x) dx, \simeq \frac{1}{N} \sum_{i=1}^N z(x_i)W(x_i),$$

where the weight function is

$$W(x) = \frac{f(x)}{f^*(x)} = \begin{cases} \frac{2L}{D}, & 0 < x < L \quad \text{or} \quad D-L < x < D, \\ 0, & \text{otherwise} \end{cases}$$

and the x_i are sampled from $f^*(x)$. Results are shown in Fig. 5.4 with and without importance sampling for the case $D/L = 10,000$. For this case, Eq. (1.1) gives the exact value of $P_{cut}^{exact} = 2/(10^4\pi) = 6.36620 \dots \times 10^{-5}$.

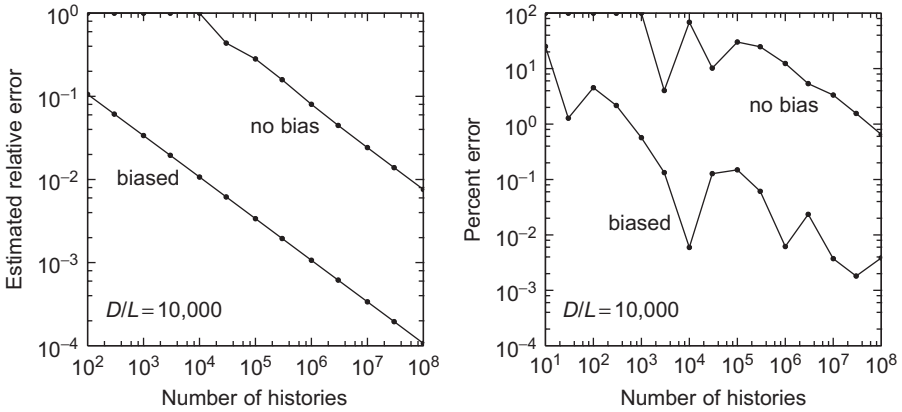


Figure 5.4 Results with and without importance sampling for the Buffon's needle integral of Eq. (5.1) discussed in Example 5.1. The left figure shows the relative error $R = s(\bar{z})/\bar{z}$, and the right figure gives the percentage error $100|\bar{z} - P_{cut}^{exact}|/P_{cut}^{exact}$.

Example 5.3 Importance Sampling for an Important Subinterval

Consider an example of the use of an important subinterval. Let x be distributed according to the PDF $f(x)$, which is defined on the interval $[a, b]$. Suppose that it is desired to preferentially select from the "important" subinterval $[c, d]$ of $[a, b]$. This can be accomplished in the following way: Form the function

$$f^*(x) = \begin{cases} \alpha f(x), & x \in [c, d] \\ \beta f(x), & x \in [a, b], x \notin [c, d] \end{cases}$$

In order to favor the preferred subinterval, choose a value for β from the interval $[0, 1]$. To make f^* a PDF, choose α to satisfy the condition $\int_a^b f^*(x) dx = 1$. Thus,

$$\alpha = \frac{1 - \beta[1 + F(c) - f(d)]}{F(d) - F(c)}.$$

Select x from $f^*(x)$ and form the weight factor

$$W(x) = \begin{cases} \frac{f(x)}{f^*(x)} = \frac{1}{\alpha}, & x \in [c, d] \\ \frac{1}{\beta}, & x \in [a, b] \quad x \notin [c, d] \end{cases}.$$

This procedure can be quite useful, although it involves a discontinuous PDF.

5.2.1 Application to Monte Carlo Integration

The general approach of using importance sampling to reduce the standard deviation of the estimated mean value can be applied to the Monte Carlo evaluation of a multi-dimensional integral

$$I \equiv \int_V g(\mathbf{x}) d\mathbf{x}. \quad (5.11)$$

Example 5.1, which showed how transforming an integral to produce a more uniform integrand, is an example of importance sampling. In this example, the term $f(x)dx = e^{\alpha x}dx$, which caused the integrand to vary rapidly, was identified as an exact differential ds that was integrable analytically and of which the integral could be inverted. Often, it is difficult to find some function that reproduces the large variation in an integrand that satisfies these two properties. However, importance sampling can still be used.

Consider the problem of evaluating, by *straightforward* Monte Carlo, the general integral of Eq. (5.11) by uniformly sampling over V , namely,

$$\langle g \rangle \pm \sigma(g) \equiv \int_V g(\mathbf{x}) d\mathbf{x} \pm \sqrt{\int_V [g - \langle g \rangle]^2 d\mathbf{x}} \simeq V\bar{g} \pm V\sqrt{\frac{\overline{g^2} - \bar{g}^2}{N}}, \quad (5.12)$$

where the $\pm$ term gives the 1-sigma standard deviation. Rather than sample the $\mathbf{x}_i$ uniformly in V , one could use a biased sampling from a PDF $f^*(\mathbf{x})$. Thus, the integral becomes

$$I = \int_V g(\mathbf{x}) d\mathbf{x} = \int_V \left(\frac{g}{f^*} \right) f^* d\mathbf{x} \simeq \left(\overline{\frac{g}{f^*}} \right) \pm \sqrt{\frac{(\overline{g^2/f^{*2}}) - (\overline{g/f^*})^2}{N}}. \quad (5.13)$$

But what choice to make for f^* ? Intuitively, a good choice would be one that makes g/f^* relatively constant over V . More rigorously, the numerator in the square-root

term in Eq. (5.13) should be made as small as possible. In particular, one seeks to minimize

$$\begin{aligned}\sigma^2(g/f^*) &\equiv \left\langle \frac{g^2}{f^{*2}} \right\rangle - \left\langle \frac{g}{f^*} \right\rangle^2 = \int_V \frac{g^2}{f^{*2}} f^* d\mathbf{x} - \left[\int_V \frac{g}{f^*} f^* d\mathbf{x} \right]^2 \\ &= \int_V \frac{g^2}{f^*} d\mathbf{x} - \left[\int_V g d\mathbf{x} \right]^2.\end{aligned}\quad (5.14)$$

Hence, to minimize $\sigma^2(g/f^*)$, f^* must be chosen to make the integral $\int_V (g^2/f^*) d\mathbf{x}$ as small as possible. This minimization is achieved when

$$f^*(\mathbf{x}) = |g(\mathbf{x})| / \int_V |g(\mathbf{x})| d\mathbf{x}. \quad (5.15)$$

To prove this result, apply Schwartz's inequality² to the integral

$$\begin{aligned}\left(\int_V |g(\mathbf{x})| d\mathbf{x} \right)^2 &= \left(\int_V \frac{|g(\mathbf{x})|}{\sqrt{f^*(\mathbf{x})}} \sqrt{f^*(\mathbf{x})} d\mathbf{x} \right)^2 \\ &\leq \left(\int_V \frac{g^2(\mathbf{x})}{f^*(\mathbf{x})} d\mathbf{x} \right) \left(\int_V f^*(\mathbf{x}) d\mathbf{x} \right) = \int_V \frac{g^2(\mathbf{x})}{f^*(\mathbf{x})} d\mathbf{x},\end{aligned}\quad (5.16)$$

where the fact that f^* is a PDF has been used to set $\int_V f^* d\mathbf{x} = 1$. The integral on the right side is precisely the integral that must be minimized. Further, when f^* is chosen according to Eq. (5.15), the right integral equals that on the left side, i.e., the equality in Eq. (5.16) holds. In other words, the integral $\int_V (g^2/f^*) d\mathbf{x}$ is minimized, and the minimum of $\sigma^2(g/f^*)$ is given by

$$\sigma(g/f^*)_{\min} = \left(\int_V |g(\mathbf{x})| d\mathbf{x} \right)^2 - \left(\int_V g(\mathbf{x}) d\mathbf{x} \right)^2. \quad (5.17)$$

In particular, if g does not change sign in V , $\sigma(g/f^*) = 0$!

This estimator of I with a greatly reduced variance (or even a zero variance if g is of one sign) is achieved if the importance is chosen as $f^* = |g| / \int_V |g| d\mathbf{x}$. But to evaluate $\int_V |g| d\mathbf{x}$ is tantamount to evaluating I , and there would be no point in using Monte Carlo to find I . But all is not lost. The variance can be substantially reduced by choosing an importance function f^* that has a shape similar to that of g . With such a choice, importance sampling tries to smooth the values of g/f^* in regions where g varies rapidly. How effective importance sampling is depends on how effective g/f^* is in reducing variations in the integrand. Also when choosing such an approximate f^* , careful consideration must be given to the problem of how to sample values of $\mathbf{x}$ from it. This sampling can be difficult if g is not a well-behaved function, e.g., if g is a function with many rapid fluctuations.

² For those who have forgotten, Schwartz's inequality for the two functions u and w is

$$\left(\int_V uw d\mathbf{x} \right)^2 \leq \left(\int_V u^2 d\mathbf{x} \right) \left(\int_V w^2 d\mathbf{x} \right).$$

5.3 Systematic Sampling

Estimation of $\langle z \rangle$ often can be obtained more accurately by subdividing the integration volume V into M contiguous mutually exclusive subregions V_m , such that

$$V = \sum_{m=1}^M V_m. \quad (5.18)$$

Let p_m be an assigned probability of sampling from region m , where, of course, $\sum_{m=1}^M p_m = 1$. In systematic sampling, p_m is the probability that $\mathbf{x}_i$ lies in V_m , namely,

$$p_m = \int_{V_m} f(\mathbf{x}) d\mathbf{x} / \int_V f(\mathbf{x}) d\mathbf{x} = \int_{V_m} f(\mathbf{x}) d\mathbf{x}, \quad (5.19)$$

because $f(\mathbf{x})$ is a PDF over V . Now define

$$f_m(\mathbf{x}) = \begin{cases} f(\mathbf{x})/p_m, & \mathbf{x} \in V_m \\ 0, & \text{otherwise} \end{cases}. \quad (5.20)$$

Note that

$$\int_{V_m} f_m(\mathbf{x}) d\mathbf{x} = \frac{1}{p_m} \int_{V_m} f(\mathbf{x}) d\mathbf{x} = \frac{p_m}{p_m} = 1, \quad (5.21)$$

as must be the case for a PDF.

The population mean is the same regardless of the way in which sampling is performed to obtain the sample mean. Nevertheless, the population mean can be written in terms of the p_m , as follows:

$$\langle z \rangle = \int_V z(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} = \sum_{m=1}^M p_m \int_{V_m} z(\mathbf{x}) f_m(\mathbf{x}) d\mathbf{x}, \quad (5.22)$$

or, more concisely,

$$\langle z \rangle = \sum_{m=1}^M p_m \langle z \rangle_m, \quad (5.23)$$

where $\langle z \rangle_m$ is defined by

$$\langle z \rangle_m = \int_{V_m} z(\mathbf{x}) f_m(\mathbf{x}) d\mathbf{x}. \quad (5.24)$$

Similarly, the population variance can be expressed as

$$\sigma^2(z) = \int_V [z(\mathbf{x}) - \langle z \rangle]^2 f(\mathbf{x}) d\mathbf{x} = \int_V [z^2(\mathbf{x}) - 2z(\mathbf{x})\langle z \rangle + \langle z \rangle^2] d\mathbf{x}. \quad (5.25)$$

By Eq. (5.20), $f(\mathbf{x}) = p_m f_m(\mathbf{x})$, $\mathbf{x} \in V_m$, and so

$$\begin{aligned}\sigma^2(z) &= \sum_{m=1}^M p_m \int_{V_m} [z^2(\mathbf{x}) - 2z(\mathbf{x})\langle z \rangle + \langle z \rangle^2] f_m(\mathbf{x}) d\mathbf{x} \\ &= \sum_{m=1}^M p_m [\langle z^2 \rangle_m - 2\langle z \rangle \langle z \rangle_m + \langle z \rangle^2],\end{aligned}\quad (5.26)$$

where

$$\langle z^2 \rangle_m = \int_{V_m} z^2(\mathbf{x}) f_m(\mathbf{x}) d\mathbf{x}. \quad (5.27)$$

Systematic sampling assumes the p_m are known and defines

$$N_m = p_m N, \quad (5.28)$$

where N is the total number of histories. The systematic sampling estimator of $\langle z \rangle$ can be expressed, using Eq. (5.28), as

$$\bar{z}_{\text{sys}} = \sum_{m=1}^M \frac{p_m}{N_m} \sum_{i_m=1}^{N_m} z(\mathbf{x}_{i_m}), \quad (5.29)$$

where the $\mathbf{x}_{i_m}$ are sampled from V_m according to the PDF $f_m(\mathbf{x})$. This can be written more concisely in the form

$$\bar{z}_{\text{sys}} = \sum_{m=1}^M p_m \bar{z}_m, \quad (5.30)$$

where

$$\bar{z}_m = \frac{1}{N_m} \sum_{i_m=1}^{N_m} z(\mathbf{x}_{i_m}). \quad (5.31)$$

It is clear that Eq. (5.30) is an unbiased estimator of $\langle z \rangle$ because, by the law of large numbers, each $\bar{z}_m$ approaches each $\langle z_m \rangle$.

The variance of this estimator is $\langle (\bar{z}_{\text{sys}} - \langle z \rangle)^2 \rangle$. Substitution of Eq. (5.23) and Eq. (5.30) into this expression yields

$$\begin{aligned}\sigma^2(\bar{z}_{\text{sys}}) &= \left\langle \left\{ \sum_{m=1}^M \frac{p_m}{N_m} \sum_{i_m=1}^{N_m} [z(\mathbf{x}_{i_m}) - \langle z \rangle_m] \right\}^2 \right\rangle \\ &= \sum_{m=1}^M \frac{p_m^2}{N_m^2} \sum_{i_m=1}^{N_m} \langle [z(\mathbf{x}_{i_m}) - \langle z \rangle_m]^2 \rangle \\ &\quad + \text{cross product terms that average to zero}\end{aligned}$$

$$= \sum_{m=1}^M \frac{p_m^2}{N_m^2} \sum_{i_m=1}^{N_m} \sigma_m^2 = \sum_{m=1}^M \frac{p_m^2}{N_m} \sigma_m^2. \quad (5.32)$$

But in systematic sampling $p_m = N_m/N$ and, thus, the variance of the estimator is

$$\sigma^2(\bar{z}_{\text{sys}}) = \frac{1}{N} \sum_{m=1}^M p_m \sigma_m^2(z), \quad (5.33)$$

where

$$\sigma_m^2(z) \equiv \int_{V_m} [z(\mathbf{x}) - \langle z \rangle_m]^2 f_m(\mathbf{x}) d\mathbf{x} \simeq \frac{1}{N_m} \sum_{i_m=1}^{N_m} z^2(\mathbf{x}_{i_m}) - \left[\frac{1}{N_m} \sum_{i_m=1}^{N_m} z(\mathbf{x}_{i_m}) \right]^2. \quad (5.34)$$

5.3.1 Comparison to Straightforward Sampling

In straightforward sampling, N values of $\mathbf{x}_i$ are sampled from the PDF $f(\mathbf{x})$ over V , and the expected value of z is estimated as [Kahn 1956]

$$\bar{z}_{sf} = \frac{1}{N} \sum_{i=1}^N z(\mathbf{x}_i) \simeq \langle z \rangle = \int_V z(\mathbf{x}) f(\mathbf{x}) d\mathbf{x}. \quad (5.35)$$

The probability a sample point $\mathbf{x}_i$ is in V_m is simply p_m . The variance of this estimator is calculated as follows:

$$\begin{aligned} \sigma^2(\bar{z}_{sf}) &= \frac{1}{N} \sigma^2(z) = \frac{1}{N} \langle [z(\mathbf{x}) - \langle z \rangle]^2 \rangle = \frac{1}{N} \int_V [z(\mathbf{x}) - \langle z \rangle]^2 f(\mathbf{x}) d\mathbf{x} \\ &= \frac{1}{N} \sum_{m=1}^M p_m \int_{V_m} [z(\mathbf{x}) - \langle z \rangle]^2 f_m(\mathbf{x}) d\mathbf{x} \\ &= \frac{1}{N} \sum_{m=1}^M p_m \int_{V_m} [(z(\mathbf{x}) - \langle z \rangle_m) + (\langle z \rangle_m - \langle z \rangle)]^2 f(\mathbf{x}) d\mathbf{x} \\ &= \frac{1}{N} \sum_{m=1}^M p_m \left\{ \int_{V_m} [z(\mathbf{x}) - \langle z \rangle_m]^2 f_m(\mathbf{x}) d\mathbf{x} \right. \\ &\quad + 2 \int_{V_m} [z(\mathbf{x}) - \langle z \rangle_m][\langle z \rangle_m - \langle z \rangle] f_m(\mathbf{x}) d\mathbf{x} \\ &\quad \left. + \int_{V_m} [\langle z \rangle_m - \langle z \rangle]^2 f_m(\mathbf{x}) d\mathbf{x} \right\}. \end{aligned}$$

The first integral is just $\sigma_m^2(z)$, the second integral integrates to zero, and the integrand of the third integral is the constant $[\langle z \rangle_m - \langle z \rangle]^2$. Thus,

$$\sigma^2(\bar{z}_{sf}) = \frac{1}{N} \sum_{m=1}^M p_m \sigma_m^2(z) + \frac{1}{N} \sum_{m=1}^M p_m [\langle z \rangle_m - \langle z \rangle]^2. \quad (5.36)$$

Comparison of Eq. (5.33) to Eq. (5.36) shows that the estimator for systematic sampling almost always has a smaller variance than does that for straightforward sampling. Only in the unlikely case that the means of z in all the subregions are the same does systematic sampling produce no advantage. The amount of reduction is seen to equal the variance in the average values of z in the subregions, and for cases in which there is a large variation in z among the different regions, systematic sampling is effective. For cases in which the p_m are known, systematic sampling should always be done. But often it is difficult to obtain good estimates of p_m .

5.3.2 Systematic Sampling to Evaluate an Integral

The following multidimensional integral can be estimated as

$$I = \int_V g(\mathbf{x}) d\mathbf{x} = V \langle g \rangle \simeq V \bar{g}. \quad (5.37)$$

To apply systematic sampling to estimate $\langle g \rangle$, the results of Section 5.3 can be used directly by making the following changes: $z(\mathbf{x}) \rightarrow g(\mathbf{x})$, $f(\mathbf{x}) \rightarrow 1/V$, $f_m(\mathbf{x}) \rightarrow 1/V_m$, and $p_m \rightarrow V_m/V$. Thus, the systematic estimator of $\langle g \rangle$ becomes

$$\bar{g}_{sys} = \sum_{m=1}^M p_m \bar{g}_m = \sum_{m=1}^M \frac{p_m}{N_m} \sum_{i_m=1}^{N_m} g(\mathbf{x}_{i_m}), \quad (5.38)$$

where the $\mathbf{x}_{i_m}$ are uniformly sampled from V_m . The variance of this estimator

$$\sigma^2(\bar{g}_{sys}) = \frac{1}{N} \sum_{m=1}^M p_m \sigma_m^2(g), \quad (5.39)$$

where

$$\sigma_m^2(g) \equiv \frac{1}{V_m} \int_{V_m} [g(\mathbf{x}) - \langle g \rangle_m]^2 d\mathbf{x} \simeq \frac{1}{N_m} \sum_{i_m=1}^{N_m} g^2(\mathbf{x}_{i_m}) - \left[\frac{1}{N_m} \sum_{i_m=1}^{N_m} g(\mathbf{x}_{i_m}) \right]^2. \quad (5.40)$$

5.3.3 Systematic Sampling as Importance Sampling

Like many variance reduction schemes, systematic sampling can be viewed as a special case of importance sampling, because it involves sampling from a slightly

modified PDF, namely,

$$f^*(\mathbf{x}) = \frac{f(\mathbf{x})}{p_m}, \quad \mathbf{x} \in V_m, \quad m = 1, \dots, M. \quad (5.41)$$

The corresponding importance function is thus

$$W(\mathbf{x}) = \frac{f(\mathbf{x})}{f^*(\mathbf{x})} = p_m, \quad \mathbf{x} \in V_m, \quad m = 1, \dots, M. \quad (5.42)$$

5.4 Stratified Sampling

Stratified sampling, also sometimes called *quota* sampling, is akin to systematic sampling in that a predetermined number of samples are taken from each of the M subregions, but the method of selection N_m is quite different. As with systematic sampling, one seeks

$$\langle z \rangle = \int_V z(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} = \sum_{m=1}^M \int_{V_m} z(\mathbf{x}) f(\mathbf{x}) d\mathbf{x}. \quad (5.43)$$

Proceeding exactly as in systematic sampling (see [Section 5.3](#)), the stratified sampling estimator of $\langle z \rangle$ is

$$\bar{z}_{strat} = \sum_{m=1}^M p_m \bar{z}_m = \sum_{m=1}^M \frac{p_m}{N_m} \sum_{i_m=1}^{N_m} z(\mathbf{x}_{i_m}). \quad (5.44)$$

From [Eq. \(5.32\)](#), the variance of this estimator is

$$\sigma^2(\bar{z}_{strat}) = \sum_{m=1}^M \frac{p_m^2}{N_m} \sigma_m^2(z), \quad (5.45)$$

with $\sigma_m^2(z)$ given by [Eq. \(5.34\)](#).

So far, this is the same as systematic sampling. But in stratified sampling the N_m are not predetermined but are chosen so as to minimize the variance of the estimator. To minimize $\sigma^2(\bar{z}_{strat})$ with the restraint that $\sum_{m=1}^M N_m = N$, one seeks the unrestricted minimum of

$$\sum_{m=1}^M \frac{p_m^2 \sigma_m^2(z)}{N_m} + \lambda \sum_{m=1}^M N_m, \quad (5.46)$$

where λ is a Lagrange multiplier. The variation of this quantity with N_j is then set to zero so that

$$0 = \frac{\delta}{\delta N_j} \left\{ \sum_{m=1}^M \left[\frac{p_m^2 \sigma_m^2}{N_m} + \lambda N_m \right] \right\} = \sum_{m=1}^M \frac{\delta}{\delta N_j} \left[\frac{p_m^2 \sigma_m^2}{N_m} + \lambda N_m \right] = -\frac{p_j^2 \sigma_j^2}{N_j^2} + \lambda. \quad (5.47)$$

From this quantity, the optimum number of samples from each region is found to be $N_j = p_j \sigma_j / \sqrt{\lambda}$. Rearranging and summing this result gives

$$\sqrt{\lambda} \sum_{j=1}^M N_j = \sqrt{\lambda} N = \sum_{j=1}^M p_j \sigma_j \equiv \langle \sigma_j \rangle, \quad (5.48)$$

from which the optimum number of samples that minimizes $\sigma^2(\bar{z}_{strat})$ is

$$N_j = \frac{p_j \sigma_j(z)}{\langle \sigma_j \rangle} N = \frac{p_j \sigma_j(z)}{\sum_{j=1}^M p_j \sigma_j(z)} N. \quad (5.49)$$

Thus, in stratified sampling, the number of points sampled in each subregion is proportional to the variance of z in the region. Regions in which z varies rapidly are sampled more often than regions in which z is relatively constant. Substitution of this result (with j replaced by m) into Eq. (5.45) gives the minimum variance of $\bar{z}_{strat}$, namely,

$$\sigma^2(\bar{z}_{strat}) = \frac{1}{N} \left(\sum_{m=1}^M p_m \sigma_m \right)^2 = \frac{1}{N} \langle \sigma_m \rangle^2. \quad (5.50)$$

Unfortunately, the parameters $\sigma_j(z)$ and $\langle \sigma_j \rangle$ are usually not known a priori. It is *not* a good idea to use some of the same histories used to determine $\bar{z}_{strat}$ to also determine σ_j , because these results are correlated and very strange results can occur. One approach to estimate the optimum values of N_j is to estimate σ_j from a small preliminary run to obtain rough estimates, with consideration given to the trade-off between the desired precision and the cost of sampling.

A more serious limitation for stratified sampling is the subdivision of V for high-dimensional integrals, say $d \geq 5$. Division of each axis into k segments causes the number of subvolumes, M^d , to explode with increasing k and greatly confounds accurate estimates of the needed variances for each subregion.

5.4.1 Comparison to Straightforward Sampling

Stratified sampling almost always produces a more efficient estimate of $\langle z \rangle$ than does straightforward sampling, i.e., a smaller variance in the estimator. Subtraction of Eq. (5.50) from Eq. (5.36) yields

$$\begin{aligned} \sigma^2(\bar{z}_{sf}) - \sigma^2(\bar{z}_{strat}) &= \frac{1}{N} \sum_{m=1}^M p_m \sigma_m^2(z) + \frac{1}{N} \sum_{m=1}^M p_m [\langle z \rangle_m - \langle z \rangle]^2 - \frac{1}{N} \langle \sigma_m \rangle^2 \\ &= \frac{1}{N} [\langle \sigma_m^2 \rangle - \langle \sigma_m \rangle^2] + \frac{1}{N} \sum_{m=1}^M p_m [\langle z \rangle_m - \langle z \rangle]^2 \\ &= \frac{1}{N} \left[\sum_{m=1}^M p_j (\sigma_m - \langle \sigma_m \rangle)^2 + \sum_{m=1}^M p_m (\langle z \rangle_m - \langle z \rangle)^2 \right]. \end{aligned} \quad (5.51)$$

This last result is always non-negative so that stratified sampling almost always is more efficient. Only in the case where all the variances and all the means of z in each subregion are equal is no advantage realized, which is a very unlikely possibility.

5.4.2 Importance Sampling versus Stratified Sampling

The differences between importance sampling and stratified sampling are quite distinct. First, importance sampling usually uses a continuous importance function to flatten the integrand, while stratified sampling always breaks the integration volume into subvolumes. Second, in importance sampling points are concentrated in regions where the magnitude of the integrand $|z(\mathbf{x})f(\mathbf{x})|$ is large, while in stratified sampling points are concentrated in regions where the variance is large. These two techniques, thus, seem at odds with each other.

Importance sampling requires an approximate PDF f^* that has a shape similar to $z(\mathbf{x})f(\mathbf{x})$ and that can also be efficiently sampled. With an imperfect f^* , the error decreases only as $N^{-1/2}$ and, in regions in which f^* cannot match rapid variations in $z(\mathbf{x})f(\mathbf{x})$, the sampled function $z(\mathbf{x})f(\mathbf{x})/f^*(\mathbf{x})$ has a large variance. The key to importance sampling is to smooth the sampled function and, only if this is achieved, is importance sampling effective.

Stratified sampling, by contrast, requires no knowledge of the values of $z(\mathbf{x})f(\mathbf{x})$ but, rather, the variances for each subregion are required. It works well if good estimates of the variances in each subregion are available and if subregions can be constructed that reduce the variance of $z(\mathbf{x})f(\mathbf{x})$ in each subregion compared to the variance of $z(\mathbf{x})f(\mathbf{x})$ over the whole volume V . This requires different knowledge of $z(\mathbf{x})f(\mathbf{x})$ than does importance sampling.

In many cases, $z(\mathbf{x})f(\mathbf{x})$ is small everywhere in V except in some small subregion. In such a subregion the standard deviation of $z(\mathbf{x})f(\mathbf{x})$ is often comparable to the mean value, so that both methods produce comparable results. As pointed out by [Press et al. \[1996\]](#), these two methods can also be used together by using one method on a coarse grid and the other in each grid cell.

[Press et al. \[1996\]](#) also describe (1) how importance sampling can be combined with stratified sampling in an adaptive algorithm for multidimensional integrals, and (2) how stratified sampling can be used recursively to refine the subdivision of V into subregions. Although these uses of variance reduction are beyond the scope of this book, they do reveal the high level of sophistication that is presently available for Monte Carlo analyses.

5.5 Correlated Sampling

An extremely powerful method for reducing the variance in a certain class of Monte Carlo problems is that of *correlated sampling*. In many problems, the difference between two almost equal situations is sought. For example, the effect of a small change in some underlying model, geometry, or other system parameter. Rather than perform two independent Monte Carlo analyses for each problem and then subtract

one result from the other to obtain an estimate of the difference, it is often far better to perform both calculations simultaneously and use the same random number sequence for both. By using the same random numbers, the two results will be highly correlated, which results in a smaller variance in the estimate of the difference.

Suppose $\langle z_1 \rangle$ and $\langle z_2 \rangle$ are given by

$$\langle z_1 \rangle = \int_{V_1} z_1(\mathbf{x}) f_1(\mathbf{x}) d\mathbf{x} \quad \text{and} \quad \langle z_2 \rangle = \int_{V_2} z_2(\mathbf{y}) f_2(\mathbf{y}) d\mathbf{y}. \quad (5.52)$$

What is sought is $\langle \Delta z \rangle = \langle z_1 \rangle - \langle z_2 \rangle$. First, generate $\{\mathbf{x}_i\}$, $i = 1, \dots, N$ and $\{\mathbf{y}_i\}$, $i = 1, \dots, N$. Second, estimate $\langle \Delta z \rangle$ as

$$\langle \Delta z \rangle \simeq \frac{1}{N} \sum_{i=1}^N z_1(\mathbf{x}_i) - \frac{1}{N} \sum_{i=1}^N z_2(\mathbf{y}_i) = \frac{1}{N} \sum_{i=1}^N \Delta z_i. \quad (5.53)$$

Now the variance of this estimator of the difference is

$$\sigma^2(\langle \Delta z \rangle) = \sigma_1^2 + \sigma_2^2 - 2\text{cov}(\bar{z}_1, \bar{z}_2), \quad (5.54)$$

where

$$\bar{z}_1 = \frac{1}{N} \sum_{i=1}^N z_1(\mathbf{x}_i), \quad \bar{z}_2 = \frac{1}{N} \sum_{i=1}^N z_2(\mathbf{y}_i), \quad \sigma_i^2 = \langle (\bar{z}_i - \langle z_i \rangle)^2 \rangle, \quad (5.55)$$

and

$$\text{cov}(\bar{z}_1, \bar{z}_2) = \langle (\bar{z}_1 - \langle z_1 \rangle)(\bar{z}_2 - \langle z_2 \rangle) \rangle. \quad (5.56)$$

If the two estimators $\bar{z}_1$ and $\bar{z}_2$ were calculated independently, then $\text{cov}(\bar{z}_1, \bar{z}_2) = 0$. But if the random variables $\mathbf{x}$ and $\mathbf{y}$ are positively correlated and if $f_1(\mathbf{x})$ and $f_2(\mathbf{y})$ are very similar in shape, then $\bar{z}_1$ and $\bar{z}_2$ are strongly correlated so that $\text{cov}(\bar{z}_1, \bar{z}_2) > 0$ and $\sigma^2(\langle \Delta z \rangle) < \sigma_1^2 + \sigma_2^2$.

From this analysis, it is apparent that to reduce the variance of the difference between $\langle z_1 \rangle$ and $\langle z_2 \rangle$ the correlation between these two quantities must be positive and as large as possible. There is no general prescription to ensure this; however, a good way is to control the use of random numbers. In parts of the Monte Carlo simulation that are the same or very similar, the same sequence of random numbers $\{\mathbf{x}_i\}$ and $\{\mathbf{y}_i\}$ should be used. A simple example of using correlated sampling is given in [Example 5.4](#).

5.5.1 Correlated Sampling With One Known Expected Value

Sometimes, when trying to estimate the expected value

$$\langle z \rangle \equiv \int_V z(\mathbf{x}) f(\mathbf{x}) d\mathbf{x}, \quad (5.57)$$

the exact expectation of a very closely related problem, often some idealization of the problem of interest, is known, i.e., one knows a priori

$$\langle z_o \rangle = \int_V z_o(\mathbf{x}) f(\mathbf{x}) d\mathbf{x}. \quad (5.58)$$

Here, $z_o(\mathbf{x})$ is some close approximation to $z(\mathbf{x})$, so that $\langle z \rangle$ is most likely close to $\langle z_o \rangle$. Thus, rather than trying to estimate $\langle z \rangle$ by calculating $\bar{z}$ directly, it is usually better to estimate the difference $\langle z \rangle - \langle z_o \rangle$ and then add this estimate to the known $\langle z_o \rangle$.

More generally, another estimator of $\langle z \rangle$, for any α , is

$$\overline{z(\alpha)} \equiv \bar{z} - \alpha(\bar{z}_o - \langle z_o \rangle). \quad (5.59)$$

This is an unbiased estimate because $\langle \overline{z(\alpha)} \rangle = \langle \bar{z} \rangle - \alpha \langle \bar{z}_o - \langle z_o \rangle \rangle = \langle z \rangle - \alpha \langle \langle z_o \rangle - \langle z_o \rangle \rangle = \langle z \rangle$ for any α . However, the variance of this estimator does depend on α and is given by

$$\sigma^2(\overline{z(\alpha)}) = \frac{1}{N} \left[\sigma^2(z) - 2\alpha \text{cov}(z, z_o) + \alpha^2 \sigma^2(z_o) \right]. \quad (5.60)$$

Here, the variances and covariance are

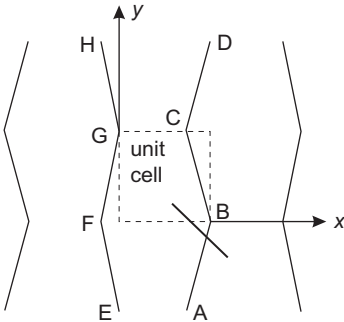
$$\sigma^2(z) \equiv \int_V (z - \langle z \rangle)^2 f(\mathbf{x}) d\mathbf{x} = \langle z^2 \rangle - \langle z \rangle^2 \simeq \bar{z}^2 - \bar{z}^2 \quad (5.61)$$

$$\sigma^2(z_o) \equiv \int_V (z_o - \langle z_o \rangle)^2 f(\mathbf{x}) d\mathbf{x} = \langle z_o^2 \rangle - \langle z_o \rangle^2 \simeq \bar{z}_o^2 - \bar{z}_o^2 \quad (5.62)$$

$$\begin{aligned} \text{cov}(z, z_o) &\equiv \int_V (z - \langle z \rangle)(z_o - \langle z_o \rangle) f(\mathbf{x}) d\mathbf{x} \\ &= \langle z z_o \rangle - \langle z \rangle \langle z_o \rangle \simeq \bar{z} \bar{z}_o - \bar{z} \bar{z}_o. \end{aligned} \quad (5.63)$$

From Eq. (5.60) it is apparent that if $2\alpha \text{cov}(z, z_o) > \alpha^2 \sigma^2(z_o)$, the variance is reduced, i.e., $\sigma^2(\overline{z(\alpha)}) < \sigma^2(z)$.

Example 5.4 A Modified Buffon's Needle Problem



To illustrate correlated sampling, consider the Buffon's needle problem with the parallel grid lines replaced by zigzag grid lines as shown on the left. A unit cell of dimension $D \times D$ is shown by the dashed lines, and each straight line segment has one end displaced in the x -direction by an amount δ with respect to the other end. Because the grid line length is greater with the zigzag grid, it is to be expected that P_{cut}^{zigzag} should increase slightly over $P_{cut}^{\parallel}$ of the standard case of parallel grid lines, with the same spacing D and needle length L . For the parallel-line grid, $P_{cut} = 2L/(\pi D)$ (see Example 1.1).

Unlike the standard Buffon's needle problem, no simple analytical solution is available, and to estimate $\Delta P_{cut} = P_{cut}^{zigzag} - P_{cut}^{\parallel}$ the difference between two simulations, one for each type of grid lines, is used. The simulation for the parallel grid lines is discussed in Example 1.3. For the zigzag grid lines, the simulation proceeds as follows. First, the coordinates (x_i, y_i) of a needle end are selected and uniformly distributed in the unit cell. The angle of the needle with respect to the positive x -axis is then chosen uniformly in $(0, 2\pi)$. Then the coordinates of and the distance to the intersection of a line through the needle and the lines through grid segments AB, BC, CD, EF, FG, and GH are computed. If any of the intersection distances are less than L and the intersection point is within the vertical coordinates of these grid segments, the needle intersects the grid. For small line displacements δ , a simulated needle that intersects the zigzag grid also has a good chance of intersecting the parallel-line grid, so that the results of the two simulations have large positive correlations.

Examples of results of the differences between simulations using the same needle drops for both grids (correlated samplings) and independent drops for each grid are shown in Fig. 5.5.

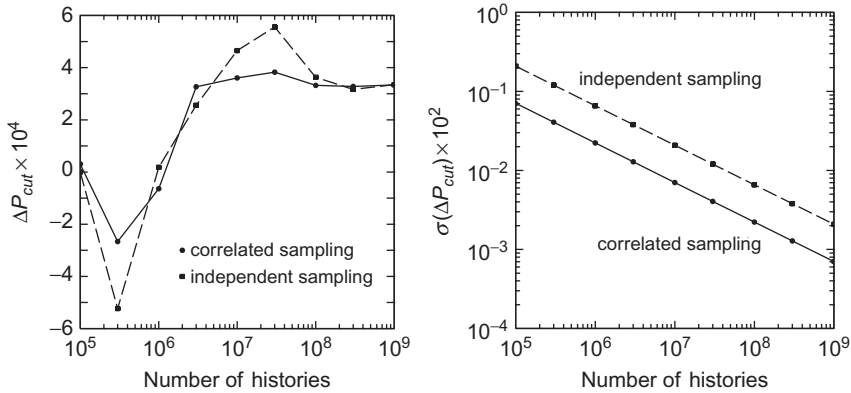


Figure 5.5 Results for correlated and independent Buffon's needle simulations to estimate the change in P_{cut} , ΔP_{cut} , when the straight parallel grid lines are replaced by zigzag grid lines. Results are for $D = 2L = 20\delta$. The left figure shows the convergence of ΔP_{cut} to the value of 0.000334. The right figure shows the estimate of the standard deviation of the estimated ΔP_{cut} . As expected, correlated sampling has a smaller standard deviation.

From Schwartz's inequality, it follows that

$$\int_V [z(\mathbf{x}) - \langle z \rangle]^2 f(\mathbf{x}) d\mathbf{x} \int_V [z_o(\mathbf{x}) - \langle z_o \rangle]^2 f(\mathbf{x}) d\mathbf{x} \geq \left[\int_V [z(\mathbf{x}) - \langle z \rangle][z_o(\mathbf{x}) - \langle z_o \rangle] f(\mathbf{x}) d\mathbf{x} \right]^2, \quad (5.64)$$

or

$$\sigma^2(z)\sigma^2(z_o) \geq \text{cov}^2(z, z_o). \quad (5.65)$$

From this result, it follows that the *correlation coefficient*, defined as

$$\rho(z, z_o) \equiv \frac{\text{cov}(z, z_o)}{\sigma(z)\sigma(z_o)}, \quad (5.66)$$

has the property $-1 \leq \rho(z, z_o) \leq 1$. In terms of this correlation coefficient,

$$\sigma^2(\overline{z(\alpha)}) = \frac{1}{N} \left[\sigma^2(z) - 2\alpha\sigma(z)\sigma(z_o)\rho(z, z_o) + \alpha^2\sigma^2(z_o) \right]. \quad (5.67)$$

The optimum value of α that minimizes $\sigma^2(\overline{z(\alpha)})$ is found by requiring

$$\frac{d\sigma^2(\overline{z(\alpha)})}{d\alpha} = 0. \quad (5.68)$$

From this requirement it is easily found that $\alpha_{opt} = \sigma(z)\rho(z, z_o)/\sigma(z_o)$. With the optimum value the minimum variance is

$$\sigma^2(\overline{z(\alpha_{opt})}) = \frac{1}{N} \sigma^2(z) [1 - \rho^2(z, z_o)]. \quad (5.69)$$

Hence, the more z is correlated with z_o , the smaller is the variance of the estimator of Eq. (5.59).

5.5.2 Antithetic Variates

Consider the simple problem of estimating the integral

$$I = \int_0^1 g(x) dx \simeq \frac{1}{N} \sum_{i=1}^N g(x_i) = I_1, \quad (5.70)$$

where the x_i are uniformly distributed in $(0, 1)$. [Hammersley and Morton \[1956\]](#) proposed that if two unbiased estimators I_1 and I_2 for some quantity I could be found that have a strong negative correlation, a better estimator of I may be $[I_1 + I_2]/2$. The variance of this estimator is

$$\sigma^2 \left(\frac{I_1 + I_2}{2} \right) = \frac{1}{4} \sigma^2(I_1) + \frac{1}{4} \sigma^2(I_2) + \frac{1}{2} \text{cov}(I_1, I_2). \quad (5.71)$$

If $\text{cov}(I_1, I_2)$ is strongly negative, then the average of two negatively correlated estimators may have a much reduced variance.

This integral can also be estimated as

$$I = \int_0^1 g(1-x) dx, \simeq \frac{1}{N} \sum_{j=1}^N g(1-x_j) = I_2, \quad (5.72)$$

where x_i are also uniform in $[0, 1]$. Then the compound estimator

$$I_{12} \equiv \frac{I_1 + I_2}{2} \simeq \frac{1}{2} \sum_{i=1}^N [g(x_i) + g(1 - x_i)] \quad (5.73)$$

is an unbiased estimator because both I_1 and I_2 are unbiased estimators. But, because the same x_i are used for both estimators, there is a negative correlation between the two estimators, and the variance of the compound estimator is usually less than that for I_1 . In fact, it can be shown that for any function $g(x)$, $\sigma^2(I_{12}) \leq \sigma^2(I_1)$ [Kahn 1956]. This is not too surprising because the estimator I_{12} uses twice as many values of $g(x_i)$ than does the I_1 estimator, for the same N .

However, because I_{12} requires twice the computational expense, this estimator is favored only if

$$\sigma^2(I_{12}) \leq \frac{\sigma^2(I_1)}{2}. \quad (5.74)$$

The inequality is indeed the case if $g(x)$ is a continuous monotonic function with a continuous first derivative, as is now demonstrated. The variance of I_{12} is

$$\begin{aligned} \sigma^2(I_{12}) &= \frac{1}{N} \left\{ \frac{1}{4} \sigma^2(I_1) + \frac{1}{4} \sigma^2(I_2) + \frac{1}{2} \text{covar}(I_1, I_2) \right\} \\ &= \frac{1}{N} \left\{ \frac{1}{4} \left[\int_0^1 g^2(x) dx - \langle I_1 \rangle^2 \right] + \frac{1}{4} \left[\int_0^1 g^2(1-x) dx - \langle I_2 \rangle^2 \right] \right. \\ &\quad \left. + \frac{1}{2} \left[\int_0^1 g(x)g(1-x) dx - \langle I_{12} \rangle^2 \right] \right\} \\ &= \frac{1}{N} \left\{ \frac{1}{2} \int_0^1 g(x)g(1-x) dx + \left[\int_0^1 g^2(x) dx - I^2 \right] \right\}, \end{aligned} \quad (5.75)$$

where the last line results from the fact $\langle I_1 \rangle = \langle I_2 \rangle = \langle I_{12} \rangle = I$. The second term of Eq. (5.75) in square brackets is $\sigma^2(I_1)$. Hence, the relation, which holds for any $g(x)$, is

$$2\sigma^2(I_{12}) - \sigma^2(I_1) = \int_0^1 g(x)g(1-x) dx - I^2. \quad (5.76)$$

To prove that $\sigma^2(I_{12}) \leq \sigma^2(I_1)/2$, it is sufficient to show that the right side of Eq. (5.76) is ≤ 0 . Toward this end, consider the function

$$h(x) = \int_0^x g(1-x') dx' - xI, \quad 0 \leq x \leq 1, \quad (5.77)$$

of which the derivative is

$$h'(x) = g(1 - x) - I. \quad (5.78)$$

Now, if $g(x)$ is a monotonic decreasing function of x , then $g(1 - x)$ and $h'(x)$ are monotonic increasing functions of x . Because I is the average value of the monotonic decreasing function $g(x)$ in the interval $[0, 1]$, it follows that $h'(0) < 0$ and $h'(1) > 0$, so that $h(x) \leq 0$ for $x \in [0, 1]$. Then, because in the interval $[0, 1]$, $h(x) \leq 0$, and $g'(x) \leq 0$,

$$\int_0^1 h(x)g'(x) dx \geq 0. \quad (5.79)$$

Integration by parts and use of the boundary values $h(0) = h(1) = 0$ yields the relation

$$\int_0^1 h'(x)g(x) dx \leq 0. \quad (5.80)$$

Substitution of [Eq. \(5.78\)](#) into this result yields

$$\int_0^1 g(1 - x)g(x) dx - \int_0^1 Ig(x) dx = \int_0^1 g(1 - x)g(x) dx - I^2 \leq 0. \quad (5.81)$$

A similar analysis produces the same result if $g(x)$ is a monotonic increasing function of x . Thus, the right side of [Eq. \(5.76\)](#) is indeed ≤ 0 , and the relation of [Eq. \(5.74\)](#) is proved.

Generalization

A more general problem is to evaluate

$$I = \int_{-\infty}^{\infty} g(x)f(x) dx, \quad (5.82)$$

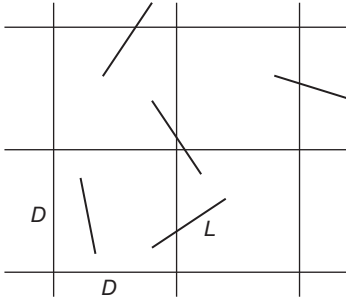
where $f(x)$ is a PDF with a cumulative distribution $F(x)$. Then by analogy to [Eq. \(5.73\)](#) this integral can be estimated as

$$I \simeq \frac{1}{2N} \sum_{i=1}^N [g(x_i) + g(y_i)], \quad (5.83)$$

where $x_i = F^{-1}(\rho_i)$ and $y_i = F^{-1}(1 - \rho_i)$. The pair x_i and y_i are negatively correlated because of the use of the same ρ_i to generate both. Thus, this estimator may have a smaller variance than that of the sample-mean estimator.

The use of antithetic variates is demonstrated in [Example 5.5](#).

Example 5.5 Buffon's Needle Problem with Antithetic Sampling



To illustrate antithetic sampling, consider the Buffon's needle problem using a square grid of vertical and horizontal lines as shown to the left. Each cell is $D \times D$ and needles of length L are randomly dropped on this grid. After dropping N needles, the number N_h cutting a horizontal grid line and the number N_v cutting a vertical grid line are recorded. Then, two independent estimators of P_{cut} are

$$\bar{P}_1 = \frac{N_h}{N} \quad \text{and} \quad \bar{P}_2 = \frac{N_v}{N}.$$

These two estimators, especially for needle lengths L comparable to the grid spacing D , are negatively correlated. For example, needles landing close to the horizontal have a good chance of cutting a vertical grid line but not a horizontal one. Similarly, vertically oriented needles often intersect horizontal lines but not the vertical lines. The compound estimator,

$$\bar{P}_{12} \equiv \frac{1}{2}[\bar{P}_1 + \bar{P}_2] = \frac{1}{2} \left[\frac{N_v + N_h}{N} \right],$$

thus has a smaller variance than either of the separate estimators. An example of how the relative error varies with the number of needle drops for the $\bar{P}_{12}$ estimator and for the $\bar{P}_1$ or $\bar{P}_2$ estimator is shown in Fig. 5.6.

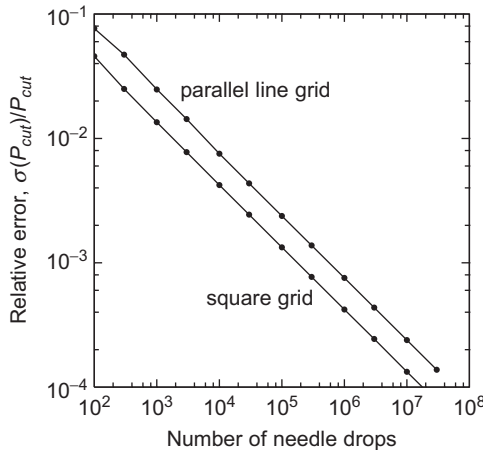


Figure 5.6 Results for a square grid with $D = L$ and for a grid of parallel lines. For the square grid the $\bar{P}_{12}$ estimator based on the antithetic variates N_v and N_h is used and is seen to have a relative error of almost one-half of that of the $\bar{P}_1$ or $\bar{P}_2$ estimator. For this example, the $\text{var}(\bar{P}_1) \simeq 3.2\text{var}(\bar{P}_{12})$.

5.6 Partition of the Integration Volume

Sometimes the integration volume V can be broken into two or more parts, and the integration over some of the subregions can be performed analytically. It is almost always better to use this partial exact information. For example, suppose $V = V_1 \cup V_2$, so that

$$I = \int_V z(\mathbf{x})f(\mathbf{x}) d\mathbf{x} = \int_{V_1} z(\mathbf{x})f(\mathbf{x}) d\mathbf{x} + \int_{V_2} z(\mathbf{x})f(\mathbf{x}) d\mathbf{x}, \quad (5.84)$$

where $I_1 = \int_{V_1} z(\mathbf{x})f(\mathbf{x}) d\mathbf{x}$ can be calculated analytically. Now define a truncated PDF for subregion V_2 as

$$h(\mathbf{x}) = \begin{cases} \frac{f(\mathbf{x})}{1-P}, & \text{if } \mathbf{x} \in V_2 \\ 0, & \text{otherwise} \end{cases}, \quad (5.85)$$

where $P = \int_{V_1} f(\mathbf{x}) d\mathbf{x}$. Eq. (5.84) can, thus, be written as

$$\begin{aligned} I &= I_1 + (1-P) \int_{V_2} z(\mathbf{x}) \frac{f(\mathbf{x})}{1-P} d\mathbf{x} \\ &= I_1 + (1-P) \int_{V_2} z(\mathbf{x}) h(\mathbf{x}) d\mathbf{x} = I_1 + (1-P) \langle z \rangle_2. \end{aligned} \quad (5.86)$$

An unbiased estimator of I is thus

$$I_{part} = I_1 + \frac{(1-P)}{N} \sum_{i=1}^N z(\mathbf{x}_i), \quad (5.87)$$

where the $\mathbf{x}_i$ are sampled from $h(\mathbf{x})$. It can be shown that the variance of this estimator is less than that of Eq. (5.3) [Rubinstein 1981].

5.7 Reduction of Dimensionality

If the n -dimensional integral

$$I = \int_V z(\mathbf{x})f(\mathbf{x}) d\mathbf{x} \quad (5.88)$$

can be decomposed as

$$I = \int_{V_1} \int_{V_2} z(\mathbf{u}, \mathbf{v}) f(\mathbf{u}, \mathbf{v}) d\mathbf{v} d\mathbf{u}, \quad (5.89)$$

where $\mathbf{u} = (x_1, x_2, \dots, x_m)$ and $\mathbf{v} = (x_{m+1}, \dots, x_n)$. Suppose the integration over $\mathbf{v}$ can be performed analytically so that the marginal PDF

$$f_{\mathbf{u}}(\mathbf{u}) = \int_{V_2} f(\mathbf{u}, \mathbf{v}) d\mathbf{v} \quad (5.90)$$

and the conditional expectation

$$\langle z(\mathbf{v}|\mathbf{u}) \rangle = \int_{V_2} z(\mathbf{u}, \mathbf{v}) f(\mathbf{v}|\mathbf{u}) d\mathbf{v} \quad (5.91)$$

can be calculated analytically. The conditional probability function is found from Eq. (2.44) as

$$f(\mathbf{v}|\mathbf{u}) = \frac{f(\mathbf{u}, \mathbf{v})}{f_{\mathbf{u}}(\mathbf{u})}. \quad (5.92)$$

Finally, I can be estimated as follows. From Eqs. (5.89) and (5.92),

$$\begin{aligned} I &= \int_{V_1} \int_{V_2} z(\mathbf{u}, \mathbf{v}) f(\mathbf{v}|\mathbf{u}) f_{\mathbf{u}}(\mathbf{u}) d\mathbf{v} d\mathbf{u}, \\ &= \int_{V_1} f_{\mathbf{u}}(\mathbf{u}) \left[\int_{V_2} z(\mathbf{u}, \mathbf{v}) f(\mathbf{v}|\mathbf{u}) d\mathbf{v} \right] d\mathbf{u}, \\ &\equiv \int_{V_1} f_{\mathbf{u}}(\mathbf{u}) \left\langle \int_{V_2} z(\mathbf{v}|\mathbf{u}) \right\rangle d\mathbf{u}, \\ &\simeq \frac{1}{N} \sum_{i=1}^N \langle z(\mathbf{v}|\mathbf{u}_i) \rangle, \end{aligned} \quad (5.93)$$

where the $\mathbf{u}_i$ are sampled from the marginal PDF $f_{\mathbf{u}}(\mathbf{u})$.

This method of reducing the dimensionality of the problem is sometimes referred to as the *expected value* and was introduced by Buslenko [Schreider 1966]. The estimator of Eq. (5.93) always has a smaller variance than the usual sample-mean estimator of Eq. (5.3) [Rubinstein 1981].

5.8 Russian Roulette and Splitting

In many variance reduction schemes, the integration region is subdivided into M contiguous subregions V_m , i.e.,

$$\langle z \rangle = \int_V z(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} = \sum_{m=1}^M \int_{V_m} z(\mathbf{x}) f(\mathbf{x}) d\mathbf{x}, \quad (5.94)$$

and the sampling from each region is modified over that predicted from $f(\mathbf{x})$. For example, in both the importance and stratified sampling techniques, regions with a

large influence on $\bar{z}$, or its variance, are sampled more frequently than unimportant regions, each subregion having an assigned probability p_m that a random value selected from the PDF $f(\mathbf{x})$ will be taken from that subregion.

However, it is often difficult to do this. Often, the p_m are not known a priori, or the p_m results from many levels of sampling so that it is difficult to make the number of samplings taken from V_m to be proportional to p_m .

To address such cases, J. von Neumann and S. Ulam developed a technique called *Russian roulette and splitting*, named after a supposedly popular game of chance played in the Russian army. In this technique, each subregion is classified as *computationally attractive* or *computationally unattractive*. In computationally attractive subregions either the variance of $z(\mathbf{x})$ is large or the cost of picking $\mathbf{x}_i$ and evaluating $z(\mathbf{x}_i)$ is relatively small. Such regions should be sampled often. The exact opposite is true in computationally unattractive subregions, i.e., the variance of $z(\mathbf{x})$ is small or the cost of computing $z(\mathbf{x}_i)$ is large, and to save computer time unattractive regions should be avoided. As an extreme example, if $z(\mathbf{x})$ were constant in a particular subregion, a single sample would give $\langle z \rangle_m$, and hence multiple samplings from such a region is a waste of time.

To implement Russian roulette and splitting, a particular subregion V_m is picked. If the selected subregion is unattractive, then Russian roulette is used. A value of $\mathbf{x}_i$ for this region is made only some of the time, with a predetermined probability q_m , and a weight of $z(\mathbf{x})/q_m$ is recorded; otherwise, with a probability $1 - q_m$, no sample is taken. The more unattractive the subregion, the smaller is q_m . As an example, if $q_m = 0.75$ no sample is made 25% of the time and 75% of the time a sample is taken.

By contrast, if an attractive subregion is selected, splitting is used. To implement this, each attractive region is assigned a value $r_m > 1$. The more attractive the region, the larger is r_m . Denote by $k_m = [r_m]$ the largest integer in r_m , and set $q_m = r_m - k_m < 1$. Then, with probability q_m , $k_m + 1$ independent samples $\mathbf{z}_i$ are taken from V_m , and, with probability $1 - q_m$, k_m samples are used. For example, if $r_m = 2.25$, 75% of time the sample is split 3 for 1 and 25% of the time the split is 2 for 1. To remove the sampling bias, the weight for each sample is set to $z(\mathbf{x}_i)/r_m$.

Russian roulette combined with splitting is one of the oldest variance reduction schemes and is widely used in Monte Carlo analyses, particularly in Monte Carlo simulations. When used properly, considerable computer time can be saved because, in unattractive regions, fewer samples are taken (in order to save computer time) and, in attractive regions, more samples are taken (in order to produce a better estimate of $\langle z \rangle$). However, oversplitting can waste computer time because, although splitting decreases the variance per selected subregion, it increases computing time. By contrast, Russian roulette increases the variance per initial sample but decreases computing time.

5.8.1 Application to Monte Carlo Simulation

Russian roulette and splitting are widely used in Monte Carlo simulations of problems for which $z(\mathbf{x})$ is not known explicitly and for which the calculation involves a complex geometry and many physical models.

Suppose, for example, one is trying to estimate the expected number of annual traffic accidents at a particular intersection in a town. Many models are needed in such a simulation to specify the probabilities of when and where car trips are started, weather conditions, speeds and accelerations, the many possible turns drivers may take, etc. For each car trip history, the time the car reaches the intersection in question is recorded. An analog simulation would require millions of histories to even find a few near coincidental arrivals at the intersection. Most histories would never reach the intersection in question and produce a zero score. For such a simulation, Russian roulette and splitting would be very effective. Histories in which a car leaves the town on a highway have a low probability of returning and reaching the intersection. Such histories can be terminated by Russian roulette. Then histories in which cars turn onto streets leading to the intersection can be split to produce more arrivals at the intersection. Likewise, trips that begin between 1 and 4 in the morning can be terminated by Russian roulette because of the low traffic density, while trips that begin at rush hour should be split. Clearly, the decision of when to kill or split a history depends on the insight of the analyst.

In such simulations, a history is given some initial weight. The history is tracked as it moves through the various subregions in the problem space. Under certain adverse conditions, one may choose to terminate histories with probability p . Those histories not terminated (with probability $1 - p$) continue with modified weight

$$W' = \frac{W}{1 - p}, \quad (5.95)$$

where W is the incoming weight and W' is the outgoing weight. Under other favorable conditions, one may choose to split histories, replacing one having weight W with m , each having weight

$$W' = \frac{W}{m}. \quad (5.96)$$

Although splitting and Russian roulette can be extremely effective in problems in which very few histories ever produce a nonzero score, it is often difficult to subdivide the problem phase space into contiguous subregions and then to assign effective importance values to each region. For simple problems of low dimensionality, regions near scoring regions are given larger importances than regions far from the scoring regions. Also, as a general rule, large changes in importance between adjacent regions should be avoided.

5.9 Combinations of Different Variance Reduction Techniques

Until now, each variance reduction technique has been examined as if it were the only biased technique being used for a particular problem. By showing that the expected value of the tally in the biased game is the same as that for the unbiased (analog) game,

the various biased games are shown to be “fair games.” Then from examination of the second moments of the biased and unbiased games, conditions can be developed so that the biased game produced a tally with a smaller variance than that of the unbiased game.

In many complex Monte Carlo analyses, particularly those based on simulations of nature, it is observed empirically that combining different variance reduction techniques can produce tallies with smaller variances than can be obtained using each single variance reduction technique alone. Yet, until relatively recently, no rigorous proof was available to guarantee that an arbitrary combination of variance reduction techniques, each of which alone produces a fair game, also yields a fair game. In other words, is any combination of fair games also a fair game?

Proofs of fairness are available for specialized cases. [Lux and Koblinger \[1991\]](#) showed that the combination of splitting and biased kernels gives unbiased scores. [Booth \[1985\]](#) showed that splitting can be combined with *any* variance reduction techniques provided the techniques do not depend on the particle weight. [Booth and Pederson \[1992\]](#) extended this result to show that any arbitrary combination of variance reduction techniques produces an unbiased tally provided the techniques are again weight independent. This proviso means that if no weight-dependent games are played, the particle random walks are independent of weight, and particles with different weights have the same random walks for the same random number sequence and produce the same score for a constant weight factor.

However, some of the most powerful variance reduction techniques do depend heavily on the particle weight (e.g., weight windows and energy cut-offs discussed in Chapter 10 for radiation transport). [Booth and Pederson \[1992\]](#) finally provided a theoretical basis to show that all possible combinations of variance reduction techniques (each of which produces an unbiased score) also give an unbiased tally. Thus, any combination of fair games is itself a fair game.

5.10 Biased Estimators

Consider the estimation of

$$I = \int_V z(\mathbf{x})f(\mathbf{x}) d\mathbf{x} = \int_V g(\mathbf{x}) d\mathbf{x}, \quad (5.97)$$

where $f(\mathbf{x})$ is a PDF. To decrease the variance of the estimate of I , importance sampling is often used to recast the problem as

$$I = \int_V \frac{g(\mathbf{x})}{f^*(\mathbf{x})} f^*(\mathbf{x}) d\mathbf{x}, \quad (5.98)$$

where $f^*(\mathbf{x})$ is a PDF that is approximately proportional to $g(\mathbf{x})$. An unbiased estimator of I is the straightforward sample-mean estimator

$$I_{sf} = \frac{1}{N} \sum_{i=1}^N \frac{g(\mathbf{x}_i)}{f^*(\mathbf{x}_i)}. \quad (5.99)$$

Sometimes the PDF $f^*(\mathbf{x})$ from which the random variables are to be generated is very complicated and the sampling is difficult. For example, in the above importance sampling, if $g(\mathbf{x})$ is a complicated function, then $f^*(\mathbf{x})$ can also become complicated and sampling from it can be difficult. [Powell and Swann \[1966\]](#) suggest a method called *weighted uniform sampling* that samples from a uniform distribution so as to avoid such difficulties.

Another estimator of I , albeit a biased one, is

$$\bar{I}_{bias} = \frac{\sum_{i=1}^N g(\mathbf{u}_i)}{\sum_{i=1}^N f^*(\mathbf{u}_i)}, \quad (5.100)$$

where the $\mathbf{u}_i$ are sampled uniformly in V . It is easy to show that $\bar{I}_{bias}$ is a biased estimator, namely,

$$\langle \bar{I}_{bias} \rangle = \left\langle \frac{\sum_{i=1}^N g(\mathbf{u}_i)}{\sum_{i=1}^N f^*(\mathbf{u}_i)} \right\rangle \neq \frac{V \langle \sum_{i=1}^N g(\mathbf{u}_i) \rangle}{V \langle \sum_{i=1}^N f^*(\mathbf{u}_i) \rangle}. \quad (5.101)$$

Clearly, $\lim_{N \rightarrow \infty} V \sum_{i=1}^N g(\mathbf{u}_i) = V \langle g(\mathbf{u}) \rangle = I$ and $\lim_{N \rightarrow \infty} V \sum_{i=1}^N f^*(\mathbf{u}_i) = 1$. Thus, $\bar{I}_{bias}$ is a *consistent* estimator because

$$\lim_{N \rightarrow \infty} \bar{I}_{bias} = \lim_{N \rightarrow \infty} \left[\frac{V \sum_{i=1}^N g(\mathbf{u}_i)}{V \sum_{i=1}^N f^*(\mathbf{u}_i)} \right] = \left[\frac{\lim_{N \rightarrow \infty} V \sum_{i=1}^N g(\mathbf{u}_i)}{\lim_{N \rightarrow \infty} V \sum_{i=1}^N f^*(\mathbf{u}_i)} \right] = I. \quad (5.102)$$

In this method, then, only uniform sampling from V is required, a great improvement if $f^*(\mathbf{x})$ is complex and difficult to sample. [Powell and Swann \[1966\]](#) show that, for large N , this method is $\sqrt{N}$ times more efficient than the sample-mean method.

5.11 Improved Monte Carlo Integration Schemes

In the usual sample-mean Monte Carlo integration

$$I = \int_a^b z(x)f(x) dx \equiv \int_a^b g(x) dx \simeq \bar{I} \equiv \frac{b-a}{N} \sum_{i=1}^N g(x_i), \quad (5.103)$$

in which points x_i are sampled uniformly from (a, b) , the error in the estimator $\bar{I}$ decreases as $1/\sqrt{N}$. However, this $1/\sqrt{N}$ behavior for Monte Carlo quadrature is not inevitable. Quadrature schemes can be found that have much faster convergence. For example, if an equispaced Cartesian grid were superimposed on the interval (a, b) and each grid point were sampled exactly once, in any order, the relative error decreases

at least as fast as $1/N$. But a grid scheme is not very attractive because one has to decide in advance how fine to make the grid, and all of the grid must be used. Unlike Monte Carlo, one cannot stop a grid-based calculation when a “good enough” result is obtained. All predetermined N points must be evaluated. The advantage of Monte Carlo methods is that N can be increased “on-the-fly” until a prescribed precision is reached. Two Monte Carlo approaches for quadrature are presented as follows that achieve a faster convergence than that based on the uncorrelated random ordinates of the sample-mean method.

5.11.1 Weighted Monte Carlo Integration

A simple Monte Carlo quadrature scheme that has faster convergence than the sample-mean approach is the so-called *weighted Monte Carlo* scheme [Yakowitz et al. 1978]. In this approach, N values of $g(x_i)$ are calculated where the x_i are uniformly distributed over the interval (a, b) . Then the $\{x_i\}$ are ordered in increasing order as $x_{(1)}, \dots, x_{(N)}$ and the trapezoid quadrature approximation is used to approximate the integral between adjacent pairs of the $x_{(i)}$ to give the estimate

$$\begin{aligned} \bar{I} &= \frac{1}{2} \sum_{i=0}^N [g(x_{(i)}) + g(x_{(i+1)})][x_{(i+1)} - x_{(i)}] \\ &= \frac{1}{2} \left\{ \sum_{i=1}^N g(x_{(i)})[x_{(i+1)} - x_{(i-1)}] + g(x_{(0)})[x_{(1)} - x_{(0)}] \right. \\ &\quad \left. + g(x_{(N+1)})[x_{(N+1)} - x_{(N)}] \right\}. \end{aligned} \quad (5.104)$$

where $x_{(0)} = a$ and $x_{(N+1)} = b$. In this scheme, the relative error decreases at least as fast as $1/N$. Moreover, if a more accurate result is needed, additional x_i and $f(x_i)$ values are calculated, included with the previously calculated values, the entire set is then resorted, and then Eq. (5.104) is used to calculate a refined value of $\bar{I}$. In this way, previous values of $f(x_i)$ are not discarded when increasing N . By contrast, if N is changed in deterministic numerical quadrature schemes, all the $\{f(x_i)\}$ must be recomputed, a terrible penalty if the evaluation of $f(x)$ is computationally expensive.

Although the weighted Monte Carlo method has at least $1/N$ convergence, it has the extra computational overhead of sorting the x_i ordinates whenever another refined estimate of the integral is desired. However, very efficient sorting routines are available [Press et al. 1996] and, generally, most of the computational effort is associated with computing $g(x)$. The major drawback with the weighted Monte Carlo integration scheme is that it is difficult to apply to multidimensional cases.

Example 5.6 Buffon's Problem: Weighted Monte Carlo Quadrature

From Eq. (1.6) the probability a dropped needle cuts a grid line is

$$P_{cut} = \int_0^D \frac{s_1(x) + s_2(x)}{2\pi LD} dx \equiv \int_0^D g(x) dx.$$

The functions $s_1(x)$ and $s_2(x)$ are given by Eqs. (1.4) and (1.5), respectively. In the usual sample-mean Monte Carlo approach, P_{cut} is estimated as

$$P_{cut} \simeq \frac{D}{N} \sum_{i=1}^N g(x_i),$$

where the random $\{x_i\}$ are uniformly distributed in the interval $(0, D)$.

With a uniform grid of x_i values over $(0, D)$, with $\Delta x = D/N$, P_{cut} can be estimated from the mid-interval values of $g(x)$ as

$$P_{cut} \simeq \Delta x \sum_{i=1}^N g\left(\left[i - \frac{1}{2}\right] \Delta x\right) = \frac{D}{N} \sum_{i=1}^N g\left(\left[i - \frac{1}{2}\right] \Delta x\right).$$

Finally, P_{cut} can be obtained using the weighted Monte Carlo scheme of Eq. (5.104). A comparison of these three methods is shown in Fig. 5.7 where it is seen that both the Cartesian grid and the weighted Monte Carlo schemes have an error that decreases as $1/N$, much faster than the $1/\sqrt{N}$ of the usual sample-mean Monte Carlo method.

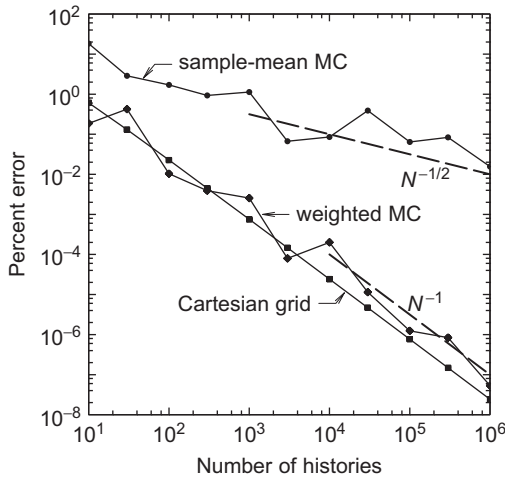


Figure 5.7 Comparison of three different quadrature schemes for the Buffon's needle problems.

5.11.2 Monte Carlo Integration with Quasirandom Numbers

It is possible to improve greatly the convergence of the sample-mean approach by using a different type of sequence for selecting the x_i . This special type of sequence is based, not on pseudorandom numbers that simulate true randomness, but on numbers that strive for uniformity of coverage over the integration volume. Generators that produce such numbers are intentionally designed to be *subrandom* so that generated values, although appearing random, try to avoid previous values, thereby producing a more even-distributedness of the generated values. In other words, numbers from a quasirandom generator try to fill in the gaps left by the previous numbers.

By using various tricks to represent numbers in different prime bases, for example, it is possible to produce long sequences of quasirandom numbers that more uniformly fill the integration volume than do true random numbers [Halton 1960]. Several other ways to generate quasirandom number sequences have been proposed, notably that of Sobol [1967]. An excellent review of such quasirandom number generators is given by Bratley and Fox [1988], and practical implementations are provided by Press et al. [1996].

Can quadrature with these quasirandom number generators achieve a true $1/N$ convergence? The answer is almost yes. For integration of a smooth function in k dimensions, a Sobol sequence has a convergence rate proportional to $(\ln N)^k/N$, which is almost as fast as $1/N$.

5.12 Summary

There are various ways to modify the sampling and scoring procedures of Monte Carlo that do not change the sample mean (in the limit) but do change the sample variance. If these variance reduction procedures are carefully implemented, the sample variance of a Monte Carlo estimate can be reduced, often significantly. However, the variance reduction schemes cannot be applied blindly or automatically. Often, several variance reduction approaches can be used simultaneously to good effect. But sometimes combining different schemes or misusing a single scheme can produce totally wrong results. The effective implementation of variance reduction for a particular problem is part of the “art” of using Monte Carlo.

Bibliography

- BOOTH, T.E. 1985, “Monte Carlo Variance Comparison for Expected-Value Versus Sampled Splitting,” *Nucl. Sci. Eng.*, **89**, 305–309.
- BOOTH, T.E. AND S.P. PEDERSON 1992, “Unbiased Combinations of Nonalog Monte Carlo Techniques and Fair Games,” *Nucl. Sci. Eng.*, **110**, 254–261.
- BRATLEY, P. AND B.L. FOX 1988, “Algorithm 659: Implementing Sobol’s Quasirandom Sequence Generator,” *ACM T. Math. Software*, **14**, 88–100.
- HALTON, J.H. 1960, “On the Efficiency of Certain Quasi-Random Sequences of Points in Evaluating Multi-Dimensional Integrals,” *Numer. Math.*, **2**, 84–90.

- HAMMERSLEY, J.M. AND K.W. MORTON 1956, "A New Monte Carlo Technique: Antithetic Variates," *P. Camb. Philol. Soc.*, **52**, 449–474.
- KAHN, H. 1956, *Applications of Monte Carlo*, AECU-3259, U.S. Atomic Energy Commission, available through Technical Information Service Extension, Oak Ridge, TN.
- LUX, I. AND L. KOBLINGER 1991, *Monte Carlo Particle Transport Methods: Neutron and Photon Calculations*, CRC Press, Boca Raton, FL.
- MARSHALL, A.W. 1956, "The Use of Multi-Stage Sampling Schemes in Monte Carlo Computations," in *Symposium on Monte Carlo Methods*, M.A. Meyer, ed., pp. 123–140, Wiley, New York.
- POWELL, M.I.D. AND I. SWANN 1966, "Weighted Uniform Sampling—A Monte Carlo Technique for Reducing Variance," *J. I. Math. Appl.*, **2**, 228–238.
- PRESS, W.H., S.A. TEUKOLSKY, W.T. VETTERLING, AND B.P. FLANNERY 1996, *Numerical Recipes*, 2nd ed., Cambridge University Press, Cambridge.
- RUBINSTEIN, R.V. 1981, *Simulation and the Monte Carlo Method*, Wiley, New York.
- SCHREIDER, Y.A. (Ed.) 1966, *The Monte Carlo Method (the Method of Statistical Trials)*, Pergamon, Elmsford, NY.
- SHULTIS, J.K. AND R.E. FAW 2000, *Radiation Shielding*, American Nuclear Society, La Grange Park, IL.
- SOBOL, I.M. 1967, Distribution of Points in a Cube and Approximate Evaluation of Integrals *USSR Comp. Maths. Math. Phys.*, **7**(4), 86–112.
- YAKOWITZ, S., J.E. KRIMMEL, AND F. SZIDAROVSKY 1978, "Weighted Monte Carlo Integration," *Soc. Indust. Appl. Math. J. Numer. Anal.*, **15**(6), 1289–1300.

Problems

1. Estimate the value of the integral

$$I = \int_0^1 e^{7x} dx$$

by straightforward sampling, i.e., with the x_i uniformly sampled over the interval $[0, 1]$. The function $3^{7x}/2$ has a shape similar to that of the above integrand. Convert this function to a PDF $f^*(x)$ and use it with importance sampling to estimate I . Estimate the standard deviation of both estimates. Finally, plot the accuracy of your estimates as a function of the number of samples used.

2. Use stratified sampling to estimate the value of the integral $I = \int_V g(\mathbf{x}) d\mathbf{x}$, where

$$f(\mathbf{x}) = f(x, y) = \begin{cases} x^2 y^2 \sin^2 x \sin^2 y, & x > 0, y > 0 \\ x^2 y^2 / 25, & x \leq 0 \text{ or } y \leq 0 \end{cases}.$$

Here, V is the interior of the circle $x^2 + y^2 = 25$. Subdivide V into the four quadrants of the circle and estimate the number of samples to be taken from each quadrant to produce an estimate of I with the smallest variance. Compare your results with that obtained by using the straightforward sample-mean result for a single region.

3. If ρ_i is a random number uniformly sampled from $[0, 1]$, show that the random numbers $x_i = a\rho_i + b$ and $x'_i = a(1 - \rho_i) + b$ have a correlation coefficient equal to -1 .

4. Consider the problem of estimating the small difference between the two integrals

$$\Delta = \int_0^b \frac{\sin \sqrt{x}}{\sqrt{x}} dx - \int_0^b \left[1 - \frac{x}{3!} + \frac{x^2}{5!} - \frac{x^3}{7!} \right] dx.$$

The right integral can be analytically integrated to give

$$I_o = b - \frac{b^2}{12} + \frac{b^3}{360} - \frac{b^4}{20160}.$$

For $b = 5$, use the correlated sampling estimators of Eqs. (5.53) and (5.59) to evaluate Δ . Also, estimate the standard deviations of your estimates.

5. Evaluate the integral

$$I = \int_1^3 x^4 dx$$

by the usual sample-mean method (see Eq. (5.103)) and the antithetic sampling estimator of Eq. (5.73). Plot the percent error and the estimated standard deviation of the estimators as a function of the number of samples used.

6. Prove Eq. (5.81) is still valid if $g(x)$ is a monotonic increasing function of x over the interval $[0, 1]$.
7. Consider evaluation of the integral $\int_V z(\mathbf{x})f(\mathbf{x}) d\mathbf{x}$ using the volume partitioning method of Section 5.6 for the function

$$z(\mathbf{x}) = z(x, y) = \begin{cases} x^2 y^2 \sin x \sin y, & x > 0, y > 0 \\ 1 & x \leq 0, \text{ or } y \leq 0 \end{cases}.$$

Here, V is the interior of the circle $x^2 + y^2 = 25$ and the PDF $f(\mathbf{x}) = 1/(25\pi)$, i.e., points $\mathbf{x} = (x, y)$ are uniformly distributed in V . Compare your estimated precision with that of the straightforward sample-mean method using a single integration volume for a different number of samples N .

8. Consider the integral

$$I = \int_0^\infty x^2 \cos x e^{-x/2} dx.$$

- Use the straightforward sample-mean method with exponential sampling to estimate this integral.
 - Use exponential antithetic samples to estimate this integral.
 - Using both methods, obtain 200 estimates using 1000 samples for each estimate and compare the variance of the two estimators.
9. Consider the integrand of problem 1. Use the biased estimator of Section 5.10 to estimate the value of the integral and compare the accuracy with the number of samples used.
10. Use weighted Monte Carlo integration method as well as the straightforward sample-mean method to evaluate

$$I = \int_{-1}^1 \cos^2 x dx$$

and plot the accuracy of the results as a function of the number of samples used.

11. Suppose that the angular distribution of a neutron-scattering cross section is represented by the truncated Legendre polynomial series

$$f(\omega) = \frac{1}{2} + \frac{1}{2} \sum_{n=1}^N (2n+1) f_n P_n(\omega).$$

Discuss the advantages and disadvantages of sampling ω using (a) the composition method and (b) using importance sampling with $\tilde{f} = 1/2$.

6 Markov Chain Monte Carlo

*A rose by any other name
would smell as sweet and look the same.
So, a chain that is Markov
can be the start of,
just another Monte Carlo game.*

The sampling methods discussed in Chapters 4 and 5 are quite effective for many cases but not for all. Consider again the expected value of a function z in the form

$$\langle z \rangle = \int_V z(\mathbf{x})f(\mathbf{x})d\mathbf{x}, \quad (6.1)$$

where $\mathbf{x}$ is a vector of random variables and f is a PDF defined on a volume or domain V . In principle, one can estimate $\langle z \rangle$ by any of a number of Monte Carlo sampling methods. Indeed, this flexibility is one of the main virtues of Monte Carlo. For instance, one can use the marginal and conditional formulation implied by Eq. (4.23) and pick samples of each variable sequentially by the inverse CDF method or by rejection sampling. Alternatively, one can pick samples from some other distribution and then weight the samples using importance sampling, as discussed in Section 5.2. These standard sampling techniques, however, may not work well in certain complex problems. For instance, the PDF may be difficult to sample in highly multidimensional cases. In other cases, one might hypothesize that the variables are distributed according to complex functions that are difficult to reduce to normalized PDFs, i.e., the shapes of the distributions may be presumed, but the normalization constants required to make the shapes into PDFs may be difficult to determine. In these and other cases, the standard sampling schemes introduced in Chapters 4 and 5 may be of limited utility.

6.1 Markov Chains to the Rescue

Metropolis et al. [1953] recognized that Markov chains could be applied to difficult sampling problems, and Hastings [1970] constructed what now is referred to as the *Metropolis-Hastings* algorithm for performing such samplings. Interest in what has come to be known as *Markov chain Monte Carlo* (MCMC) has exploded in the last few decades. Markov chain Monte Carlo is simply Monte Carlo in which the samples are obtained from a Markov chain. But what is a Markov chain?

A Markov chain is a sequence of random samples $\mathbf{x}_i$, $i = 1, 2, \dots, N$, for which the value $\mathbf{x}_{i+1}$ depends only on the previous value $\mathbf{x}_i$ and not on any of the values $\mathbf{x}_k$, where $k < i$. The terminology often is expressed in a temporal analogy, in which the “next

state” (random sample) $\mathbf{x}_{i+1}$ depends on the “current state” $\mathbf{x}_i$ but not on any earlier states. A chain of states (sequence of random samples) for which this dependence is true is called a Markov chain. The concept of Markov chains was introduced by the Russian mathematician A.A. Markov [Markov 1906], and these chains were given the name Markov chains by Bernstein [1926].

Consider first a single discrete random variable x that moves from the current state x_i to the new state x_j with a certain “transition probability” T_{ij} , which depends only on the current state x_i . If there are m possible states, then an $m \times m$ transition matrix

$$\mathbf{T} = \{T_{ij}\}, \quad i, j = 1, 2, \dots, m \quad (6.2)$$

gives the various probabilities of transitioning among all possible states. Because after each transition one of the m states must be selected, this matrix is normalized, such that $\sum_{j=1}^m T_{ij} = 1$, for any i . Note that T_{ii} is the probability of remaining in state x_i and also that certain transition probabilities may be zero, i.e., some transitions may be prohibited. Each transition from one state to another depends only on the current state and not on how that state happened to evolve. Thus, a sequence of states sampled from $\mathbf{T}$ forms a Markov chain.

The transition matrix is inextricably related to the discrete PDF for the random variable x . Let $p_{ij}^{(n)}$ be the probability of transitioning from state x_i to state x_j after n steps and, for simplicity, let $p_{ij}^{(1)} = p_{ij}$, where p_{ij} is the probability of transitioning from state x_i to state x_j in one step. Now, let

$$\boldsymbol{\pi}^{(n)} = \{\pi_1^{(n)}, \pi_2^{(n)}, \dots, \pi_m^{(n)}\}, \quad (6.3)$$

where $\pi_i^{(n)}$ is the probability that x has the value x_i on the n th step. The probability that the chain assumes state x_i at step $n + 1$ is given by MacKay [2007]

$$\pi_i^{(n+1)} = \sum_{k=1}^m \pi_k^{(n)} T_{ki}. \quad (6.4)$$

The matrix form of this equation

$$\boldsymbol{\pi}^{(n+1)} = \boldsymbol{\pi}^{(n)} \mathbf{T} \quad (6.5)$$

can be used to infer that

$$\boldsymbol{\pi}^{(n)} = \boldsymbol{\pi}^{(n-1)} \mathbf{T} = (\boldsymbol{\pi}^{(n-2)} \mathbf{T}) \mathbf{T} = \boldsymbol{\pi}^{(n-2)} \mathbf{T}^2. \quad (6.6)$$

Thus, the probability of transitioning from state x_i to state x_j in one step is just $p_{ij} = T_{ij}$, and the probability of going from state x_i to state x_j in two steps is given by

$$p_{ij}^{(2)} = \sum_{k=1}^m T_{ik} T_{kj}, \quad (6.7)$$

which is just the product of the i th row of $\mathbf{T}$ with the j th column of $\mathbf{T}$, or the ij th element of $\mathbf{T}^2$.

This result can be easily generalized to show that the probability of transitioning from state x_i to state x_j after n steps, i.e., $p_{ij}^{(n)}$, is simply the ij th element of $\mathbf{T}^n$. This is equivalent to the result, derivable from Eq. (6.5), that

$$\boldsymbol{\pi}^{(n)} = \boldsymbol{\pi}^{(0)} \mathbf{T}^n. \quad (6.8)$$

Eventually, as n grows large enough, $\boldsymbol{\pi}^{(n)}$ approaches the *stationary* distribution or PDF on which the transition matrix is based, as does each row of $\mathbf{T}^n$, i.e., $f_i = \lim_{n \rightarrow \infty} \pi_i^{(n)}$.

The initial state x_i is specified by the row vector $\boldsymbol{\pi}^{(0)}$, all elements of which are zero except the i th element, of which the value is unity. For instance, if the initial state is x_i for $i = 3$ out of $m = 5$ states, the initial row vector would be $\boldsymbol{\pi}^{(0)} = (0, 0, 1, 0, 0)$. Multiplication of $\mathbf{T}^n$ by the initial state row vector produces the i th row of $\mathbf{T}^n$, which gives the probabilities of transitioning from state x_i to each of the m states. [Example 6.1](#) demonstrates some of these concepts for a particular situation.

Not all random variables are discrete and Markov chains can be formed for continuous random variables as well. When x is continuous, the transition probability is a function $T(x \rightarrow u)$, such that the quantity $T(x \rightarrow u)du$ gives the probability of transitioning from state x to a state that is within du about u in one step. If $x, u \in [a, b]$, then, because some state must be selected in each transition, this function is normalized, such that $\int_a^b T(x \rightarrow u) du = 1$. The probability of going from state x to state u in two steps is given by

$$T^{(2)}(x \rightarrow u) = \int_a^b T(x \rightarrow v_1)T(v_1 \rightarrow u)dv_1, \quad (6.9)$$

and the probability of transitioning from state x to state u after n steps is given by the integral over the $n - 1$ intervening steps

$$\begin{aligned} T^{(n)}(x \rightarrow u) &= \int_a^b T(x \rightarrow v_1)T(v_1 \rightarrow v_2) \int_a^b T(v_2 \rightarrow v_3) \dots \\ &\quad \int_a^b T(v_{n-2} \rightarrow v_{n-1}) \int_a^b T(v_{n-1} \rightarrow u)dv_1 dv_2 \dots dv_{n-2} dv_{n-1}. \end{aligned} \quad (6.10)$$

As in the discrete case, the transition function $T(x \rightarrow u)$ inherently contains information about $f(x)$, the PDF of x . In particular, it can be shown [[Tierney 1994](#)] that

$$f(u) = \lim_{n \rightarrow \infty} T^{(n)}(x \rightarrow u). \quad (6.11)$$

In general, a state may be specified by a vector $\mathbf{x}$ of n random variables, each component of which may be either discrete or continuous. The possible values of $\mathbf{x}$ define

Example 6.1 The Weather in Transylvania

Consider that the daily weather in Transylvania is categorized into $m = 3$ possibilities, namely, “sunny” ($x_1 = s$), “cloudy” ($x_2 = c$), and “rainy” ($x_3 = r$). Assume that the transition matrix is known and is given by

$$\mathbf{T} = \begin{pmatrix} T_{ss} & T_{sc} & T_{sr} \\ T_{cs} & T_{cc} & T_{cr} \\ T_{rs} & T_{rc} & T_{rr} \end{pmatrix} = \begin{pmatrix} 0.300 & 0.400 & 0.300 \\ 0.300 & 0.200 & 0.500 \\ 0.400 & 0.100 & 0.500 \end{pmatrix},$$

where the first subscript on T_{ij} indicates the initial state and the second subscript indicates the final state.

Today is a cloudy day in Transylvania. Out of the next three days, on which day is it least likely to rain? This is easily determined by comparing $p_{23}^{(n)}$ for $n = 1, 2, 3$. From $\mathbf{T}$, the probability of rain tomorrow is 50.0%; from

$$\mathbf{T}^2 = \begin{pmatrix} 0.330 & 0.230 & 0.440 \\ 0.350 & 0.210 & 0.440 \\ 0.350 & 0.230 & 0.420 \end{pmatrix},$$

the probability of rain on the second day is 44.0%; and from

$$\mathbf{T}^3 = \begin{pmatrix} 0.3434 & 0.2254 & 0.4312 \\ 0.3430 & 0.2258 & 0.4312 \\ 0.3430 & 0.2254 & 0.4316 \end{pmatrix},$$

the probability of rain on the third day is 43.12%. (Thus, you might plan your golf outing for the third day.) Further, each row of

$$\mathbf{T}^4 = \begin{pmatrix} 0.343137 & 0.225490 & 0.431372 \\ 0.343137 & 0.225490 & 0.431372 \\ 0.343137 & 0.225490 & 0.431373 \end{pmatrix}$$

gives the PDF f_i , $i = 1, 2, 3$ to five decimal places. Finally, multiplication of, say, $\mathbf{T}^3$ by the initial state row-vector $\pi^{(0)} = (0, 1, 0)$ produces the second row of $\mathbf{T}^3$, which gives the probabilities that the day three days hence will be sunny (34.3%), cloudy (22.58%), or rainy (43.12%).

the states of the system and, again, transition probabilities of going from one state to another can be defined and used to create Markov chains. Each discrete variable has a transition matrix and each continuous variable has a transition function. Methods to sample from multidimensional random variables are discussed later in [Section 6.1.5](#).

6.1.1 Ergodic Markov Chains

Every Markov chain is based either on a single distribution or on a cycle of distributions in the sense that the chain samples converge to a single PDF or to multiple PDFs. An *ergodic* Markov chain is one for which the probability of transitioning from one state to another is independent of when the transition takes place [MacKay 2007].

In other words, a Markov chain for which every time a specific state x is encountered, the probability of transitioning to any state u (for discrete variables) or to any state within du about u (for continuous variables) is the same is an ergodic Markov chain. In this case, the underlying distribution does not change and every ergodic Markov chain converges to a single distribution, often called the *stationary* or *invariant* distribution. Every ergodic Markov chain, thus, is based on a single underlying stationary PDF. Samples produced from an ergodic Markov chain of which the stationary distribution is f are in fact random samples from f .

An important feature of MCMC is that it provides a relatively simple way to sample from complex distributions. In fact, MCMC is typically used to sample multidimensional distributions and can be applied when only the shape of the distribution is known. This ability is a very powerful characteristic of MCMC because the normalizing factors in simple, joint, or conditional PDFs sometimes may be difficult to determine. Thus, MCMC allows one to sample from *unnormalized* distributions. Nevertheless, MCMC is, in essence, another way to obtain Monte Carlo samples and, as such, could have been included in Chapter 4. Indeed, MacKay [2007] treats MCMC in just this way. However, MCMC has such unique features and such a devoted user base that it deserves to be discussed as its own Monte Carlo method.

6.1.2 The Metropolis-Hastings Algorithm

Metropolis et al. [1953] recognized that one could simulate molecular motion using Markov chains.¹ This is a highly multidimensional problem that depends on the number of molecules being considered. Previously, statistical sampling had been used to estimate the *minimum* potential energy of a collection of particles. One could do this by starting with a specific spatial distribution of molecules, calculating the energy of this state, changing particle positions (either individually or all together), calculating the potential energy of the new state, and retaining whichever state had the lower energy. The minimum-energy state is identified after repeating this process until no new state with lower energy is found. However, the Metropolis group took things another step. Suppose one wants to find the *average* energy state (or some other quantity such as average pressure) of a collection of particles. The procedure employed, essentially the first MCMC calculation conducted on a digital computer, is outlined as follows.

Metropolis and colleagues (including the renowned physicist Edward Teller) started with a collection of $N = 224$ particles at specified positions in a two-dimensional cell. The potential energy of the system can be easily calculated for this, and any, known state. They assumed a completely ordered initial state, in which each row had 14 particles at equal spacing d_0 , and the even rows were displaced to the right by $d_0/2$. However, each of the N particles can move in random directions (within the plane), and thus the *expected* potential energy E requires evaluation of an integral over $2N$ dimensions (because each of the N particles can move in either of two directions) and is difficult to evaluate by normal means. The difficulty increases as N gets larger.

¹ Metropolis is the same man who co-authored the first paper [Metropolis and Ulam 1949] to formally use the term *Monte Carlo*.

The Metropolis group proposed a Monte Carlo procedure in which a new position of each of the particles was sampled and then the potential energy of the new state was calculated. An acceptance/rejection criterion then was applied. If the energy E' of the new system was lower than the original energy, the particles were moved to the new sampled positions. If the energy of the new system was higher than the original, then the particles were moved only with a certain probability; otherwise, the move was rejected and a new history was run. In their case, the acceptance probability was given by $e^{-E'/(kT)} / e^{-E/(kT)} = e^{-\Delta E/(kT)}$, where $\Delta E = E' - E$ is the change in potential energy between the two states, T is the temperature of the system, and k is Boltzmann's constant.

It is reasonable to suppose that the next state will depend only on the previous state, i.e., the next state will be independent of how the last state happened to evolve. This, then, is a Markov process, although Metropolis and colleagues did not use the term *Markov* in their paper. In their simulation, the Metropolis group ran “something less than sixteen cycles” [Metropolis et al. 1953] before beginning their estimation of the sample mean, in order to “forget” the original ordered state. It is interesting that they obtained useful results running thereafter “about forty-eight to sixty-four cycles” [Metropolis et al. 1953].

Almost two decades after Metropolis and colleagues introduced their novel simulation approach, Hastings [1970] expanded their work and formalized Markov chain Monte Carlo into the procedure now known as the *Metropolis-Hastings algorithm*. The procedure is summarized as follows for a single continuous random variable. Extension to discrete variables is straightforward and methods to deal with multidimensional variables are considered later in this chapter.

Let $f(x)$, defined on some domain $D = [a, b]$, be the stationary PDF from which random samples are desired. The Metropolis-Hastings algorithm for sampling from f proceeds as follows:

1. Choose any PDF $h(u|x)$ defined on the domain D . The PDF so chosen is called the *proposal* PDF, the *proposal* distribution, or the *proposal* function. The proposal distribution can be conditioned on the current state x but need not be.
2. Specify a starting value within D . This value should not be one that is highly unlikely for reasons discussed later. Treat this value as the starting state x_i with $i = 1$.
3. Pick a sample u_i from h using any appropriate procedure. For instance, if the CDF of h can be inverted, use the inverse CDF method.
4. Evaluate the quantity R , generally called the *Metropolis ratio* but sometimes the *Hastings ratio*,

$$R(x_i, u_i) = \frac{f(u_i) h(x_i|u_i)}{f(x_i) h(u_i|x_i)}. \quad (6.12)$$

5. Set

$$\alpha(x_i, u_i) = \min[1, R(x_i, u_i)], \quad (6.13)$$

and pick a value ρ from the uniform PDF on the unit interval, $\mathcal{U}(0, 1)$.

6. Accept u_i as the new sample with probability α or reject it with probability $1 - \alpha$, i.e., set

$$x_{i+1} = \begin{cases} u_i, & \text{if } \rho \leq \alpha \\ x_i, & \text{if } \rho > \alpha \end{cases}. \quad (6.14)$$

7. Replace i by $i+1$ and repeat steps 3 to 7 until all the desired samples have been generated.

If $R \geq 1$, the proposed state is always accepted, whereas if $R < 1$, the proposed state is accepted only with probability $\alpha = R$. In effect, the transition probability from state x to state u in MCMC is given by

$$T(x \rightarrow u) = h(u|x)\alpha(x, u). \quad (6.15)$$

There are several interesting features about the Metropolis-Hastings algorithm:

- The acceptance/rejection criterion depends only on the current state x_i , the critical requirement for the samples to form a Markov chain.
- The algorithm contains elements of importance sampling because samples are chosen from a distribution h other than f and, in essence, weights of the form f/h are formed, although they are not used in the same way as in traditional importance sampling.
- The algorithm also exhibits qualities of rejection sampling because samples are accepted or rejected based on a condition involving samples ρ selected from the uniform distribution on the unit interval. In the rejection sampling discussed in Chapter 4, the rejection criterion can be applied many times before a new sample x_{i+1} is obtained. In the Metropolis-Hastings algorithm, however, a new sample is obtained on the first application of the rejection criterion (as long as the sample is inside the correct interval); sometimes the new sample is updated and sometimes the new sample is the same as the previous sample. In fact, many samples in a row may be exactly the same!
- MCMC provides a relatively simple way to sample PDFs that are difficult to sample by other means because one samples from a proposal distribution h of one's choosing in order to generate samples from the stationary PDF f . If the stationary PDF is easy to sample directly, there is no need to employ MCMC, which requires generating at least two pseudorandom numbers to obtain each new random sample.
- Finally, the Metropolis-Hastings algorithm naturally handles samples for which the stationary PDF is zero. For instance, if a value of u_i is proposed at a location where $f(u_i) = 0$, then $R = \alpha = 0$ and the sample will be rejected. Alternatively, if, for instance, a starting value x_1 is specified where the stationary PDF is zero, then R is undefined and $\alpha = 1$ and the candidate state will be accepted, i.e., the chain will move to a new state.

Samples drawn according to the Metropolis-Hastings algorithm form a Markov chain because they are random samples that depend only on the current state and not on any other previous states. The Markov chain is ergodic if the transition probabilities do not change.

Reversibility of the transition function

The Metropolis-Hastings algorithm leads to the *detailed balance* or *reversibility* condition [Gilks et al. 1996]

$$f(x)T(x \rightarrow u) = f(u)T(u \rightarrow x). \quad (6.16)$$

To show why this reversibility property occurs, combine Eqs. (6.13) and (6.15) to obtain

$$f(x)T(x \rightarrow u) = f(x)h(u|x)\alpha(x, u) = f(x)h(u|x) \min[1, R(x, u)]. \quad (6.17)$$

With Eq. (6.12), this result can be rewritten as

$$\begin{aligned} f(x)T(x \rightarrow u) &= \min[f(x)h(u|x), R(x, u)f(x)h(u|x)] \\ &= \min[f(x)h(u|x), f(u)h(x|u)]. \end{aligned} \quad (6.18)$$

Because $\min[a, b] = \min[b, a]$ and from Eq. (6.12), the right side of this result can be developed as

$$\begin{aligned} f(x)T(x \rightarrow u) &= \min[f(u)h(x|u), f(x)h(u|x)] \\ &= \min[f(u)h(x|u), f(u)h(x|u)/R(x, u)]. \end{aligned} \quad (6.19)$$

Finally, because $1/R(x, u) = R(u, x)$, it follows that $\alpha(u, x) = \min[1, R(u, x)]$ so that Eqs. (6.19) and (6.15) give

$$\begin{aligned} f(x)T(x \rightarrow u) &= f(u)h(x|u) \min[1, 1/R(x, u)] \\ &= f(u)h(x|u)\alpha(u, x) = f(u)T(u \rightarrow x), \end{aligned} \quad (6.20)$$

which validates Eq. (6.16).

Distribution of Metropolis-Hasting samples

Now consider why samples formed according to the Metropolis-Hastings algorithm are samples from the stationary PDF $f(x)$. As before, assume the PDF $f(x)$ is defined on the domain $D = [a, b]$ and further let D^+ specify the domain over which $f(x) > 0$. Next, assume that the starting point is specified within D^+ . In general, the transition probability from state x to state x may be nonzero, i.e., it may be possible to transition to the same state. Thus, define the function

$$p(x, u) = \begin{cases} h(u|x)\alpha, & \text{if } u \neq x \\ 0, & \text{if } u = x \end{cases}, \quad (6.21)$$

which, in general, is not a PDF and, thus, $\int_{D^+} p(x, u) du$ may not be unity. Let

$$r(x) = 1 - \int_{D^+} p(x, u) du, \quad (6.22)$$

which is the probability that a state x remains at x . With these definitions the transition function $T(x \rightarrow u)$ can be written as the sum of two pieces, one giving the transition probability of moving to a different state and the other giving the probability of staying in the same state, namely,

$$T(x \rightarrow u)du = p(x, u)du + r(x)\delta_x(du), \quad (6.23)$$

where

$$\delta_x(du) = \begin{cases} 1, & \text{if } du \text{ contains } x \\ 0, & \text{otherwise.} \end{cases} \quad (6.24)$$

Now, the probability P of moving from any state x to any state within a subset S of D^+ can be expressed as

$$\begin{aligned} P &= \int_{D^+} \int_S f(x)T(x \rightarrow u) du dx \\ &= \int_{D^+} f(x) \left\{ \int_S p(x, u) du \right\} dx + \int_{D^+} f(x)r(x)\delta_x(S) du \\ &= \int_S \left\{ \int_{D^+} f(x)p(x, u) dx \right\} du + \int_S f(x)r(x) dx. \end{aligned} \quad (6.25)$$

Because of the way $p(x, y)$ was defined, the reversibility condition implies that

$$f(x)p(x, u) = f(u)p(u, x) \quad (6.26)$$

and, thus, with the definition of Eq. (6.22),

$$\begin{aligned} P &= \int_S \left\{ \int_{D^+} f(u)p(u, x) dx \right\} du + \int_S f(x)r(x) dx \\ &= \int_S [1 - r(u)]f(u) du + \int_S f(x)r(x) dx \\ &= \int_S f(u) du. \end{aligned} \quad (6.27)$$

This result demonstrates that the chain states are samples from the stationary PDF f .

Hence, the reversibility condition of Eq. (6.16) is sufficient to show that the samples produced by the Metropolis-Hastings algorithm are samples from the stationary PDF. In fact, any α for which the reversibility condition is satisfied generates samples from the stationary distribution. A more complete discussion concerning the fact

that Markov chain samples are samples from the stationary PDF is given by Tierney [1994].

Example 6.2 Evaluate a One-Dimensional Integral Using MCMC

Consider the integral given by

$$I = \int_0^{\infty} (x - \alpha\beta)^2 f(x) dx,$$

where $f(x)$ is the gamma PDF (see Appendix A.1.3)

$$f(x) = \frac{1}{\beta\Gamma(\alpha)} \left(\frac{x}{\alpha}\right)^{\alpha-1} e^{-x/\beta}, \quad x \geq 0,$$

with $\Gamma(\alpha)$, the gamma function. This integral is of the form of Eq. (6.1) with $z(x) = (x - \alpha\beta)^2$ and defines the variance of the gamma PDF, which is known to be $\sigma^2 = \alpha\beta^2$. Compare the use of MCMC and straightforward sampling to estimate this integral for the case $\alpha = 2$ and $\beta = 1$, for which the correct answer is $I = 2$.

One MCMC approach would be to select the proposal PDF

$$h(u) = 0.5e^{-0.5u}, \quad u \geq 0,$$

and set $x_i = 1$ for $i = 1$. Then sample from the proposal PDF, obtaining candidate samples $u_i = -2\ln(\rho)$. Form the Metropolis ratio

$$R = \frac{f(u_i)h(x_i)}{f(x_i)h(u_i)} = \left(\frac{u_i}{x_i}\right)^{\alpha-1} e^{(u_i-x_i)/\beta} e^{-0.5(x_i-u_i)}$$

and update the samples as follows:

$$x_{i+1} = \begin{cases} u_i, & \text{if } \rho \leq \min[1, R] \\ x_i, & \text{if } \rho > \min[1, R] \end{cases}.$$

These samples are then used to form the estimate

$$\bar{I}_N = \frac{1}{N} \sum_{i=1}^N (x_i - \alpha\beta)^2 = \frac{1}{N} \sum_{i=1}^N (x_i - 2)^2.$$

Alternatively, one can sample the gamma distribution using the multistep process described in Appendix A.1.3. Results for four values of N are shown in Table 6.1 for both approaches.

Both results appear to converge to the correct value of $I = 2$, suggesting that MCMC is indeed a valid alternative way to obtain Monte Carlo samples.

Table 6.1 Estimates of the variance of the gamma PDF using MCMC and direct sampling of $f(x)$. Here, $\alpha = 2$, $\beta = 1$, and the exact value is $I = 2$.

N	MCMC	Direct Sampling
	$\bar{I}_N \pm \sigma(\bar{I}_N)$	$\bar{I}_N \pm \sigma(\bar{I}_N)$
10^4	1.946 ± 0.051	2.060 ± 0.047
10^5	2.004 ± 0.014	1.985 ± 0.014
10^6	1.997 ± 0.0047	2.003 ± 0.0045
10^7	1.999 ± 0.0014	1.999 ± 0.0014

6.1.3 The Myth of Burn-in

It is common practice to express the MCMC estimate of a population mean $\langle z \rangle = \int_{-\infty}^{\infty} z(\mathbf{x})f(\mathbf{x})d\mathbf{x}$ in the form

$$\bar{z} = \frac{1}{N} \sum_{i=m+1}^{m+N} z(\mathbf{x}_i), \quad (6.28)$$

where the $\mathbf{x}_i$ are sampled from $f(\mathbf{x})$ and m , called the *burn-in length*, is the number of samples x_i , $i = 1, 2, \dots, m$ at the beginning of the chain that are discarded. For a process governed by a stable PDF that does not change while samples are being drawn, no burn-in is necessary, i.e., $m = 0$ is sufficient. [Metropolis et al. \[1953\]](#) used burn-in only because they started from a highly unlikely state and they reasoned that several wasted Markov steps would result in a more typical state from which to start actual scoring. However, this was not strictly required because a long enough chain will generate samples having the proper distribution of the governing PDF. It is true that starting from states that are highly unlikely, as Metropolis and colleagues originally did, may lead to repeated samples of that state. However, samples will be distributed with the correct PDF, without burn-in, *if* a sufficiently large number of histories is run.

If MCMC is used to sample from a fixed distribution, then the samples obtained by application of the Metropolis-Hastings algorithm will generate an ergodic Markov chain and no burn-in is required. Only if the underlying PDF changes during sampling will the chain of random samples generated by MCMC not be ergodic. But even in this case, burn-in may not be necessary. For instance, consider that a process is being modeled in which a change of phase occurs and the probabilities of certain outcomes change as a result of the phase change. Samples generated by the Metropolis-Hastings

algorithm yield valid samples for the entire process. Only if samples are desired from the PDF that results after phase change would burn-in be useful. Whenever it is easy to start from an initial state that is not highly unlikely, then this should be done and no burn-in is required. It is more important that N be large enough so that the samples assume their proper frequency than it is that burn-in be used. Nevertheless, the myth that burn-in is required, even for ergodic Markov chains, somehow survives.

6.1.4 The Proposal Distribution

There is a presumed advantage to using a proposal PDF that is conditioned on the current value of x ; this has to do with reducing the number of duplicate samples. However, the proposal distribution does not have to depend on x and can have the simple form $h(u)$. Application of MCMC with a proposal distribution that is independent of x is discussed by Tierney [1994] and is commonly called the *independence sampler*.

In principle, the choice of the proposal distribution is quite general, but in practice one wants a proposal distribution of which the shape is similar to the true PDF. Consider, for purposes of illustration, the case in which one seeks to generate samples from the simple PDF $f(x) = 0.2$, $0 \leq x \leq 5$ and uses an exponential proposal of the form $h(u) = e^{-u}$. The first 1000 sampled values of x_i generated by the Metropolis-Hastings algorithm for this case are shown in Fig. 6.1. Here, large values of x_i tend to repeat, often many times, because samples of u_i are likely to be smaller than a large x_i so that $h(x_i)/h(u_i) \ll 1$. Such repeats of large values of x_i are necessary, however, in order for the distribution of the $\{x_i\}$ to approach that of f . The distribution of the x_i

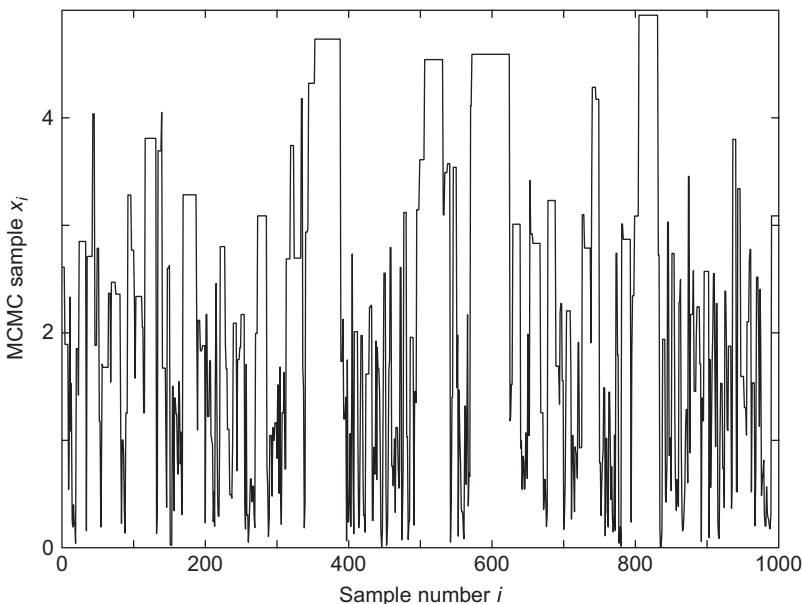


Figure 6.1 The first 1000 samples of x_i for the MCMC sampling from the proposal $h(u) = e^{-u}$. Notice that large values of x_i often are successively repeated, while small values of x_i are seldom repeated.

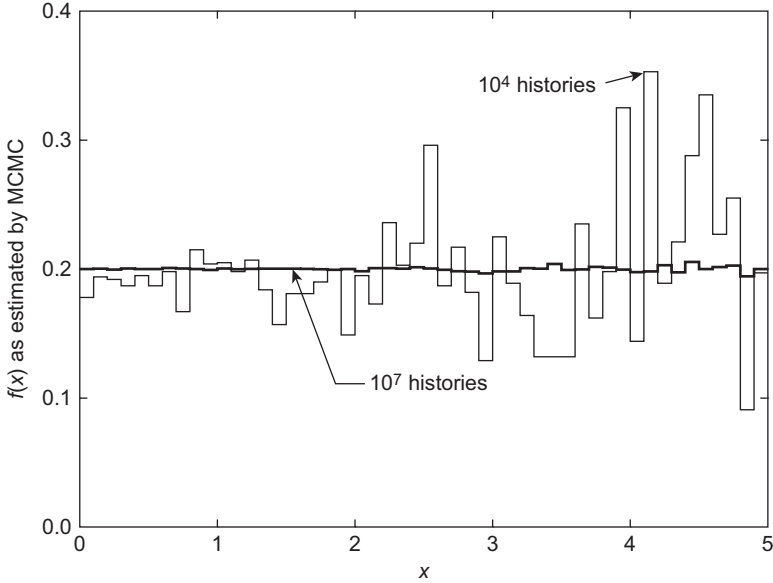


Figure 6.2 The distribution of the $\{x_i\}$ obtained by MCMC using the proposal $h(u) = e^{-u}$. A large number of samples is required to reproduce the distribution $f(x) = 1/D = 0.2$, especially for large values of x . This slow convergence for large x is a consequence of the low probability of obtaining a large sample value from the proposal distribution, which would normally be a poor choice but is useful for demonstration purposes.

samples are shown in Fig. 6.2 for two values of N . For $N = 10^4$, there is significant variation from the true PDF $f(x) = 0.2$, especially for large x . However, the samples approximate the true PDF remarkably well for $N = 10^7$ histories.

It seems obvious that duplication of the sampled points should be minimized. The term *mixing* is used to indicate the rate at which the sample space is sampled; good or rapid mixing means that the space of possible states is quickly sampled and few states are repeated. A good proposal function exhibits good mixing. This can be achieved by choosing a proposal distribution h that approximates well the shape of the stationary distribution f , so that R is generally near unity. Application of MCMC where $h = f$ results in perfect mixing because $R = 1$ always. However, if f can be sampled easily by other means, then MCMC is unnecessary.

Samples over infinite domains

There are a number of distributions that are defined over the infinite interval $(-\infty, \infty)$. Methods for obtaining samples from some of these distributions exist (see Appendix A), but MCMC can also be used. For instance, the proposal distribution given by

$$h(u) = (\lambda/2)e^{-\lambda|u|} \quad (6.29)$$

can be used to obtain candidate values of x drawn from any PDF defined on $(-\infty, \infty)$. One can sample a pseudorandom number ρ_i uniformly from the unit interval and then

for $\rho_i < 0.5$, the sampled value of u_i is negative and is given by

$$u_i = \ln(2\rho_i)/\lambda \quad (6.30)$$

for $\rho_i = 0.5$, the sampled value of u_i is zero, and for $\rho > 0.5$, the sampled value of u_i is positive and is given by

$$u_i = -\ln(2 - 2\rho_i)/\lambda. \quad (6.31)$$

This procedure is illustrated next.

Assume that one wishes to generate samples from the stationary PDF f that is normally distributed with mean zero and standard deviation 2.5, i.e.,

$$f(x) = \mathcal{N}(0, 2.5) = \frac{e^{-x^2/[2(2.5)^2]}}{2.5\sqrt{2\pi}}. \quad (6.32)$$

Consider the two proposal distributions

$$h_1(u) = \mathcal{N}(0, 1) = \frac{e^{-u^2/2}}{\sqrt{2\pi}} \quad (6.33)$$

and

$$h_2(u) = 0.2e^{-0.4|u|}. \quad (6.34)$$

The PDFs f , h_1 , and h_2 are shown in Fig. 6.3. Proposal distribution h_1 is of the same form as f but is more peaked; proposal distribution h_2 has longer tails than either f or

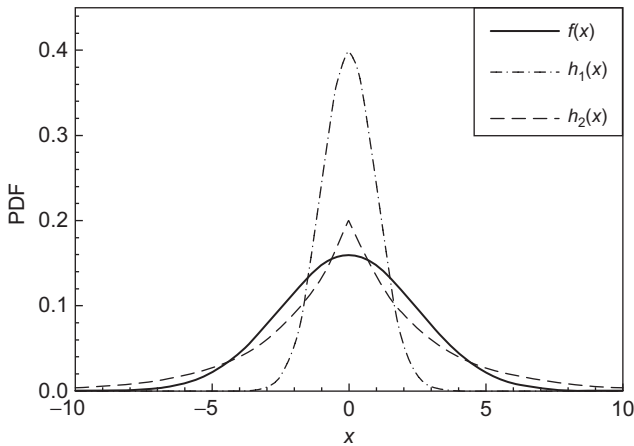


Figure 6.3 The PDF $f(x) = \mathcal{N}(0, 2.5)$ (solid line) and two proposal distributions h_1 (dot-dash) and h_2 (long dashes).

h_1 but conforms more closely to the shape of f than does h_2 over the important range $-6 \leq x \leq 6$, the range where most of the samples are generated. The Metropolis-Hastings algorithm was applied to generate samples from f using each proposal distribution and a starting value of $x_1 = -3.0$. These samples were then collected into 80 bins from -10.125 to 10.125 , for both the cases $N = 10^3$ and $N = 10^5$. The results are shown, with the true PDF f , in Figs. 6.4 and 6.5. Note that $N = 10^3$ samples is

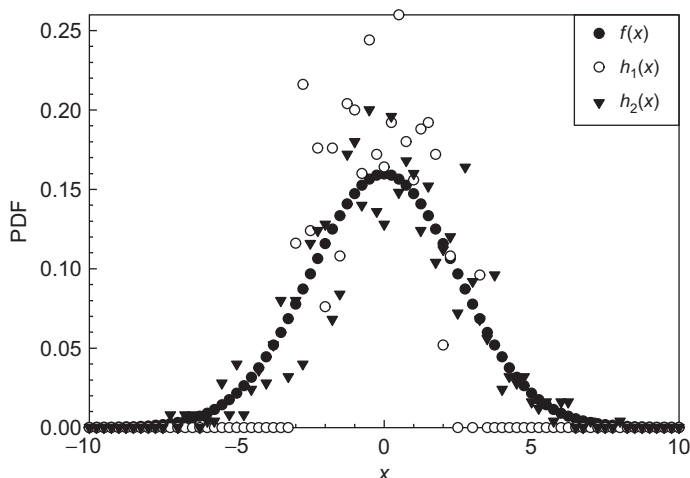


Figure 6.4 Histograms of 1000 samples from f using the two proposal distributions h_1 (open circles) and h_2 (triangles), with $m = 0$. Filled circles are used to show the actual PDF.

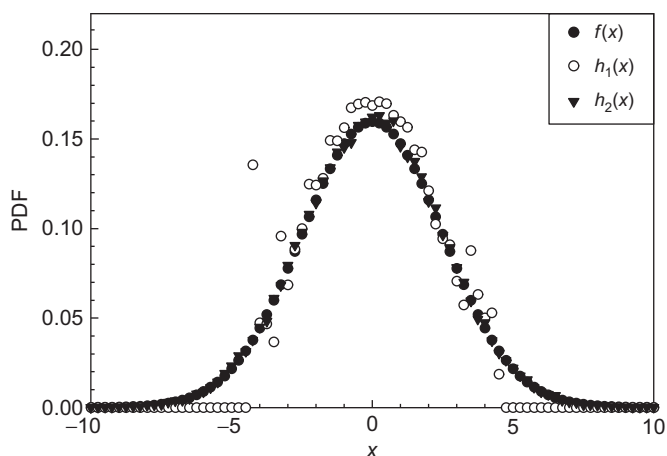


Figure 6.5 Histograms of 100,000 samples from f using the two proposal distributions h_1 (open circles) and h_2 (triangles), with $m = 0$. Filled circles are used to show the actual PDF.

insufficient for either proposal function to reproduce the desired distribution very well, although h_1 does a poorer job than does h_2 . On the other hand, after 10^5 histories, samples picked using both proposal functions generate samples distributed approximately according to the desired PDF. However, of the two proposals, h_2 , which more closely resembles the stationary distribution, does a considerably better job, especially for the tails of $f(x)$. This demonstrates that it is best to use a proposal that resembles the stationary distribution. However, if *enough* histories are drawn, essentially any proposal, defined on the proper interval, will suffice. But there is no way to estimate a priori what constitutes enough histories. In practice, N usually is increased until fluctuations in $\bar{z}$ are within acceptable limits.

6.1.5 Multidimensional Sampling

In highly multidimensional cases, sampling by traditional methods can be very problematic. However, the extension of MCMC to many variables is conceptually straightforward. Let $\mathbf{x} = (x_1, x_2, \dots, x_n)$ be a collection of n random variables with joint PDF f . Often MCMC is applied in such cases to obtain random samples $\mathbf{x}^{(i)}$, $i = 1, 2, \dots, N$, where N is the number of histories. Two basic methods are commonly employed. The first is to sample all components of the multidimensional PDF $f(\mathbf{x})$ on every history, using a multidimensional candidate distribution $h(\mathbf{x})$ and applying the Metropolis-Hastings algorithm. This approach is sometimes called the *global* approach (e.g., see Neal [1993]). Thus, all n random variables are sampled on each history, so that $\mathbf{x}^{(i)} = (x_1^{(i)}, x_2^{(i)}, \dots, x_n^{(i)})$.

The second (*local*) method is to select one component at a time and sample it from a candidate distribution using the Metropolis-Hastings method. In this case, the proposal distribution is a one-dimensional PDF. For instance, to sample the j th component one would sample from a proposal PDF of the form $h_{x_j}(x_j)$. In this case, if variable j is sampled on history i , then

$$\mathbf{x}^{(i)} = (x_1^{(i-1)}, x_2^{(i-1)}, \dots, x_{j-1}^{(i-1)}, x_j^{(i)}, x_{j+1}^{(i-1)}, \dots, x_{n-1}^{(i-1)}, x_n^{(i-1)}). \quad (6.35)$$

Of course, the local approach requires that a proposal $h_{x_j}(x_j)$ be given for each component. It is interesting that the determination of which component or variable to sample can be determined either sequentially or at random. This procedure gives rise to the concept of a “random walk,” because a move is made in one direction only in a multidimensional space at each sampling point. Twelve steps of a random walk starting at $\mathbf{x}^{(i)}$ in a three-dimensional space are depicted in Fig. 6.6 in which the component direction is selected sequentially (x_1 first then x_2 next, and so forth). Figure 6.7 depicts 12 steps in which the component direction is selected at random. Either method is valid in that the entire space eventually will be sampled if enough steps are taken. Figure 6.8 shows 12 steps in which all three components are updated during each step.

The great attractiveness of MCMC is that multidimensional PDFs need not be sampled directly; it is only necessary to evaluate these PDFs at candidate values of the

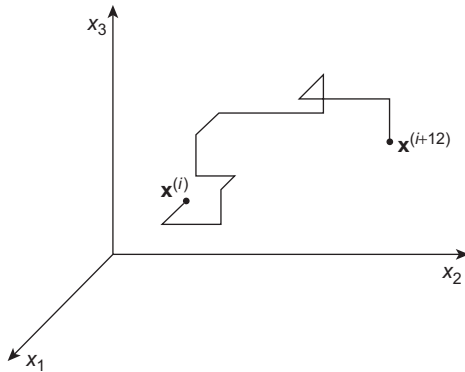


Figure 6.6 Twelve hypothetical random-walk steps taken by sequentially sampling x_1 , then x_2 , then x_3 , then x_1 , and so forth.

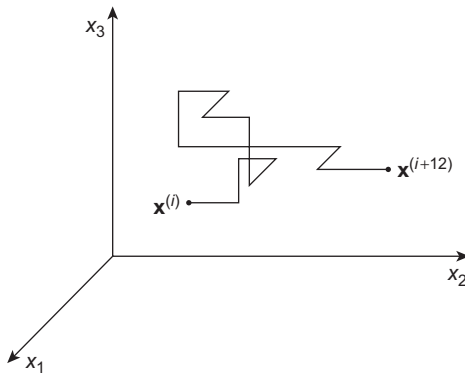


Figure 6.7 Twelve hypothetical random-walk steps taken by sampling components of $\mathbf{x}$ at random.

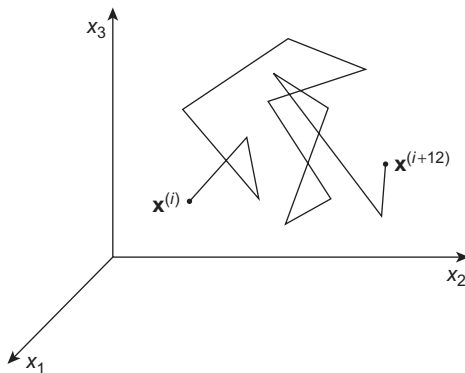


Figure 6.8 Twelve hypothetical random-walk steps taken by global sampling.

random vector. The Metropolis-Hastings algorithm for an n -dimensional PDF $f(\mathbf{x})$ proceeds as follows:

1. Choose a proposal distribution, $h(\mathbf{u}|\mathbf{x})$, which can be easy to sample. Recall that this distribution need not be conditioned on $\mathbf{x}$, but the case of a conditional PDF is treated here.
2. Choose an initial value of $\mathbf{x}_i$, for $i = 1$. All components of $\mathbf{x}$ should have values that are not highly unlikely, so as to avoid repeating samples early in the process.
3. Sample from h to obtain a new sample $\mathbf{u}_i$. In global sampling, all components u_j , $j = 1, 2, \dots, n$ are sampled from h on each history i , $i = 1, 2, \dots, N$ and $\mathbf{u}_i = (u_1, u_2, \dots, u_n)_i$. In local sampling, one component at a time is updated, either sequentially or at random. Thus, if the j th component is sampled, $\mathbf{u}_i = (x_1, x_2, \dots, x_{j-1}, u_j, x_{j+1}, \dots, x_n)_i$.
4. Evaluate the Metropolis ratio

$$R = \frac{f(\mathbf{u}_i) h(\mathbf{x}_i|\mathbf{u}_i)}{f(\mathbf{x}_i) h(\mathbf{u}_i|\mathbf{x}_i)}. \quad (6.36)$$

5. Pick a pseudorandom number ρ from the unit rectangular distribution and update the vector $\mathbf{x}$ as follows:

$$\mathbf{x}_{i+1} = \begin{cases} \mathbf{u}_i, & \text{if } \rho \leq \min[1, R] \\ \mathbf{x}_i, & \text{if } \rho > \min[1, R] \end{cases}. \quad (6.37)$$

6. Use this vector to score this history, increment i by one, and repeat the process until all histories have been completed.

A two-dimensional case is considered in [Example 6.3](#).

Example 6.3 A Modified Buffon's Needle Problem

Consider a modified Buffon's needle game in which needles of length L are dropped by a biased machine. In the original Buffon's needle problem, the PDF for the position of the needle endpoints is $f(x) = 1/D$, $0 \leq x < D$, the PDF for the orientation angle is $g(\theta) = 1/\pi$, $0 \leq \theta < \pi$, and P_{cut} is given by Eq. (1.7). The biased machine, however, drops needles such that the endpoints x and the orientation angles θ are distributed, within a unit cell, according to a bivariate normal distribution (see Appendix A.3.1), i.e.,

$$f(x, \theta) = \frac{K}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{z(x, \theta)}{2(1-\rho^2)}\right], \quad 0 \leq x < D, \quad 0 \leq \theta < \pi,$$

where K is a normalization constant,

$$z(x, \theta) = \frac{(x - \mu_1)^2}{\sigma_1^2} - \frac{2\rho(x - \mu_1)(\theta - \mu_2)}{\sigma_1\sigma_2} + \frac{(\theta - \mu_2)^2}{\sigma_2^2},$$

and

$$\rho = \frac{\sigma_{12}}{\sigma_1\sigma_2}.$$

Employing an analysis similar to that taken in Example 1.1, the cut probability for this machine dropping needles on a grid with $D \geq L$ can be expressed as

$$P_{cut} = \int_0^\pi \int_0^D \frac{L \sin \theta}{D} f(x, \theta) dx d\theta.$$

How might one estimate this cut probability for a given set of parameters?

Consider the case where $D = 5$, $L = 2$, $\mu_1 = 3$, $\mu_2 = 1$, $\sigma_1 = 1$, $\sigma_2 = 1$, and $\sigma_{12} = 0.1$. One could drop n needles into the machine and observe k crossings on the grid below, and approximate P_{cut} as k/n . However, it is much more efficient to run an experiment on a computer using MCMC.

Choose as proposal distributions functions that are easy to sample, say $h_1(u) = 1/D$ and $h_2(v) = 1/\pi$. Then, sample u and v from these proposal functions. In this case, the Metropolis ratio is simply

$$R = \frac{f(u, v)}{f(x, \theta)}$$

because the proposals are both constants. After running 10^6 histories an estimate of $P_{cut} = 0.2837 \pm 0.0001$ was obtained (a value statistically slightly higher than for uniform drops). Notice that the two-dimensional integral for P_{cut} is evaluated without having to sample from the bivariate normal distribution or having to determine the normalization constant K . All that is required is that the ratio of bivariate normals be evaluated at both current and candidate points sampled from the simple proposal functions.

Results generated by global MCMC sampling from a multidimensional distribution are shown in Figs. 6.9 to 6.11, in which a bivariate normal PDF (see Appendix A.3.1) is sampled by MCMC for three different values of N . The large spikes in the tails evident for 10^3 histories are “softened” by 10^5 histories and almost completely removed by $N = 10^7$ histories.

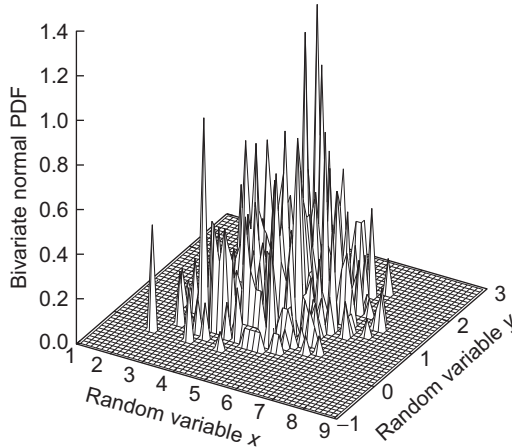


Figure 6.9 Bivariate normal PDF sampled using MCMC after 10^3 histories.

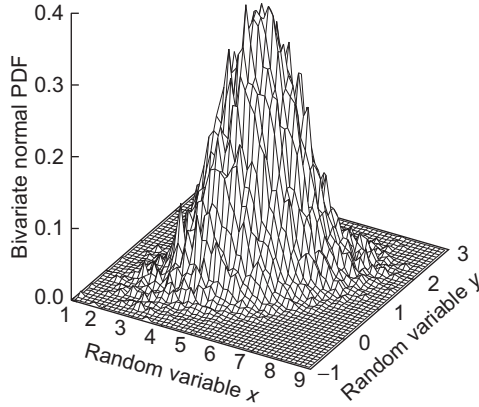


Figure 6.10 Bivariate normal PDF sampled using MCMC after 10^5 histories.

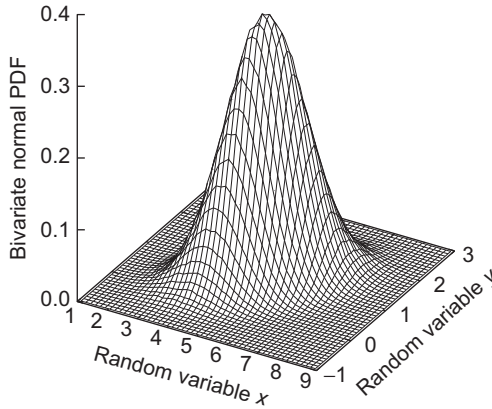


Figure 6.11 Bivariate normal PDF sampled using MCMC after 10^7 histories.

6.1.6 The Gibbs Sampler

The “Gibbs sampler” is a special case of Metropolis-Hastings in which the single-component proposal distribution is the conditional PDF for that component given the other components. The notation $\mathbf{x}_{[-j]}$ is used to represent the vector $\mathbf{x}$ with the j th component missing, i.e.,

$$\mathbf{x}_{[-j]} = (x_1, x_2, \dots, x_{j-1}, x_{j+1}, x_{j+2}, \dots, x_n). \quad (6.38)$$

Then $f(x_j | \mathbf{x}_{[-j]})$ is the conditional PDF for x_j given $\mathbf{x}_{[-j]}$. For the Gibbs sampler it is always true that $\alpha = 1$ and, thus, no samples are ever rejected.

Consider an n -dimensional random variable $\mathbf{x}$, for which the conditional PDFs $f(x_j|\mathbf{x}_{[-j]})$, $j = 1, 2, \dots, n$ are known. Local Gibbs sampling after i histories would proceed by selecting a j , either sequentially or at random, and then sampling the j th component of $\mathbf{x}$ for history $i + 1$, i.e., $x_j^{(i+1)}$, from $f(x_j|\mathbf{x}_{[-j]}^{(i)})$. However, global sampling is preferred because it is straightforward to sample all n components, each from its conditional PDF.

One could proceed by sampling $x_1^{(i+1)}$, from $f(x_1|\mathbf{x}_{[-1]}^{(i)})$, then $x_2^{(i+1)}$ from $f(x_2|\mathbf{x}_{[-2]}^{(i)})$, and so on, for $j = 3, 4, \dots, n$. However, it is better to use the most recent sampled value of each variable, so that sampling would proceed as follows:

1. Sample $x_1^{(i+1)}$ from $f(x_1|\mathbf{x}_{[-1]}^{(i)})$.
2. Next, sample $x_2^{(i+1)}$ from $f(x_2|x_1^{(i+1)}, x_3^{(i)}, x_4^{(i)}, \dots, x_n^{(i)})$.
3. Then, sample $x_3^{(i+1)}$ from $f(x_3|x_1^{(i+1)}, x_2^{(i+1)}, x_4^{(i)}, \dots, x_n^{(i)})$.
4. Continue in a similar manner for each $x_j > 3$ until you sample $x_n^{(i+1)}$ from $f(x_n|\mathbf{x}_{[-n]}^{(i+1)})$.

This procedure then provides an updated sample for $\mathbf{x}^{i+1}$, which is used in the MCMC estimator.

Geman and Geman [1984] were concerned with Gibbs distributions when their work exposed this sampling technique to wide attention and, thus, the terminology “Gibbs sampling” survives. The Gibbs sampler is a staple of MCMC and often is used when all of the conditional PDFs are known.

6.2 Brief Review of Probability Concepts

MCMC often is applied in statistical inference. Statistical inference in recent times has become associated with Bayesian analysis. Thus, a brief review of probability theory and Bayes theorem are given here and the potential use of MCMC in statistical inference and related applications, such as decision theory, is briefly discussed.

Bayesian analysis is in stark contrast to the frequentist approach. Frequentists interpret probability in terms of the frequency with which an event occurs among a number of trials. To a frequentist, the probability of an event is the relative frequency of occurrence. For example, if a die is thrown 100 times and a three is obtained 20 times, a frequentist is apt to say that the best estimate of the probability of throwing a three is 20/100 or 20%. Proponents of the Bayesian approach interpret probability as a measure of *likelihood* of an outcome in the absence of complete knowledge. To a Bayesian, probability is an inherent quantity, which may not be known but can be estimated. Bayesians often talk of probability as a measure of their degree of certainty about their knowledge, with 0 being completely uncertain and 1 being completely certain. A Bayesian can take the same data (e.g., 20 threes out of 100 throws) and produce a different result than the frequentist. The debate between frequentists and Bayesians rages in the literature and in public and private forums. Although both approaches are widely used, it appears that the cadre of Bayesians is growing at the expense of the ranks of the frequentists.

Let $\text{Prob}\{A\}$ represent the probability of A , where A may be an outcome (such as “heads or tails”), a conjecture (such as “it will rain today in Kansas City”), or an hypothesis (such as “the moon resulted from the impact over 4 billion years ago of a body about one-third the size of the current earth with a larger body that was the proto-earth”). It is useful to express the fact that probability models generally are based on some assumptions or other information, denoted here by the symbol O (denoting “other”). Thus, the expression for the probability of A more generally can be written in the conditional form $\text{Prob}\{A|O\}$, which in words reads “the probability of event A given other information O .” It is important to recognize that models and hypotheses are based on assumptions or other information and the notation “ $|O$ ” formally reminds one of this.

The symbol $\bar{A}$ is interpreted as “not A ,” i.e., the complement of A such that $\text{Prob}\{\bar{A}|O\} = 1 - \text{Prob}\{A|O\}$ or, as it is more often presented,

$$\text{Prob}\{A|O\} + \text{Prob}\{\bar{A}|O\} = 1. \quad (6.39)$$

Further, $\text{Prob}\{A + B|O\}$ is the probability of either A or B given O , and $\text{Prob}\{AB|O\}$ is the probability of both A and B given O . If A and B are mutually exclusive and collectively exhaustive events (e.g., A stands for “heads,” B for “tails,” and O assumes an unbiased throw of a fair coin), then $\text{Prob}\{A + B\} = 1$ (the result must be one or the other) and $\text{Prob}\{AB\} = 0$ (you cannot get both a head and a tail on one throw).

Now consider two basic tenets of probability theory. First, the product rule

$$\text{Prob}\{AB|O\} = \text{Prob}\{A|BO\}\text{Prob}\{B|O\} \quad (6.40)$$

states that the probability that both A and B are true, given O , is the product of

- the *conditional probability* that A is true given both B and O , and
- the *marginal probability* that B is true, given O .

The second tenet is the commutativity property

$$\text{Prob}\{AB|O\} = \text{Prob}\{BA|O\}, \quad (6.41)$$

which holds because AB and BA are identical results (AB means both A and B and BA means both B and A).

6.3 Bayes Theorem

An equivalent form of Eq. (6.40) for $\text{Prob}\{BA|O\}$ is

$$\text{Prob}\{BA|O\} = \text{Prob}\{B|AO\}\text{Prob}\{A|O\}. \quad (6.42)$$

It follows from Eqs. (6.40), (6.41), and (6.42) that

$$\text{Prob}\{A|BO\}\text{Prob}\{B|O\} = \text{Prob}\{B|AO\}\text{Prob}\{A|O\}, \quad (6.43)$$

which can be rearranged into the form known as Bayes theorem, namely,

$$\text{Prob}\{A|BO\} = \text{Prob}\{A|O\} \frac{\text{Prob}\{B|AO\}}{\text{Prob}\{B|O\}}. \quad (6.44)$$

Thus, Bayes theorem results from the basic commutativity and product rules of probability theory.

Bayes theorem often is written in terms of PDFs. The derivation is slightly different, because PDFs are not probabilities but rather probability densities. Consider the case of two random variables x and y . The marginal PDF for x given y was given in Chapter 2 (see Eq. (2.38)) as

$$f(x|y) = \frac{f(x, y)}{f_y(y)}, \quad (6.45)$$

where $f(x, y)$ is the joint PDF and $f_y(y)$ is the marginal PDF of y . But Eq. (2.38) indicates that the joint PDF can be expressed as

$$f(x, y) = f(y|x)f_x(x). \quad (6.46)$$

Substitution of this result into Eq. (6.45) results in a form of Bayes theorem for two random variables, namely,

$$f(x|y) = f_x(x) \frac{f(y|x)}{f_y(y)}. \quad (6.47)$$

In this form, Bayes theorem relates the marginal and conditional distributions of $f(x, y)$, a result embarrassing to frequentists. To Bayesians, however, x and y need not be repeatable outcomes but may be propositions or hypotheses, and Bayes theorem becomes a tool for making consistent inferences. Often $f_x(x)$, also called the *prior distribution*, can be based on experience, hypothesis, or subjective opinion about some property of a system. For example, let x represent the probability a particular diesel-fired emergency electric generator will start when needed, and let y represent the observation that the generator in question has started as required in all N previous tests. The prior distribution $f_x(x)$ represents the probability that similar generators start upon demand, and may be based on data from many similar generators or even upon informed subjective opinion. Then, Bayes theorem is used to combine this prior “guess” with the observed data for the generator in question to obtain a refined estimate of the probability $f(x|y)$ that this generator will start the next time when needed.

More generally, consider a model of a process or system with any number n of parameters θ and a limited set of observations or “data” $\mathbf{D}$. Now, replace the variable x in Eq. (6.47) with the parameter vector θ and the variable y with the observed data $\mathbf{D}$.

Then Bayes theorem can be expressed in the general form

$$f(\boldsymbol{\theta}|\mathbf{D}) = g(\boldsymbol{\theta}) \frac{\ell(\mathbf{D}|\boldsymbol{\theta})}{M}, \quad (6.48)$$

where the PDF $f(\boldsymbol{\theta}|\mathbf{D})$ is called the *posterior distribution*, the PDF $g(\boldsymbol{\theta})$ is called the *prior distribution*, and $\ell(\mathbf{D}|\boldsymbol{\theta})$ is called the *likelihood function*. The quantity M has a variety of names and is variously called the global likelihood, the *marginal distribution*, or the *evidence*. The global likelihood is just the probability of obtaining the observed data and is a constant. This constant, which may or may not be known, can be expressed as

$$M = \int_{\Theta} \ell(\mathbf{D}|\boldsymbol{\theta}) g(\boldsymbol{\theta}) d\boldsymbol{\theta}, \quad (6.49)$$

where Θ is the range of possible values of $\boldsymbol{\theta}$. Thus, Bayes theorem states that the posterior distribution (the conditional probability of $\boldsymbol{\theta}$ given the data) is proportional to the product of the prior distribution (the PDF $g(\boldsymbol{\theta})$) and a likelihood function (the probability of obtaining the data, given $\boldsymbol{\theta}$). Bayes theorem is a result of classic statistics and agrees with frequentist conclusions whenever the prior distribution is known, as is demonstrated in [Example 6.4](#).

Example 6.4 A Classical Statistical Analysis

Consider three bowls containing 30 candies each. The first bowl has 10 red, 10 yellow, and 10 blue candies, the second bowl contains 5 red, 15 yellow, and 10 blue candies, and the third bowl contains 10 red, 15 yellow, and 5 blue candies. Thus, there are 25 red, 40 yellow, and 25 blue candies distributed among the three bowls. A young child goes into the room with the three bowls and comes back with a yellow candy. One asks the simple question “What are the probabilities that the child got the candy from bowl one, from bowl two, and from bowl three”? Let us use the shortcut notation that $\text{Prob}\{i|c\}$ is interpreted as the conditional probability that bowl i is chosen given that color c was selected, that $\text{Prob}\{c|i\}$ is interpreted as the conditional probability that color c was selected given bowl i , and that $\text{Prob}\{i\}$ is interpreted as the probability that bowl i is selected. A frequentist might divide the number of yellow candies in each bowl by the total of all yellow candies, which leads to

$$\begin{aligned} \text{Prob}\{1|\text{yellow}\} &= \frac{10}{40} = 0.25 \\ \text{Prob}\{2|\text{yellow}\} &= \frac{15}{40} = 0.375 \\ \text{Prob}\{3|\text{yellow}\} &= \frac{15}{40} = 0.375. \end{aligned} \quad (6.50)$$

Bayes theorem provides another way to make this estimate, based on the datum of a single candy. The probability of randomly selecting each bowl is $1/3$, i.e., the prior distributions are known for each bowl and are given by $\text{Prob}\{1\} = \text{Prob}\{2\} = \text{Prob}\{3\} = 1/3$.

However, the normalized likelihood for bowl n is $\text{Prob}\{\text{yellow}|n\}/M$, where M is the global likelihood

$$\begin{aligned} M &= \text{Prob}\{\text{yellow}|1\}/3 + \text{Prob}\{\text{yellow}|2\}/3 + \text{Prob}\{\text{yellow}|3\}/3 \\ &= \frac{10/30 + 15/30 + 15/30}{3} = 0.4444. \end{aligned}$$

The Bayesian estimate of the probability that the child retrieved the candy from bowl one is

$$\text{Prob}\{1|\text{yellow}\} = \text{Prob}\{1\}\text{Prob}\{\text{yellow}|1\}/M = \frac{(1/3)(10/30)}{0.4444} = 0.25,$$

from bowl two is

$$\text{Prob}\{2|\text{yellow}\} = \text{Prob}\{2\}\text{Prob}\{\text{yellow}|2\}/M = \frac{(1/3)(15/30)}{0.4444} = 0.375,$$

and from bowl three is

$$\text{Prob}\{3|\text{yellow}\} = \text{Prob}\{3\}\text{Prob}\{\text{yellow}|3\}/M = \frac{(1/3)(15/30)}{0.4444} = 0.375.$$

The conclusion, using Bayes theorem in its classic formalism, agrees with the strict frequentist approach because the prior distribution is known.

Several things about Bayes theorem are noteworthy:

- First, both the parameters and the data are treated as random variables; the posterior f is the PDF for the parameters and the likelihood ℓ is proportional to the PDF for the data.
- Second, [Eq. \(6.48\)](#) holds for any number of random variables. It also holds whether the variables are discrete or continuous.
- Third, the data can be values either of a single measured quantity (such as temperature measured at different times) or of many quantities (such as temperature, pressure, and humidity measured at different times).
- Fourth, [Eq. \(6.49\)](#) is in the form of an expected value and expresses the fact that the probability a certain data set is obtained is the population mean of getting the data given the model parameters averaged over the range of the parameters.

Thus, estimation of the posterior distribution can be reduced to estimation from a prior distribution and a likelihood function.

The term *Bayesian statistics* has come to mean the use of Bayes theorem in a manner different from that discussed in [Example 6.3](#). Rather, experience, hypothesis, or some other means is used to assume a prior distribution when one is not known. This Bayesian approach is illustrated in [Example 6.5](#).

Example 6.5 A Bayesian Statistical Analysis

A different application of Bayes theorem occurs when one makes an educated guess at what an appropriate prior might be. To illustrate such an application, consider the problem faced by two statisticians, one a frequentist and the other a Bayesian, who are asked by a Buffon needle dropper to interpret the results from three experiments, each using a different needle length L and grid spacing D . In particular, the statisticians are asked to estimate the probability a dropped needle cuts a grid line.

In the first experiment 682 of 1000 dropped needles intersect a grid line. The frequentist immediately says the probability that the next dropped needle will cut a grid line is $682/1000 = 0.682$, because probability is interpreted as being solely based on observed frequency.

The Bayesian's estimate depends on what prior is assumed. The Bayesian might argue that experience indicates that most Buffon needle droppers use nearly equal needle lengths and grid spacings, i.e., $L \simeq D$, but there is some chance that $L \neq D$. Let p be the probability a needle cuts a grid line. One might assume the prior distribution is a beta distribution (see Appendix A.1.4), namely,

$$g(p|\alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1} (1-p)^{\beta-1},$$

where α and β are parameters. The mean and variance of this distribution are

$$\mu = \frac{\alpha}{\alpha + \beta} \quad \text{and} \quad \sigma^2 = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}. \quad (6.51)$$

The Bayesian decides it is reasonable to assume $\mu = 0.636620$ (P_{cut} for $L = D$) and $\sigma^2 = 0.01$ (to allow some variation in L/D). Solving Eq. (6.51) for α and β , the Bayesian estimates the parameters of the prior distribution as

$$\alpha = \frac{\mu^2}{\sigma^2}(1 - \mu) - \mu = 14.0906 \quad \text{and} \quad \beta = \frac{\mu}{\sigma^2}(1 - \mu)^2 + \mu - 1 = 8.0429.$$

Now, to apply Bayes theorem, one must consider the likelihood ℓ and the global likelihood M . If one conducts n trials of an experiment that has constant probability p of success (such as a needle cutting a grid line), then the PDF that k successes are obtained is given by the binomial distribution (see Appendix A.2.2)

$$\ell(k|n, p) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}, \quad k = 0, 1, 2, \dots, n.$$

The global likelihood is given by integrating the likelihood function weighted by the prior distribution, see Eq. (6.49), and in this case assumes the form

$$M = \int_0^1 \ell(k|n, p) g(p|\alpha, \beta) dp,$$

which can be shown to yield Johnson and Kotz [1970]

$$M = \frac{n!}{k!(n-k)!} \frac{\Gamma(\alpha + \beta + n)}{\Gamma(\alpha + n)\Gamma(\beta + n - k)} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)}.$$

It can be shown that the posterior distribution is also a beta distribution²

$$\begin{aligned}
 f(p|k, n, \alpha, \beta) &= g(p|\alpha, \beta) \frac{\ell(k|n, p)}{M} \\
 &= \frac{\Gamma(\alpha + \beta + n)}{\Gamma(\alpha + \beta)\Gamma(\beta + n - k)} p^{\alpha+k-1} (1-p)^{\beta+n-k-1} \\
 &= g(p|\alpha + k, \beta + n).
 \end{aligned} \tag{6.52}$$

The Bayesian statistician, thus, estimates the probability that the next needle drop cuts a grid line, given the data for the first experiment, as the expected value of $f(p|k, n, \alpha, \beta)$, which is just the mean value of the beta distribution $g(p|\alpha + k, \beta + n)$, namely,

$$\bar{p} = \frac{\alpha + k}{\alpha + \beta + n} = \frac{14.0906 + 682}{14.0906 + 8.0429 + 1000} = 0.681.$$

This value is very close to the frequentist's estimate of 0.682; and this agreement occurs because the data are reasonable, in the sense that the results obtained are within one standard deviation of the expected results if $L = D$ (i.e., 637 cuts in 1000 needle drops).

However, in the second experiment, astonishingly, none of the 1000 dropped needles cuts a grid line. The frequentist, therefore, would estimate the probability of a grid intersection on the next needle drop as $\bar{p} = 0/1000 = 0.0$. However, the Bayesian would estimate

$$\bar{p} = \frac{\alpha + k}{\alpha + \beta + n} = \frac{14.0906 + 0}{14.0906 + 8.0429 + 1000} = 0.014.$$

The Bayesian would, thus, agree that it is very unlikely to have a grid intersection on the next needle drop, but it would still be possible with a probability, given the data, of about 1 in 72.

So a Bayesian approach appears to give results very similar to the classic frequentist approach for the first two experiments. But this is not always the case. Consider now that the needle dropper begins the third experiment, drops three needles, recording 0 cuts, and then promptly leaves for the local tavern. A frequentist analyzing these data would again assume $\bar{p} = 0/3 = 0.0$. However, the Bayesian, who uses the same beta prior, would obtain

$$\bar{p} = \frac{\alpha + k}{\alpha + \beta + n} = \frac{14.0906 + 0}{14.0906 + 8.0429 + 3} = 0.561.$$

The Bayesian incorporates experience that Buffon experiments frequently use, $L \simeq D$, and obtains a result that is not too different from 64% (the mean of the beta distribution if it is assumed that $D = L$), because only three trials were performed.

6.4 Inference and Decision Applications

The Bayesian formulation provides a powerful way to draw inferences. The goal of Bayesian inference often is to estimate some quantity that depends on an unknown

² This is true because the beta and binomial distributions are so-called *conjugate* distributions.

distribution. For instance, one might want to estimate the expected value of a function $z(\boldsymbol{\theta})$ with respect to a distribution of which the parameters are $\boldsymbol{\theta}$. Bayes theorem allows us to express the unknown distribution as a posterior and then the expected value of z can be written as

$$\langle z \rangle = \frac{1}{M} \int_{\Theta} z(\boldsymbol{\theta}) f(\boldsymbol{\theta} | \mathbf{D}) d\boldsymbol{\theta} = \int_{\Theta} z(\boldsymbol{\theta}) g(\boldsymbol{\theta}) \ell(\mathbf{D} | \boldsymbol{\theta}) d\boldsymbol{\theta}, \quad (6.53)$$

where Θ is the domain of $\boldsymbol{\theta}$ and $\mathbf{D}$ are the data. The true PDF, call it $f_T(\boldsymbol{\theta})$, is approximated by the posterior PDF f , which is written in terms of a prior distribution (which the user can specify) and a likelihood function (which depends on the data). In some cases, burn-in may be required in order to reach an ergodic chain.

Consider a complex situation, such as modeling the spread of a disease. The infection rate among a population may be dependent on many variables, such as the population of carriers, the proximity of carriers to transportation portals, the behavior characteristics of carriers, the age distributions among carrier and infected populations, and even ambient conditions such as temperature and humidity. These variables form a set of parameters $\boldsymbol{\theta}$, the distribution of which is not fully known. Now consider that an investigator forms a complex model $z(\boldsymbol{\theta})$ for the number of cases of the disease in the some country at some future time. The expected number of cases then assumes a form (at least conceptually) of the expectation shown in Eq. (6.53). The investigator can then infer the expected number of cases, subject to the assumptions of the model, by MCMC. The formulation of Eq. (6.53) can be only notional, i.e., the model may not be a specific mathematical function but rather is a complex physical process. In attempting to construct the estimate of $\langle z \rangle$, the investigator averages a number of Monte Carlo histories. In each history, values of $\boldsymbol{\theta}$ are selected from a proposal distribution and accepted or rejected based on the Metropolis ratio, which depends on the assumed proposal, the assumed prior, and the likelihood function, which depends on the given data. The histories then progress through the hypothesized populations and conditions according to the proposed model z .

Another form of inference occurs when one uses observations of one set of variables to infer the values of another set of variables. For instance, [Mosegaard and Sambridge \[2002\]](#) discuss an interesting case in which seismometer data were collected on the moon. The seismic record was generated by moonquakes, meteor impacts, and impacts of spacecraft. The seismic data were reduced to the first arrival times of P- and S-waves at the various seismometer locations. Given these data, a complex model was used that depended on the assumed locations and epicenter times of the various events. The model was not an analytic one, in which a functional form for arrival times was assumed. Rather, a complex ray-tracing algorithm that depended on moon crust characteristics was simulated and a posterior distribution of P- and S-wave velocities was generated. These velocity distributions are clearly inferred from the data and are subject to considerable uncertainty. However, they do give plausible information regarding, for instance, lunar stratigraphy.

The MCMC formalism also has been applied in decision theory. Many decisions must be made in the presence of significant uncertainty. How does one decide among

the alternatives when so much is uncertain? One way is to employ MCMC. Let θ now represent a decision and θ^* represent the “true scenario,” i.e., the decision that should be made if there were no uncertainty. Further, let A represent an action that results from a decision. Then, it is possible to define a “loss function” $L(\theta, A)$ that is a measure of how much is lost when a decision θ leads to action A . In this case, the best decision can be thought of as the decision that leads to an action that minimizes the loss function L . Mathematically the value of θ is sought that minimizes

$$\Omega(\theta) = \int_{\mathcal{A}} L(\theta, A)p(A|\theta)dA, \quad (6.54)$$

where $p(A|\theta)$ is the PDF for all actions that might result from decision θ and $\mathcal{A}$ is the domain of A . This equation is in the form of a definite integral and thus is amenable to solution by Monte Carlo techniques, including MCMC.

Bayes theorem indicates that the posterior PDF f can be expressed in terms of a prior PDF g , a likelihood function ℓ , and a constant M . The Metropolis-Hastings algorithm involves ratios of the posterior and proposal PDFs and hence the constant term M (when expressing the posterior in terms of the prior) cancels. This leads to the important result that the global likelihood M need not be known. This implementation extends the reach of Monte Carlo into areas where conventional Monte Carlo methods are difficult to apply.

6.4.1 Implementing MCMC with Data

Assume one is given a data vector $\mathbf{D} = \{d_j, j = 1, 2, \dots, J\}$ of measured values of a random variable $\mathbf{x}$ and seeks to estimate the quantity

$$\langle z \rangle = \int_{-\infty}^{\infty} z(\mathbf{x})f_T(\mathbf{x})d\mathbf{x}, \quad (6.55)$$

where $z(\mathbf{x})$ is some specified function and $f_T(\mathbf{x})$ is the true (but unknown) PDF underlying the distribution of $\mathbf{x}$. In principle, one can estimate $\langle z \rangle$ by using Bayes theorem to form the posterior

$$f(\mathbf{x}|\mathbf{D}) = \frac{1}{M}g(\mathbf{x})\ell(\mathbf{D}|\mathbf{x}), \quad (6.56)$$

where $g(\mathbf{x})$ is an assumed prior distribution, $\ell(\mathbf{D}|\mathbf{x})$ is the likelihood of the data given a sample of $\mathbf{x}$, and M is the global likelihood (a normalizing constant). The likelihood function ℓ uses the data $\mathbf{D}$ to correct the prior g so that the posterior f better approximates the true PDF f_T than does the prior. One can find an estimate of $\langle z \rangle$ from

$$\bar{z} = \frac{1}{N} \sum_{i=1}^N z(\mathbf{x}_i), \quad (6.57)$$

where N is the number of histories on which the estimate is based, and the $\mathbf{x}_i$ are sampled from Eq. (6.56) by MCMC. Note that in order to apply MCMC, the value of M need not be known and, thus, can be treated as unity because it will always cancel out when the Metropolis ratio is formed.

The prior should be a good approximation to the true PDF. In fact, of course, the best prior *is* the true PDF (see Example 6.4). Thus, investigators can and should use judgment and experience to select reasonable priors. It is interesting that the prior need not be a function that is easily sampled, because in MCMC candidate samples are drawn from the proposal distribution h , not from g . In fact, the prior need not even be a normalized PDF, it only needs to have the shape of the desired distribution.

After first selecting a prior distribution g , one then selects a proposal distribution $h(\mathbf{x}_{i+1}|\mathbf{x}_i)$, which is defined on the volume over which the random variable $\mathbf{x}$ is assumed to be defined. Note that the proposal does not have to be conditional. However, it should be a distribution that is easily sampled and better mixing is achieved if it is reasonably close in shape to the true PDF $f_T(\mathbf{x})$. From this point, an MCMC procedure, with data, can proceed as follows:

1. Let $i = 1$ and specify an initial value of $\mathbf{x}$, within the appropriate volume, and call it $\mathbf{x}_i$.
2. Sample a candidate value of $\mathbf{x}$ from the proposal and call it $\mathbf{u}_i$.
3. Form the Metropolis ratio

$$R = \frac{g(\mathbf{u}_i)\ell(\mathbf{D}|\mathbf{u}_i)h(\mathbf{x}_i|\mathbf{u}_i)}{g(\mathbf{x}_i)\ell(\mathbf{D}|\mathbf{x}_i)h(\mathbf{u}_i|\mathbf{x}_i)}. \quad (6.58)$$

4. Select a pseudorandom number ρ from the unit interval.
5. Set

$$\mathbf{x}_{i+1} = \begin{cases} \mathbf{u}_i, & \text{if } \rho \leq \min [1, R] \\ \mathbf{x}_i, & \text{if } \rho > \min [1, R] \end{cases}. \quad (6.59)$$

6. Increment i by one and repeat steps 2 to 6 until all histories have been run.

The samples $\mathbf{x}_i$ can be used in Eq. (6.57) to provide an estimate of the desired quantity.

The likelihood function

There are several ways to incorporate data through the likelihood function. Some rely on the intuition and guile of the investigator, while others are somewhat pedantic. Four ways are outlined as follows and two examples are given for illustration:

1. For discrete random variables, one can use a binomial distribution as the likelihood function if the data are derived from successive trials with constant probability of success, as shown in Example 6.5. More generally, a multinomial distribution (see Appendix A.3.3.2) may apply in multidimensional cases.
2. One can attempt to use the data to construct a histogram of ℓ ; this works better as the number of data points increases. Then, any time a sample is drawn from within a given

histogram interval, the appropriate histogram value of ℓ is used to “weight” the prior. This is demonstrated in [Example 6.6](#).

3. The likelihood is specified according to the “physics” of the problem. With knowledge of something about the specific problem at hand, a likelihood function often can be inferred.
4. Another way to incorporate the data is to specify a likelihood function of the form

$$\ell(\mathbf{D}|\mathbf{x}) = Ce^{-S(\mathbf{x})}, \quad (6.60)$$

where C is a constant and S is the “misfit function” given, on the i th history, by

$$S(\mathbf{x}_i) = \sum_{j=1}^J \frac{|d_j - x_{j,i}|}{\sigma_j}, \quad (6.61)$$

with σ_j the standard deviation of d_j and $x_{j,i}$ the value of x_j on the i th history. If the MCMC samples from f are drawn from the true distribution, then the misfit function should tend toward unity and ℓ should approach the constant C/e . The more the samples differ from the data, the larger is S and thus the smaller is ℓ . This approach is discussed, for instance, by [Mosegaard and Sambridge \[2002\]](#).

Example 6.6 A One-Dimensional Problem with Data

Consider that the time intervals between events (such as tornado impacts within a state) are measured and used to form a data vector $\mathbf{D} = \{d_j, j = 1, 2, \dots, J\}$. You are asked to estimate the average time interval ([Eq. \(6.55\)](#) with $z(x) = x$). Of course, this can be estimated by simply averaging the data, but it is instructive to ignore this, for the time being, and demonstrate that it can be estimated from [Eq. \(6.55\)](#). Assume that $J = 1000$ time intervals have been measured, with the maximum time interval being 15.6513 weeks. For simplicity, select the uniform prior

$$g(x) = \frac{1}{T}, \quad 0 \leq x \leq T,$$

where T is any time greater than or equal to the maximum of the measured times. Consider a proposal distribution with the simple form

$$h(u) = 0.5e^{-0.5u},$$

which is easy to sample and provides samples on the interval $[0, \infty)$. Thus, for each history,

$$u = -\frac{1}{0.5} \ln(\rho), \quad (6.62)$$

where ρ is a pseudorandom number drawn uniformly from the unit interval. Grouping the measured data into K equal bins, each of width $\delta = T/K$, a histogram likelihood function

can be formed

$$\ell(\mathbf{D}|x_j) = \begin{cases} \ell_1, & 0 < x_j \leq \delta \\ \ell_2, & \delta < x_j \leq 2\delta \\ \vdots & \\ \ell_K, & (K-1)\delta < x_j \leq K\delta \end{cases}, \quad (6.63)$$

where ℓ_k is the number of data points that are within the interval $[(k-1)\delta, k\delta]$. It obviously is undesirable to have the likelihood function be zero over any interval that can contain samples, and so any of the ℓ_k that are zero are arbitrarily set to $\ell_k = 1$. Further, samples can be selected on the interval $x > T$, and so the histogram contains another entry ℓ_{K+1} , which in this example is arbitrarily set to $\ell_{K+1} = 0.001$. This likelihood function clearly is not normalized, but the MCMC procedure applies despite this fact.

It is easy to see that for this simple model the Metropolis ratio on the i th history takes the form

$$R = \frac{\ell_{ku}}{\ell_{kx}} e^{-0.5(x_i - u_i)}.$$

Here, ℓ_{ku} is the value of the likelihood function in the interval in which u_i is located and ℓ_{kx} is the value of the likelihood function in the interval in which x_i is located.

The results of implementing the procedure described here for $N = 10^6$ histories and $K = 50$ are as follows:

$$\bar{x} = 2.602 \pm 0.007, \quad \text{with } m = 0$$

and

$$\bar{x} = 2.616 \pm 0.007, \quad \text{with } m = 1000.$$

The average of the data is $\bar{d} = 2.506$, and it turns out that the data were derived by sampling an exponential distribution of the form $f_T(x) = 0.4e^{-0.4x}$, for which the average value is $\bar{x} = 1/0.4 = 2.5$. This very approximate procedure leads to an estimate of the mean that is within 5% of the true mean. Note that the estimate of $\bar{x}$ obtained with no burn-in ($m = 0$) is just as good as the estimate with burn-in ($m = 1000$), indicating, though not proving, that burn-in is not required.

Example 6.7 Another One-Dimensional Problem with Data

For similar time-interval data as in the previous example, select a value Δt . For each d_j find the n_j for which $(n_j - 1)\Delta t \leq d_j < n_j\Delta t$. This creates a new data vector $\mathbf{D}_n = (n_1, n_2, \dots, n_J)$. Here, n_j is the number of the time interval within which d_j is located. The radioactive decay process is known to be governed by an exponential distribution. Thus, the probability p_j that a single observation will be within the j th interval, i.e., $p_j = \text{Prob}\{(n_j - 1)\Delta t \leq d_j < n_j\Delta t\}$, is given by

$$p_j = \int_{(n_j-1)\Delta t}^{n_j\Delta t} \lambda e^{-\lambda t} dt = e^{-\lambda(n_j-1)\Delta t} - e^{-\lambda n_j\Delta t}.$$

For J observations, the probability of observing $\mathbf{D}$ is the likelihood, which is simply

$$\ell(\mathbf{D}|\lambda, \Delta t) = \prod p_j = \prod_{j=1}^J e^{-n_j \lambda \Delta t} \left[e^{\lambda \Delta t} - 1 \right]. \quad (6.64)$$

Now, the problem is to find the “decay constant” λ of the true exponential distribution that produced the observed data. The true distribution of λ is, of course, a delta function, i.e., $f_T(\lambda) = \delta(\lambda - \lambda_T)$, where λ_T is the desired but unknown decay constant.

An estimate of the true decay constant can be obtained, using the data, by forming the posterior

$$f(\lambda|\mathbf{D}) = \frac{1}{M} g(\lambda) \ell(\mathbf{D}|\lambda),$$

where g is an assumed prior, ℓ is given by Eq. (6.64), and M is the global likelihood (which need not be known). Then, the estimate of λ_T is given by

$$\bar{\lambda} = \int_0^\infty \lambda f(\lambda|\mathbf{D}) d\lambda.$$

The MCMC procedure converts the prior into a posterior that approximates the true delta function. Assume a prior in the form of a gamma distribution (see Appendix A.1.3) with parameters α and β , i.e.,

$$g(\lambda) = \frac{1}{\beta \Gamma(\alpha)} \left(\frac{\lambda}{\beta} \right)^{\alpha-1} e^{-\lambda/\beta}. \quad (6.65)$$

Also, assume a proposal distribution of the simple form

$$h(u) = 0.5e^{-0.5u},$$

as in the previous example. Then, specify an initial value of λ_i , for $i = 1$, sample a candidate new value u_i from the proposal, and form the Metropolis ratio

$$R = \frac{y_i}{\lambda_i} e^{-(y_i - \lambda_i)/\beta} \frac{1 - e^{-y_i \Delta t} \prod_{j=1}^J e^{-n_j y_i \delta t}}{1 - e^{-\lambda_i \Delta t} \prod_{j=1}^J e^{-n_j \lambda_i \delta t}} e^{-0.5(\lambda_i - y_i)}. \quad (6.66)$$

The candidate samples are accepted or rejected using the criterion of Eq. (6.59), and the procedure is repeated until all desired histories have been run.

Two hundred samples were generated from an exponential with parameter $\lambda = 0.4$. Using the first 40 of these samples as given data, the results of implementing the procedure described above for $N = 10^6$ histories, with no burn-in, yielded $\bar{\lambda} = 0.4507 \pm 0.00001$. Using the first 120 of these samples as given data and again running $N = 10^6$ histories, with no burn-in, the estimate $\bar{\lambda} = 0.3991 \pm 0.000003$ was obtained. Using all 200 samples as given data, the estimate $\bar{\lambda} = 0.3967 \pm 0.000002$ was obtained. The posteriors, scored as histograms over 100 intervals between zero and one, are shown in Fig. 6.12. It is obvious

that, as the number of data points is increased, the posterior more closely resembles a delta function and its average more closely approximates the true value.

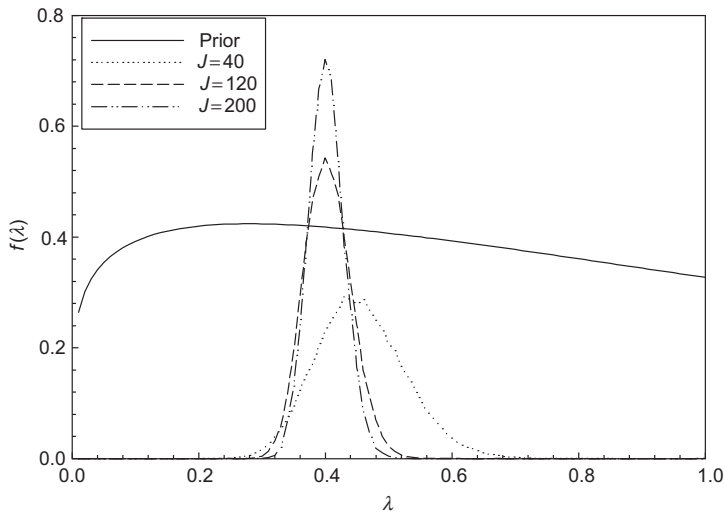


Figure 6.12 The posterior distributions for three values of J after $N = 10^6$ histories (with no burn-in) compared to the prior distribution, which is a gamma PDF with $\alpha = 1.2$ and $\beta = 1.4$. The nonprior distributions were scaled by a factor of 5.

The MCMC method has been applied in very many situations. For example, [Gilks et al. \[1996\]](#) describe the use of MCMC in many medical applications (including the study of disease risks, imaging, genetics, and medical modeling), [Hugtenburg \[1999\]](#) applies MCMC to radiation treatment planning, and [Litton and Buck \[1996\]](#) consider application of MCMC to radiocarbon dating in archaeological samples. Yet another application of MCMC includes the additive white Gaussian noise (AWGN) problem, discussed by [Wang \[2005\]](#). This problem considers an attempt to determine a transmitted signal from the received signal, in the presence of noise. Wang treats this by constructing an MCMC model in which samples are accepted and rejected on the basis of the Metropolis-Hastings algorithm. Another way to look at this particular problem is as an inverse problem in which one wishes to invert a set of measurements to obtain the signal that led to those measurements. It is this class of problems that is considered in the next chapter on Monte Carlo applications to inverse problems.

6.5 Summary

Markov chain Monte Carlo is, in essence, a particular way to obtain random samples from a PDF. Thus, it is simply a method of sampling. The method relies on using properties of Markov chains, which are sequences of random samples in which each

sample depends only on the previous sample. Some chains of random samples form ergodic Markov chains from any initial state and, thus, require no burn-in. Ergodic Markov chains have transition probabilities that are constant and produce samples from a single stationary PDF.

Some of the principal advantages of MCMC include the following:

- It is relatively simple to implement.
- It can be applied when only the shape of the distribution is specified, i.e., the normalization constants required to make functions into PDFs need not be known.
- It can be applied to problems for which Bayes theorem is employed, allowing one to sample from a PDF f (the posterior) using another PDF g (the prior) and a likelihood function.
- It can be applied to very complex and highly multidimensional sampling problems.

These advantages must be weighed against some disadvantages, which include the following:

- The acceptance/rejection criterion employed in the Metropolis-Hastings algorithm results in retaining the same random sample on some successive samplings; because of this, the sample standard deviation is artificially low, until N becomes sufficiently large.
- The production of a new sample value requires that at least two pseudorandom samples be generated, a requirement that can be somewhat inefficient.

Nevertheless, MCMC is a powerful form of Monte Carlo and has gained widespread use, especially in sampling highly multidimensional distributions or in situations where other Monte Carlo methods encounter difficulties.

Bibliography

- BERNSTEIN, S.N. 1926, "Sur l'Extension du Théorème Limite du Calcul des Probabilités aux Sommes de quantités dépendentes," *Math. Ann.*, **97**, 1–59.
- FISHMAN, G.S. 2000, *Monte Carlo Concepts, Algorithms, and Applications*, Springer-Verlag, New York.
- GELFAND, A.E. AND A.F.M. SMITH 1990, "Sampling-based Approaches to Calculating Marginal Densities," *J. Am. Stat. Assoc.*, **85**, 398–409.
- GEMAN, S. AND D. GEMAN 1984, "Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images," *IEEE T. Pattern. Anal.*, **6**, 721–741.
- GILKS, W.R., S. RICHARDSON, AND D.J. SPIEGELHALTER, eds. 1996, *Markov Chain Monte Carlo in Practice*, Chapman & Hall, New York.
- HASTINGS, W.K. 1970, "Monte Carlo Sampling Methods Using Markov Chains and Their Applications," *Biometrika*, **57**(1), 97–109.
- HUGTENBURG, R.P. 1999, "A Perturbation Based Monte Carlo Method for Radiotherapy Treatment Planning," *Radiother. Oncol.*, **51**, S30.
- JOHNSON, N.L. AND S. KOTZ 1970, *Continuous Univariate Distributions*, Vol. 2, Houghton Mifflin Co., Boston, MA.
- LITTON, C. AND C. BUCK 1996, "An Archaeological Example: Radiocarbon Dating," in *Markov Chain Monte Carlo in Practice*, W.R. Gilks, S. Richardson, and D.J. Spiegelhalter, eds., Chapman & Hall, New York.

- MACKEY, D.J.C. 2007, *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press, Cambridge, UK (sixth printing).
- MARKOV, A.A. 1906, "Rasprostranenie zakona bol'shih chisel na velichiny, zavisyaschie drug ot druga [Extension of the Law of Large Numbers to Dependent Quantities]," *Izv. Fiz.-Matem. Obsch. Kazan Univ.*, (2nd Ser.), **15**(94), 135–156.
- METROPOLIS, N. AND S. ULAM 1949, "The Monte Carlo Method," *J. Am. Stat. Assoc.* Special Issue, 125–130.
- METROPOLIS, N., A.W. ROSENBLUTH, M.N. ROSENBLUTH, A.H. TELLER, AND E. TELLER 1953, "Equation of State Calculations by Fast Computing Machines," *J. Chem. Phys.*, **21**(6), 1087–1092.
- MOSEGAARD, K. AND M. SAMBRIDGE 2002, "Monte Carlo Analysis of Inverse Problems," *Inverse Probl.*, **18**, R29–R54.
- NEAL, R.M. 1993, Probabilistic Inference Using Markov Chain Monte Carlo Met Technical Report CRG-TR-93-1, Dept. of Computer Science, University of Toronto.
- POPP, J.G. AND D.B. WILSON 1996, "Exact Sampling with Coupled Markov Chains and Applications to Statistical Mechanics," *Random Struct. Algor.* **9**(1–2), 223–252.
- TIERNEY, L. 1994, "Markov Chains for Exploring Posterior Distributions (with Discussion)," *Ann. Statist.*, **22**(4), 1701–1762.
- WANG, X. 2005, Monte Carlo Signal Processing for Digital Communications: Principles and Applications, Ch. 26 in *Signal Processing for Mobile Communications Handbook*, M. Ibnkahia, ed., CRC Press, New York.

Problems

1. Use MCMC to estimate the value of the definite integral

$$I = \int_0^1 x^2 f(x|\alpha, \beta) dx,$$

where $f(x|\alpha, \beta)$ is the beta distribution (see Appendix A.3.1). Consider the case where $\alpha = 1.5$ and $\beta = 1.2$. (a) Start at $x_1 = 0.5$ and employ the burn-in lengths of $m = 0$, $m = 1000$, and $m = 10,000$ to estimate I using 10^6 histories. What can you conclude from your results? (b) What are the standard deviations of your estimates?

2. Use MCMC to estimate the integral

$$I = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x + y) f(x, y) dx dy,$$

where $f(x, y)$ is the bivariate normal distribution (see Appendix A) with the parameters $\langle x \rangle = \mu_1 = 5$, $\sigma(x) = \sigma_1 = 1$, $\langle y \rangle = \mu_2 = 1$, $\sigma(y) = \sigma_2 = 0.5$, and $\sigma_{12} = 0.5$. Sample both x and y on each history and use the proposal functions $g(u) = 2e^{-4|u|}$ for x and $h(v) = 0.4e^{-0.8|v|}$ for y .

3. Use MCMC to estimate the value of the definite integral

$$I = \int_{-\infty}^{\infty} \int_0^{\infty} \int_2^8 (x/y)^2 f(x) g(y) h(z) dx dy dz,$$

where $f(x) = 4/9 - x/18$, $g(y) = 4e^{-4y}$, and

$$h(z) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right],$$

with $\mu = 1$, $\sigma = 0.4$.

4. With $\mu = 10.0$ and $\sigma = 5.0$, solve the problem of [Example 6.2](#) using the proposal function

$$h(y|x) = xe^{-xy}$$

and compare your results with those given in [Example 6.2](#).

5. Estimate the triple integral

$$I = \int_0^1 \int_0^{10} \int_0^\infty \cos(xy) \frac{3y}{\sqrt{z}} e^{-4x} dx dy dz$$

by MCMC using the proposals for x , y , and z given by

$$h_1(u) = 3e^{-3u}, \quad u > 0,$$

$$h_2(v) = 1/10, \quad 0 \leq v \leq 10,$$

and

$$h_3(w) = 2w, \quad 0 \leq w \leq 1,$$

respectively. Compare results for $N = 10^6$ histories for the following cases. (a) Update x , then y , and then z in successive histories. (b) Next, update x , y , or z randomly on any given history. (c) Then, update x , y , and z globally on each history. Compare your results.

7 Inverse Monte Carlo

*Karen had the right answer, you see;
but didn't know what the question might be.
After looking around
she finally found.
she could use symbolic IMC.*

Most of this book has focused on “forward” Monte Carlo methods that can be applied to solving “direct” problems. In direct problems, the response of a specified process is sought, i.e., one seeks to determine the output of a given system for a certain input. A definite integral poses a direct problem, but so does a model of a complex system, whether or not the model is in the form of a single mathematical formula or is expressed as a complicated algorithm. Monte Carlo solves direct problems by using samples of independent variables in order to estimate values of dependent variables. The independent variables can be discrete, continuous, or a combination of the two. Often, the independent variables are stochastic but this need not be the case. In direct problems, Monte Carlo can be considered to be a form of quadrature or, alternatively, a means to simulate a process in which the “physics” (the underlying processes) of the problem are known or specified.

“Inverse” problems, by contrast, arise when the outcome or response is given and the conditions that might have led to these results is sought. Meteorologists measure microwave intensity distributions from which they seek to determine atmospheric concentration profiles in order to make weather predictions. Doctors measure temperature and blood pressure in order to assess a patient’s health. Cooks use a variety of measures (some rather subjective) to ascertain when a dish is ready to serve. Our eyes are sensors that detect light, allowing us to recognize each other. Our brains perform inverse analyses to identify words from sounds produced by others or to recognize specific symphonies from the sounds produced by musical instruments. People are detectors who make other detectors to enhance their ability to perceive the world. The responses of these detectors provide data that are turned into information. Thus, inverse problems differ from direct problems in that the values of the independent variables that could have led to a set of observations are sought.

In many direct engineering problems, initial and/or boundary conditions are specified and some conditions at later times and/or at internal positions are desired in a system with specified properties. For example, a direct problem might be stated that given an initial surface temperature distribution on an object of known composition, what is the internal temperature distribution at future times? There are many forms of inverse problems. In some engineering inverse problems, system responses are measured and one wants to determine either the properties of the system or the initial

and/or boundary conditions that could have led to these responses. In other inverse problems, a desired response to a given input is specified and one seeks the properties of a system that would produce this desired response. Unlike direct problems, however, inverse problems are plagued by questions of *existence* and *uniqueness*.

The existence of solutions to some inverse problems may not always be guaranteed in advance. It is certainly possible, in some cases, to specify a set of final or internal conditions that cannot be achieved by any set of system properties and, thus, no solution exists. For example, one may want the surface temperature of an object to be constant under some specified heating conditions, but no geometry may exist that will achieve this result for the specified material. In direct problems, the system is responding, according to natural laws, to a set of imposed conditions. In inverse problems, the parameters of some hypothetical system are sought that would produce some specified conditions. It is certainly possible to specify conditions that cannot be realized physically (and perhaps not even mathematically).

Even if a solution does exist, that solution may not be unique. There may be several sets of parameters that could produce the same specified output. Alternatively, some solutions may be allowed mathematically that are not physically realizable. For instance, negative mass densities have no physical meaning and temperatures that are expressed as complex numbers may solve an equation but make no physical sense. Thus, uniqueness also is an issue in inverse problems. Moreover, it often is difficult to prove existence and uniqueness. However, for many practical problems, it is highly likely that at least one solution exists, and it is important to devise methods to find the physically reasonable solutions, from which an optimum solution generally can be determined.

Consider, for example, the backscatter response of a gamma-ray density gauge, which is a function of the density of the interrogated medium of which the composition is known. A possible model for the response R in terms of density ρ for a medium with known composition is of the form

$$R(\rho) = \rho e^{\alpha - \beta \rho} \quad (7.1)$$

and behaves as shown in Fig. 7.1. A direct problem is to calculate R , given α and β , for a specified density. An inverse problem is to calculate ρ , given α and β , for a measured response. The uniqueness problem is concerned first with whether any solution for ρ exists. For instance, the response R_1 shown in Fig. 7.1 does not correspond to any density. Second, if one solution exists, do others? For instance, the response R_2 leads to two positive densities. One must determine which of the two is most physically reasonable.

The situation in practice generally is even more complicated. Measurements are imprecise and the uncertainty in measurements leads to uncertainty in the recovered density. The response R_3 in Fig. 7.1 corresponds to a single density. However, an actual response from a sample with that density is likely either slightly above R_3 , for which there is no inverse solution, or slightly below R_3 , for which there are two solutions. Also, responses generally are functions of several variables, some of which are sought and others of which “interfere” with the measurement. Suffice it to say

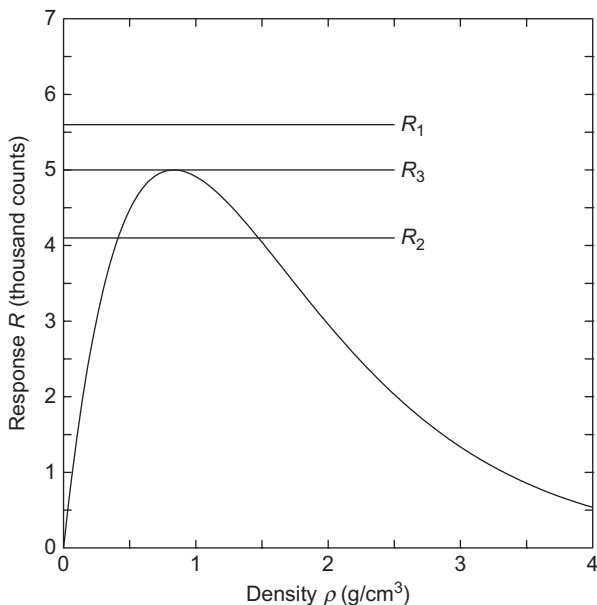


Figure 7.1 Plot of backscatter response versus density.

that existence and uniqueness are generic issues for inverse problems with which all solution techniques must deal. Monte Carlo approaches to solving inverse problems usually just assume that one or more solutions exist and then try to deduce the most physically reasonable solution among the solutions obtained.

A “well-posed” inverse problem provides sufficient information to determine if at least one solution exists, even if that solution is unknown and not unique (perhaps several sets of parameters could lead to the observed result). An “ill-posed” inverse problem is one for which not enough information is given to specify whether or not a solution exists, for instance, because there are more variables to determine than there are equations relating these variables. In this chapter, it is assumed that the inverse problems are well posed and, thus, that at least one solution exists. However, multiple solutions may exist, within the uncertainties of the data provided, and hence uniqueness is not guaranteed.

7.1 Formulation of the Inverse Problem

Two general approaches to formulating inverse problems are briefly reviewed. The first is the formal mathematical approach, which parallels the integral formulation of direct Monte Carlo. The other is the practical or physical approach, the analog of which is the simulation approach of direct Monte Carlo. Neither approach is inherently better than the other. Some problems are easily specified in equation form, while others are more easily posed in terms of the physical processes. This does not mean that there are two

classes of inverse problems, only that there are two ways to specify these problems. The situation is rather like wave-particle duality. Photons can be considered waves for some experiments (the two-slit experiment) and particles for other experiments (photoelectric absorption). So far as is known, both models are correct and just illuminate different aspects of the behavior of photons. The two general approaches for inverse problems are outlined in the following two sections. This is followed by a brief section that identifies certain optimization problems as inverse problems.

7.1.1 Integral Formulation

Consider the expected value of some *kernel* z expressed in the form

$$\langle z_k \rangle = \langle z(\mathbf{y}_k) \rangle = \int_V z(\mathbf{x}, \mathbf{y}_k, \boldsymbol{\theta}) f(\mathbf{x}, \boldsymbol{\theta}) d\mathbf{x}, \quad k = 1, 2, \dots, K, \quad (7.2)$$

where $\mathbf{x}$ is an n -dimensional random variable distributed according to the joint PDF f , $\mathbf{y}$ is a vector of independent parameters that can take on specified values $\mathbf{y}_k$, with $k = 1, 2, \dots, K$, V is the domain of $\mathbf{x}$, and $\boldsymbol{\theta}$ is a set of $J \geq 1$ parameters of the integrand. In direct problems, the $\boldsymbol{\theta}$ are assumed known or specified and the object is to estimate the expectations of Eq. (7.2) for $K \geq 1$ values of $\mathbf{y}_k$. Analog Monte Carlo does this by using the sample mean defined by

$$\bar{z}(\mathbf{y}_k) = \frac{1}{N} \sum_{i=1}^N z(\mathbf{x}_i, \mathbf{y}_k, \boldsymbol{\theta}), \quad (7.3)$$

where the $\mathbf{x}_i$ are sampled from $f(\mathbf{x}, \boldsymbol{\theta})$.

A similar formulation exists for inverse problems. Let $R(\mathbf{y}_k, \boldsymbol{\theta})$ be the response model for a system when an independent variable $\mathbf{y}$ has the specific value $\mathbf{y}_k$ and let $\boldsymbol{\theta}$ be a vector of $J \geq 1$ unknown parameters of the system. Then, the equations

$$R(\mathbf{y}_k, \boldsymbol{\theta}) = \int_V z(\mathbf{x}, \mathbf{y}_k, \boldsymbol{\theta}) f(\mathbf{x}, \boldsymbol{\theta}) d\mathbf{x}, \quad k = 1, 2, \dots, K, \quad (7.4)$$

where the kernel z is determined by the physics of the problem, pose a classic inverse problem in the form of a Fredholm integral equation of the first kind to find the parameters $\boldsymbol{\theta}$. This chapter addresses how Monte Carlo offers a means to solve inverse problems such as those posed in the form of Eq. (7.4).

For instance, it is known that the response of a radiation detector can be expressed as a function of energy in the form (e.g., see Dunn and Shultis [2009])

$$R(E) = Q_o \int_0^\infty \chi(E_o) K(E_o \rightarrow E) dE_o, \quad (7.5)$$

where Q_o is the number of radiation particles incident on the detector during some counting time, $\chi(E_o)$ is the PDF defining the energy distribution of source particles incident on the detector, $K(E_o \rightarrow E)$ is the response of the detector at energy E due to

a unit source at energy E_o , and $R(E)$ is the response of the detector per unit energy at energy E . In practice, this formulation is usually discretized and the response in a finite number of contiguous energy bins or “channels” is recorded. In radiation detection, it is a direct problem to estimate the detector response for a known source and a detector with specified response kernel K . However, Eq. (7.5) is an inverse problem if the measured response $R(E)$ is given and one seeks, for instance, the energy spectrum $\chi(E_o)$ of a source of known intensity Q_o , or the intensity of a source having a known spectrum, or both the intensity and the energy spectrum.

7.1.2 Practical Formulation

Many inverse problems are not easily cast into the form of Fredholm integral equations. Direct Monte Carlo can be viewed and implemented as a simulation of a process, such as estimating the neutron density at some position in a reactor core, given the core composition and size, without resorting to writing a definite integral. Similarly, inverse Monte Carlo can be viewed as using simulation to solve inverse problems for which no integral equation is formally considered. One can in fact use Monte Carlo to find the size of a reactor core given its composition or, alternatively, the composition of a reactor given its size, without ever considering Fredholm integral equations.

It is easy to imagine many other examples. Say that you find a metal artifact in the shape of a rectangular plate, of which the thickness L is much smaller than the other two dimensions, and want to determine the thermal conductivity, in an effort to identify the metal. You introduce a constant heat flux q'' on one face and measure the temperature difference ΔT between that face and the opposite face in steady state. You know that in this simple case, Fourier’s law of heat conduction can be written approximately as

$$q'' = k \frac{\Delta T}{L}, \quad (7.6)$$

where k is the thermal conductivity of the metal. You can then use the measured temperature difference and thickness to estimate the thermal conductivity, and thus infer the material out of which the plate is made. This problem differs from the direct problem of estimating the temperature difference for a plate of known material and thickness. The inverse problem in this case takes measured temperature values and seeks a property of the medium under consideration.

Many other inverse problems are posed as physical or conceptual inverse problems for which a process can be modeled. The retrieval of P- and S-wave velocities from lunar seismometer data discussed by Mosegaard and Sambridge [2002] is one example. The first arrival times at dispersed seismometers on the lunar surface were used with a ray-tracing model for wave propagation through unknown lunar strata to reconstruct possible wave velocities, which also demonstrated approximate lunar stratigraphy. Similarly, assume a known radiation source is placed on one side of a slab shield and n radiation detectors are placed on the other side of the shield. The thickness of the shield is known but its composition is not. One may, in principle, determine the unknown shield composition, expressed as weight fractions of $m \leq n$ elements, given

the responses from the n detectors. This is an inverse problem that is stated without writing an integral equation.

7.1.3 Optimization Formulation

Inverse problems also arise from attempts to design or optimize systems. A design optimization problem might be posed in a form such as “What system design produces a device for which the error in results using this device is minimized when measuring a set of samples?” Optimization problems typically seek to find extrema of objective functions, which are identified by locations where the partial derivatives have the desired value of zero. Thus, quite often, it is possible to consider an optimization problem as an inverse problem in which the measurements are replaced by the constraints that derivatives vanish.

Consider a system for which the location of an extremum of the objective function given by

$$\Omega(\theta) = \int_V z(\mathbf{x}, \theta) f(\mathbf{x}, \theta) d\mathbf{x} \quad (7.7)$$

is desired, where the variables are as defined in [Section 7.1.1](#). Then, the values of θ_j at which the partial derivatives of the objective function vanish, i.e.,

$$\frac{\partial \Omega(\theta)}{\partial \theta_j} = 0, \quad j = 1, 2, \dots, J, \quad (7.8)$$

identify the positions of the extrema (maxima or minima). In the remainder of this chapter, optimization problems are treated as inverse problems in which the “response models” are objective functions of the form of [Eq. \(7.8\)](#), and the “measured responses” are all zero.

7.1.4 Monte Carlo Approaches to Solving Inverse Problems

The term *inverse Monte Carlo* (IMC) has been used (e.g., see [Dunn \[1981\]](#)) to identify a specific approach, which is noniterative in the Monte Carlo simulations, to solving a class of inverse problems. Here, a broader perspective is taken and IMC refers to any Monte Carlo method applied to inverse or optimization problems. Certain implementations of MCMC that involve data to infer parameters of models to complex problems can be viewed as using Monte Carlo to solve inverse problems. In addition to using MCMC, there are two other ways to proceed. One involves IMC by iterative simulation. This approach exploits the fact that computer simulation is relatively inexpensive and, thus, many scenarios can be considered and the one of which the simulated results match the observed results most closely can be selected. The other approach, which may be advantageous when the cost of doing a simulation is expensive or the number of simulations is large (or both), converts the inverse problem into a system of algebraic equations that are then solved by standard numerical methods. The conversion involves scoring histories with symbols and numbers rather than only numbers and, hence, is called *symbolic Monte Carlo* (SMC). The SMC approach often avoids iterative simulations.

7.2 Inverse Monte Carlo by Iteration

Consider the problem of obtaining solutions to Eq. (7.4) assuming K responses $\mathcal{R}_k$, $k = 1, 2, \dots, K$ are known. In general, a “brute force” Monte Carlo approach to solving such inverse problems can be implemented. This approach involves iterative simulations, as unknown parameter values are varied. Only when simulated responses match measured or specified responses to an acceptable degree do the simulations cease. The procedure might be implemented as follows:

1. Choose an initial estimate of θ and call it θ_m with $m = 1$.
2. Construct estimates $\bar{R}_k$, $k = 1, 2, \dots, K$, by direct Monte Carlo methods that generally require sampling from a PDF using the assumed θ_m .
3. Compare the estimates $\bar{R}_k$ with the known values $\mathcal{R}_k$.
4. Choose a new estimate of θ (call it θ_{m+1}) by any of several techniques, such as by using a fixed grid superimposed over the volume V , a Fibonacci search, the simplex method [Press et al. 1996], or some other search procedure.
5. Use direct Monte Carlo to construct new estimates of the $\bar{R}_k$, $k = 1, 2, \dots, K$, and go to step 3.

The procedure is repeated until the Monte Carlo estimates $\bar{R}_k$ agree with the given responses $\mathcal{R}_k$ to within an acceptable tolerance, for all k . This process requires that a direct Monte Carlo simulation be run for each assumed value of θ and, thus, can become cumbersome and “expensive” in terms of simulation resources as the complexity of the problem or the dimensionality of θ increases. Nevertheless, many inverse problems can be solved by this iterative IMC approach, as Example 7.1 demonstrates.

Example 7.1 A Two-Parameter Inversion Using Iterative IMC

Assume that the expected response of a process involves an exponentially distributed stochastic variable and can be written as follows:

$$R(y) = \int_0^\infty \frac{x}{\sqrt{\pi y + ax}} b e^{-bx} dx.$$

Assume further that you are given values of $\mathcal{R}_1 \simeq R(y_1 = 1.0) = 0.2973 \pm 0.0002$ and $\mathcal{R}_2 \simeq R(y_2 = 5.0) = 0.1781 \pm 0.0002$ and are asked to determine the values of a and b . One way to proceed is to select initial values a and b and apply the iterative procedure just described. Starting with $a = b = 1.0$ and estimating the above integral for both $y_1 = 1$ and $y_2 = 5$ using $N = 100,000$ histories one might obtain the values given in the first row of Table 7.1. One of the parameters, say b , is then incremented to obtain the results shown in rows two through four. The results in row three are closest to the given responses and so one might next increment a for $b = 1.5$, generating the results in rows five through ten. Proceeding along similar lines, results are generated for various cases until the results in the final row of Table 7.1 are obtained. The parameter values are acceptably close to the given responses. Note that one must either evaluate the output of each run and decide on new values for the parameters or establish a close-spaced grid of many parameter pair values.

Table 7.1 Iteration toward a solution to a two-parameter inverse problem. The value of N was 10^5 in all cases. The correct values are $a = 3.23$ and $b = 1.24$.

j	a_j	b_j	$\bar{R}_{1j} \pm \sigma(\bar{R}_{1j})$	$\bar{R}_{2j} \pm \sigma(\bar{R}_{2j})$
1	1.00	0.50	0.7877 ± 0.0019	0.4540 ± 0.0013
2	1.00	1.00	0.4510 ± 0.0012	0.2377 ± 0.0007
3	1.00	1.50	0.3198 ± 0.0009	0.1617 ± 0.0005
4	1.00	2.00	0.2482 ± 0.0007	0.1224 ± 0.0004
5	0.50	1.50	0.3418 ± 0.0010	0.1640 ± 0.0005
6	1.50	1.50	0.3003 ± 0.0008	0.1582 ± 0.0005
7	2.00	1.50	0.2872 ± 0.0007	0.1567 ± 0.0005
8	2.50	1.50	0.2722 ± 0.0007	0.1529 ± 0.0004
9	3.00	1.50	0.2620 ± 0.0006	0.1509 ± 0.0004
10	3.50	1.50	0.2531 ± 0.0006	0.1492 ± 0.0004
11	3.00	1.25	0.3011 ± 0.0007	0.1782 ± 0.0005
12	3.25	1.25	0.2949 ± 0.0007	0.1767 ± 0.0005
13	3.30	1.25	0.2937 ± 0.0007	0.1764 ± 0.0005
14	3.20	1.25	0.2953 ± 0.0007	0.1764 ± 0.0005
15	3.20	1.24	0.2974 ± 0.0007	0.1779 ± 0.0005

A few things are apparent from [Example 7.1](#). First, the retrieved parameter values may not be very precise. In fact, the best one can do is determine by (a) the uncertainties inherent in the measured values $\mathcal{R}_j$, and (b) the precision of the Monte Carlo simulation, which is dependent on the number of histories N . Second, a relatively small number of histories ($N = 100,000$ in this case) often is sufficient to obtain results of practical relevance. Finally, however, many iterations might be required, especially as the number of unknown parameters increases.

The example above is in the form of an integral. However, it is just as likely that a random process, such as X-ray attenuation by an unknown shield, could produce one or more responses. Understanding the processes for X-ray scattering and absorption (see [Chapter 9](#)), one can in principle simulate the process for assumed sample variables (such as shield composition and thickness) until the simulated responses match the measured responses sufficiently well. Thus, it is possible to employ iterative IMC to solve many inverse problems. Many of these applications require a knowledge of the specific processes that are involved. A few such applications are briefly discussed in [Section 7.6](#).

7.3 Symbolic Monte Carlo

An alternative approach, previously referred to as inverse Monte Carlo or IMC [[Dunn 1981](#); [Dunn 2006](#)] but here and elsewhere [[Dunn and Shultis 2009](#)] called *symbolic Monte Carlo* or SMC, proceeds formally as follows. Instead of sampling from the PDF $f(\mathbf{x}, \boldsymbol{\theta})$ for the value of $\boldsymbol{\theta}$ selected on a given iteration, sample from an assumed

PDF $f^*(\mathbf{x})$, also defined on V , and weight the histories appropriately, as in importance sampling. The weight factor for history i is of the form $f(\mathbf{x}_i, \boldsymbol{\theta})/f^*(\mathbf{x}_i)$ and thus contains the unknown parameters $\boldsymbol{\theta}$ as symbols. Proceeding formally, one obtains

$$\bar{R}(\mathbf{y}_k, \boldsymbol{\theta}) = \frac{1}{N} \sum_{i=1}^N z(\mathbf{x}_i, \mathbf{y}_k, \boldsymbol{\theta}) \frac{f(\mathbf{x}_i, \boldsymbol{\theta})}{f^*(\mathbf{x}_i)}, \quad k = 1, 2, \dots, K. \quad (7.9)$$

Substitution of measured or specified responses $\mathcal{R}_k$ for $\bar{R}(\mathbf{y}_k, \boldsymbol{\theta})$ produces a system of equations in which all quantities are known except the $\boldsymbol{\theta}$. Note that the assumed PDF f^* is independent of $\boldsymbol{\theta}$ and, thus, can be sampled without regard to the unknown $\boldsymbol{\theta}$. In principle, Eq. (7.9) forms a system of K algebraic equations for the J unknown parameters of $\boldsymbol{\theta}$.

In the underdetermined case, where $K < J$, there are an infinite number of solutions and additional constraints must be applied to obtain a single solution. If $K = J$, then, in principle, these equations can be solved for the unknown $\boldsymbol{\theta}$. In the overdetermined case, where $K > J$, methods such as least-squares can be applied (see Press et al. [1996]) to obtain a solution that best fits all the responses. The system of K equations given by Eq. (7.9) may require that a numerical procedure, such as trial and error or least-squares, be employed in order to find the unknown $\boldsymbol{\theta}$; however, in SMC only *one* simulation needs to be performed. When simulations are lengthy, such as in highly multidimensional cases, the SMC approach offers an attractive alternative to the iterative IMC approach.

At first, it may seem odd that symbols are inserted into history scores because this has not been the case in other methods of Monte Carlo considered to this point. However, there is no reason that this cannot be done. Of course, the insertion of the symbols $\theta_1, \theta_2, \dots, \theta_J$ into the Monte Carlo scores can mean that the system of Eq. (7.9) is a very complex set of equations. Nevertheless, as is to be seen, there are cases where this procedure can be successfully implemented.

7.3.1 Uncertainties in Retrieved Values

The SMC method can be used not only to generate a system of equations for the desired variables but also to estimate, in principle, the uncertainties in the estimated values of the parameters [Dunn 1981]. The system of equations used to determine values of the parameters θ_j in the SMC process are obtained by equating Eq. (7.9) to measured or specified responses $\mathcal{R}_k$, i.e.,

$$\bar{R}(\mathbf{y}_k, \boldsymbol{\theta}) = \mathcal{R}_k, \quad k = 1, 2, \dots, K. \quad (7.10)$$

The individual estimates of θ_j , $j = 1, 2, \dots, J$, can then be written, at least conceptually, as

$$\theta_j = G_j(\mathbf{y}, \bar{\mathbf{R}}, \mathcal{R}), \quad j = 1, 2, \dots, J, \quad (7.11)$$

where G_j is the function, usually unknown, that defines the solution of Eq. (7.10) for the parameter θ_j .

Even if it is assumed that the y_k have no uncertainty, the measurements $\mathcal{R}_k$ presumably have some uncertainty, quantified by their standard deviations $\sigma(\mathcal{R}_k)$, that affects the uncertainty in the retrieved values of the θ_j . Likewise, the MC evaluation of $\bar{R}_k$ has an inherent uncertainty that affects the uncertainty in the calculated values of θ_j . Recall that the central limit theorem indicates that, for sufficiently large N , a direct Monte Carlo estimate $\bar{R}_k$ of the integral of Eq. (7.4), with θ given, has a standard deviation $\sigma(z)/\sqrt{N}$, where σ is the population standard deviation of $z(\mathbf{x}, \mathbf{y}_k, \theta)$, with respect to the PDF f . Substitution of the sample standard deviation for the population standard deviation leads to an expression for the standard deviation of the Monte Carlo estimate of the form

$$\sigma(\bar{R}_k) = \frac{\sigma(z)}{\sqrt{N}} \simeq \left[\frac{1}{N-1} \left(\overline{R_k^2} - \bar{R}_k^2 \right) \right]^{1/2}, \quad k = 1, 2, \dots, K. \quad (7.12)$$

With these uncertainties for $\mathcal{R}_k$ and $\bar{R}_k$ the uncertainty in θ_j can be estimated. If the K measurements $\mathcal{R}_k$ and the K values of $\bar{R}_k$ are all independent, then one can use the standard propagation of errors formula (e.g., see Bevington [1969]) to estimate the standard deviations of the individual recovered parameters due to uncertainties in the measured responses and the $\bar{R}_k$ by

$$\sigma(\theta_j) = \left\{ \sum_{k=1}^K \left(\frac{\partial G_j}{\partial \mathcal{R}_k} \right)^2 \sigma^2(\mathcal{R}_k) + \left(\frac{\partial G_j}{\partial \bar{R}_k} \right)^2 \sigma^2(\bar{R}_k) \right\}^{1/2}, \quad j = 1, 2, \dots, J.$$

(7.13)

However, in inverse problems, θ is not known until the inverse problem has been solved, making estimation of the $\sigma(\bar{R}_k)$ problematic. Further, the function G_j often is quite complex or it may be impossible even to formulate it, except through the algorithmic process used to solve Eq. (7.10) numerically. In these cases the evaluation of $\partial G_j / \partial \mathcal{R}_k$ and $\partial G_j / \partial \bar{R}_k$ must rely on approximations. Only for the simplest cases can G_j be explicitly determined (e.g., see Examples 7.2 and 7.7). Thus, the above formal treatment for estimating the uncertainties in the retrieved parameters is generally difficult to implement fully. Instead, approximate methods must be employed. For instance, in some cases it may be easy to evaluate $\partial G_j / \partial \mathcal{R}_k$, but difficult or impossible to evaluate $\partial G_j / \partial \bar{R}_k$. In such cases, the errors due only to uncertainties in the measurements $\mathcal{R}_k$, $k = 1, 2, \dots, K$ can be estimated as lower bounds to the $\sigma(\theta_j)$, $j = 1, 2, \dots, J$, i.e.,

$$\sigma(\theta_j) \geq \left\{ \sum_{k=1}^K \left(\frac{\partial G_j}{\partial \mathcal{R}_k} \right)^2 \sigma^2(\mathcal{R}_k) \right\}^{1/2}, \quad j = 1, 2, \dots, J. \quad (7.14)$$

Of course, this estimate will be largely unaffected by the number of histories run because the uncertainties in the Monte Carlo summations are ignored. Nevertheless, lower bounds on the uncertainties in the retrieved parameters can be estimated, and these estimates may be fairly good if N is large. An example that implements this process is given later in [Example 7.7](#).

An obvious alternate approach to determine how the parameter values θ_j depend on uncertainties in $\mathcal{R}_k$ and $\bar{\mathcal{R}}_k$ is to perform a sensitivity analysis. This approach requires that the appropriate summations used to form the function $\bar{\mathcal{R}}_k$ in the SMC simulation be saved. Then, one determines how the resulting values of θ_j , $j = 1, 2, \dots, J$, vary with changes in the $\mathcal{R}_k$, and these summations are varied by small amounts (say, 1 or 2 percent). Such a procedure provides an indication of how sensitive the retrieved parameters are to variations in uncertainties in the response values and the Monte Carlo model terms.

7.3.2 The PDF Is Fully Known

The particular implementation of SMC depends on where and how the unknown parameters are located in the kernel and the PDF. If all of the unknown parameters are contained only in the kernel, the implementation is inherently straightforward. If, on the other hand, the PDF is dependent on some or all of the unknown parameters, then the implementation is more complex; this situation is considered in [Section 7.3.3](#).

When the PDF is fully known, i.e., f is independent of θ and only the kernel z in [Eq. \(7.4\)](#) is dependent on the unknown parameters, the inverse problem simplifies considerably. In fact, any of a number of quadrature schemes, including Monte Carlo, can be applied to such inverse problems. If the PDF is easy to sample, then one might as well sample it directly, in which case [Eq. \(7.9\)](#) simplifies to

$$\bar{\mathcal{R}}_k(\theta) = \frac{1}{N} \sum_{i=1}^N z(\mathbf{x}_i, \mathbf{y}_k, \theta), \quad k = 1, 2, \dots, K. \quad (7.15)$$

In those cases where $f(\mathbf{x})$ is difficult to sample directly, MCMC can be employed or one can use another PDF $f^*(\mathbf{x})$ that is more easily sampled and [Eq. \(7.9\)](#) reduces to the form

$$\bar{\mathcal{R}}_k(\theta) = \frac{1}{N} \sum_{i=1}^N z(\mathbf{x}_i, \mathbf{y}_k, \theta) \frac{f(\mathbf{x}_i)}{f^*(\mathbf{x}_i)}, \quad k = 1, 2, \dots, K. \quad (7.16)$$

In some cases, the system of equations given by [Eqs. \(7.15\)](#) or [\(7.16\)](#) can be solved by some direct procedure, such as matrix inversion. More often, an iterative numerical procedure must be employed. However, the iterations do not involve further Monte Carlo simulations.

A simple linear kernel

For the purpose of illustration, consider the simple case where the random variable $\mathbf{x}$ is a single variable x , the vector of unknowns is a single parameter θ , and z has the

simple linear form

$$z(\mathbf{x}, \mathbf{y}_k, \boldsymbol{\theta}) = z(x, \theta) = \theta x, \quad (7.17)$$

which indicates that z has no dependence on $\mathbf{y}$. Further, assume the PDF $f(x)$ is uniform over $[a, b]$. Then Eq. (7.4) reduces to

$$R(\theta) = \int_a^b \theta x \frac{1}{b-a} dx. \quad (7.18)$$

It is trivial to sample from $f(x) = 1/(b-a)$, and each history is scored with the product θx . Then, Eq. (7.15) takes the simple form

$$\bar{R}(\theta) = \frac{\theta}{N} \sum_{i=1}^N x_i = \theta \bar{x}. \quad (7.19)$$

The known or measured $\mathcal{R}$ can be substituted for $\bar{R}(\theta)$, and the resulting equation can be solved for θ , namely,

$$\theta = \frac{\mathcal{R}}{\bar{x}}. \quad (7.20)$$

In this trivial case, the estimation of $\sigma(\theta)$ by Eq. (7.13) is straightforward and leads to

$$\sigma(\theta) = \frac{1}{\bar{x}} [\sigma^2(\bar{R}) + \sigma^2(\mathcal{R})]^{1/2}, \quad (7.21)$$

where

$$\sigma^2(\bar{R}) = \left[\frac{\theta^2}{N-1} (\overline{x^2} - \bar{x}^2) \right]. \quad (7.22)$$

Numerical results for this simple case are given in Example 7.2.

Example 7.2 A One-Dimensional Inversion

Consider the particular case where $a = 1$, $b = 3$, and $\theta = 4$. Then, Eq. (7.18) can be used to show that $R(\theta) = 8$. However, assume a measured value of $\mathcal{R} = 7.98 \pm 0.05$. Then, a single forward Monte Carlo simulation can be used to estimate $\bar{x}$ and $\overline{x^2}$, and it is possible to solve Eq. (7.20) for θ and Eq. (7.21) for $\sigma(\theta)$. This problem was simulated and the results are shown in Table 7.2 for three different values of N . It is seen that the solution approaches the correct value of $\theta = 4$ and the uncertainties decrease as $1/\sqrt{N}$.

Table 7.2 A symbolic Monte Carlo solution for θ .

N	$\bar{x}$	θ	$\sigma(\theta)$	N	$\bar{x}$	θ	$\sigma(\theta)$	N	$\bar{x}$	θ	$\sigma(\theta)$
10^2	1.987	4.015	0.115	10^4	2.001	3.989	0.012	10^6	2.000	3.990	0.003

The above procedure can be extended, generally, to any kernel that is linear in the unknown parameters θ . In most cases, estimation of the $\sigma(\theta_j)$ is more difficult, but estimation of the unknown parameters is numerically straightforward, as [Example 7.3](#) illustrates.

Example 7.3 A Three-Parameter Linear Kernel

Consider a system of which the response can be modeled as

$$R(y, \theta_1, \theta_2, \theta_3) = \int_1^4 \int_0^2 z(x_1, x_2, y, \theta_1, \theta_2, \theta_3) f(x_1, x_2) dx_2 dx_1,$$

where the kernel has the form

$$z(x_1, x_2, y, \theta_1, \theta_2, \theta_3) = \theta_1 x_1 / y + \theta_2 y e^{-2x_2} + \theta_3 x_1 x_2 y^2.$$

The joint PDF is the bivariate normal (see Appendix A.3.1)

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{\xi}{2(1-\rho^2)}\right\},$$

with

$$\xi = \frac{(x_1 - \mu_1)^2}{\sigma_1^2} - \frac{2\rho(x_1 - \mu_1)(x_2 - \mu_2)}{\sigma_1\sigma_2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2},$$

and ρ is the correlation coefficient. Here, it is assumed that $\mu_1 = 0$, $\mu_2 = 1$, $\sigma_1 = 1$, $\sigma_2 = 2.5$, and $\rho = 0.5$. How might one use SMC to obtain θ_1 , θ_2 , and θ_3 , given measurements $\mathcal{R}_1 = 0.17407$, $\mathcal{R}_2 = 0.36406$, and $\mathcal{R}_3 = 0.73845$ for $y_1 = 1$, $y_2 = 2$, and $y_3 = 3$, respectively?

The SMC formulation of [Eq. \(7.9\)](#) leads to

$$\bar{R}(y_k | \theta_1, \theta_2, \theta_3) = \frac{1}{N} \sum_{i=1}^N z(x_{1i}, x_{2i}, y_k | \theta_1, \theta_2, \theta_3) \frac{f(x_{1i}, x_{2i})}{f^*(x_{1i}, x_{2i})}, \quad k = 1, 2, 3.$$

If, for simplicity, x_1 and x_2 are sampled uniformly over their respective intervals, then $f^*(x_1, x_2) = (1/3)(1/2) = 1/6$, and these equations can be written in terms of the measurements as

$$\mathcal{R}_k = \frac{6}{N} \left\{ \frac{\theta_1}{y_k} \mathcal{S}_1 + \theta_2 y_k \mathcal{S}_2 + \theta_3 y_k^2 \mathcal{S}_3 \right\}, \quad k = 1, 2, 3,$$

where the summations $\mathcal{S}_1 = \sum_{i=1}^N x_{1i} f(x_{1i}, x_{2i})$, $\mathcal{S}_2 = \sum_{i=1}^N e^{-2x_{2i}} f(x_{1i}, x_{2i})$, and $\mathcal{S}_3 = \sum_{i=1}^N x_{1i} x_{2i} f(x_{1i}, x_{2i})$ are easily scored. A single simulation can be run in which the sums $\mathcal{S}_1$, $\mathcal{S}_2$, and $\mathcal{S}_3$ are accumulated, and the solution can be obtained by matrix inversion of $\mathcal{R} = \mathcal{A}\mathcal{S}$. For $N = 10^6$ histories, the following results were obtained: $\theta_1 = 1.505$, $\theta_2 = 0.802$, and $\theta_3 = 1.004$. The actual responses were developed assuming $\theta_1 = 1.500$, $\theta_2 = 0.800$, and $\theta_3 = 1.000$ and, thus, the SMC procedure obtained results good to within 0.5%. This problem demonstrates that, if the kernel is linear in the unknown parameters θ and the PDF is known, the SMC procedure can be reduced to a matrix inversion.

It is possible to solve the equations in the two previous examples directly or by matrix inversion because the kernels z are linear in the unknown parameters and f is completely known. However, in many cases the kernel is nonlinear in θ , and the SMC approach leads to a system of equations that must be solved by some numerical technique. An extreme case is one in which each history leads to a score of which the value is unique and thus the SMC model produces a system of equations each of which has N terms. Such a case is illustrated in [Example 7.4](#).

Example 7.4 A Nonlinear Kernel

Consider the definite integral

$$R(y, \theta_1, \theta_2) = \frac{1}{\sqrt{2\pi}\theta_2} \int_a^b e^{-[(x-y\theta_1)/(\sqrt{2}\theta_2)]^2} \lambda e^{-\lambda x} dx,$$

which can be converted to the form

$$R(y, \theta_1, \theta_2) = \frac{\lambda}{\sqrt{2\pi A\theta_2}} \int_a^b e^{-[(x-y\theta_1)/(\sqrt{2}\theta_2)]^2} A e^{-\lambda x} dx,$$

where λ is known and $A = \lambda/(e^{-\lambda a} - e^{-\lambda b})$ is a normalization constant. Samples from the PDF $f(x) = A e^{-\lambda x}$ are of the form

$$x_i = -\frac{1}{\lambda} \ln[e^{-a\lambda} - \rho_i(e^{-a\lambda} - e^{-b\lambda})],$$

where ρ_i is a pseudorandom number uniformly distributed on the unit interval. Next, form the SMC estimates, from [Eq. \(7.15\)](#), to obtain

$$\bar{R}_k(\theta_1, \theta_2) = \frac{\lambda}{\sqrt{2\pi A\theta_2}} \frac{1}{N} \sum_{i=1}^N e^{[(x_i - y_k \theta_1)/(\sqrt{2}\theta_2)]^2}, \quad k = 1, 2, \dots, K.$$

Substitution of the known $\mathcal{R}_k$ for $\bar{R}_k$ yields K equations in two unknowns. The model is not linear in the unknowns θ_1 and θ_2 , and the x_i assume arbitrary values over the interval $[a, b]$. Thus, each term in the summations has a unique value. If the x_i values are stored, then the above equations, each with N terms on the right side, can be solved numerically for θ_1 and θ_2 . This solution requires numerical iteration, but only one Monte Carlo simulation needs to be implemented.

For the particular case that $a = 2, b = 10, \theta_1 = 3, \theta_2 = 2, \lambda = 1, K = 2, y_1 = 1$, and $y_2 = 2$, calculated responses are $\mathcal{R}_1 = 0.02458$ and $\mathcal{R}_2 = 0.009157$. Given these response values, an SMC model was run and the above two equations were solved for θ_1 and θ_2 . The results shown in Table 7.3 were obtained by specifying values of θ_1 and θ_2 and comparing the calculated $\bar{\mathcal{R}}_k$ to the specified $\mathcal{R}_k$. The search was stopped when $[(\bar{\mathcal{R}}_1 - \mathcal{R}_1)^2 + (\bar{\mathcal{R}}_2 - \mathcal{R}_2)^2]^{1/2}/(\mathcal{R}_1 + \mathcal{R}_2) < 0.01$. Of course, other quadrature formulas could be used to obtain similar results. However, this problem demonstrates that even if each history leads to a unique term in the SMC model, it still may be possible to utilize SMC to obtain solutions to inverse problems utilizing only a single simulation.

Table 7.3 The retrieved parameters θ_1 and θ_2 for various values of N .

N	θ_1	θ_2	N	θ_1	θ_2	N	θ_1	θ_2
10^2	2.980	2.035	10^4	2.998	1.996	10^6	2.998	1.997

7.3.3 The PDF Is Unknown

The SMC solution for the general case, in which both the kernel and the PDF depend on some set of unknown parameters θ , is treated by Eq. (7.9). Section 7.3.2 considered the simpler case where the PDF has no dependence on the unknown parameters. The case, in which the kernel is independent of the θ but the PDF is not, is now considered. When some parameters of the PDF are unknown or the dependence of the PDF on the parameters is unknown, implementation of SMC becomes more involved. The complexity depends on the specifics of the problem. For instance, a situation in which the PDF is linear in the unknown parameters (not necessarily in the random variables) may be simpler than a situation in which the dependence on the parameters is nonlinear. Some approaches that can be applied in certain cases are considered next.

Approximating an unknown PDF as a histogram

In some cases, it is possible to linearize the search for an unknown PDF by using a histogram approach. For instance, consider a response model of the form

$$R(y) = \int_a^b z(x, y) f(x) dx, \quad (7.23)$$

where the PDF and its dependence on hidden parameters are unknown but responses are known at K values of y . In this inverse problem one seeks to infer the shape of the PDF from the K responses. If K is large enough, then one can divide the interval $[a, b]$ into $J \leq K$ subintervals and express the PDF in histogram form. For instance, if each subinterval of a one-dimensional PDF is of width $\Delta = (b - a)/J$, then the PDF can be

expressed as

$$\begin{aligned}
 f(x) &= f_1, & a \leq x < a + \Delta \\
 &= f_2, & a + \Delta \leq x < a + 2\Delta \\
 &\vdots \\
 &= f_J, & a + (J - 1)\Delta \leq x \leq b.
 \end{aligned} \tag{7.24}$$

Sampling from a known PDF f^* , an SMC solution can be expressed in the form

$$\bar{R}_k = \frac{1}{N} \sum_{i=1}^N z(x_i, y_k) \frac{f(x_i)}{f^*(x_i)}. \tag{7.25}$$

On any given history, x_i is in only one subinterval; if this subinterval is called j , then the contribution to the k th response for that history is of the form

$$\delta R_k = z(x_i, y_k) \frac{f_j}{f^*(x_i)}. \tag{7.26}$$

Everything in this equation is known except f_j . Thus, a system of linear equations of the form

$$\mathbf{A}\mathbf{f} = \mathbf{R} \tag{7.27}$$

results, where $\mathbf{f} = \{f_j, j = 1, 2, \dots, J\}$ and

$$\mathbf{A} = \{S_{kj}\}, \quad k = 1, 2, \dots, K, \quad j = 1, 2, \dots, J,$$

with

$$S_{kj} = \frac{1}{N} \sum_{i=1}^N \frac{z(x_i, y_k)}{f^*(x_i)}, \quad a + (j - 1)\Delta \leq x < j\Delta. \tag{7.28}$$

For the case $J = K$, Eq. (7.27) can be solved, in principle, by direct matrix inversion. However, it may be the case that many measurements are available (K is large) and fewer histogram values ($J < K$) are adequate to characterize the PDF. In such cases, it may be possible to find values of the $J < K$ parameters of the PDF by a numerical procedure such as least-squares [Press et al. 1996]. In practice, the uncertainties in the data and the uncertainties inherent in the Monte Carlo summations often lead to a system of equations of which the solution by these standard techniques fits the data fairly well but is not physically reasonable. For instance, some PDF values may be negative.

Example 7.5 A Linearized Quadratic PDF

For purposes of illustration, consider Eq. (7.23) with $a = 0$, $b = \pi$, and $z(x, y) = 1000/[y + 0.5 \cos(x)]$. If the PDF has the quadratic form

$$f(x) = 0.640819 - 0.025633(x - 5)^2,$$

then it can be shown, by numerical integration, that for $J = K = 4$ this model leads to the following responses: $\mathcal{R}_1(y_1 = 1.0) = 1,367.25$, $\mathcal{R}_2(y_2 = 1.5) = 791.007$, $\mathcal{R}_3(y_3 = 2.0) = 561.927$, and $\mathcal{R}_4 = \bar{\mathcal{R}}(y_4 = 2.5) = 436.943$.

Assume one is unaware of how these responses were obtained and one is asked to estimate f_j , $j = 1, 2, 3, 4$. One might choose to sample from the uniform PDF $f^*(x) = 1/\pi$, $0 \leq x \leq \pi$ and then obtain a system of equations of the form

$$\mathcal{S}_{k1}f_1 + \mathcal{S}_{k2}f_2 + \mathcal{S}_{k3}f_3 + \mathcal{S}_{k4}f_4 = \mathcal{R}_k, \quad k = 1, 2, 3, 4,$$

where

$$\mathcal{S}_{kj} = \frac{\pi}{N} \sum_{i=1}^N \frac{1000}{y_k + 0.5 \cos(x_i)}, \quad (j-1)\pi/4 \leq x_i \leq j\pi/4.$$

With $N = 10^7$ histories sampled from the uniform PDF on $[0, \pi]$ and evaluation of the summations $\mathcal{S}_{ki}$, $i = 1, \dots, 4$, the following results were obtained: $f_1 = 0.07104$, $f_2 = 0.2848$, $f_3 = 0.3939$, and $f_4 = 0.5234$. These values are plotted, along with the actual PDF, in Fig. 7.2. This example demonstrates that a reasonable approximation can be obtained by this linearization procedure.

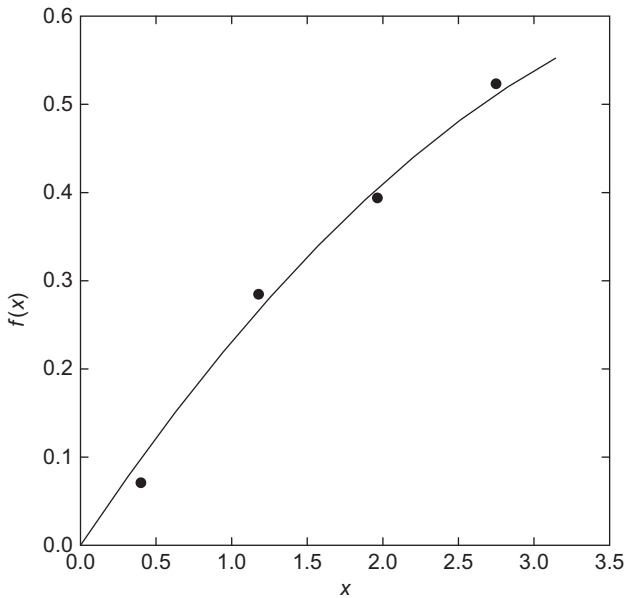


Figure 7.2 The true quadratic PDF (line) and the four-component histogram values (points) recovered using SMC.

If, for example, $K = 10$ is used in [Example 7.5](#) and ten responses are given, one would expect to be able to retrieve ten histogram values. However, standard numerical procedures employed to find $4 < J < 10$ PDF histogram values did not give results better than those shown in [Fig. 7.2](#). It is obvious that solutions that contain physically unreasonable values should be ignored and other solutions sought until all the PDF values appear to be reasonable. Even in cases where all of the PDF parameters are positive, it is possible, because of uncertainties in the data and the Monte Carlo summations, that the retrieved parameters are not very accurate. It may be useful to try several values of J and see if the trend in the histogram of the PDF remains consistent. It also deserves mention that various variance reduction methods can, and often should, be employed. For instance, rather than sampling over $[a, b]$ and then determining the subinterval within which the x_i exists, it is better to allocate $1/J$ of the histories to each subinterval, in the systematic sampling manner.

Many quantities have a PDF-like character in that they can be decomposed into a magnitude times a shape function of which the integral is normalized to unity. For instance, in spectroscopy, a spectrum (often consisting of $K > 10^3$ response values) can be written in terms of the source strength, the PDF on source energy, and the detector response function, as in [Eq. \(7.5\)](#). In cases such as these, one can treat the magnitude (e.g., Q_o in [Eq. \(7.5\)](#)) as an additional unknown and solve the system of K equations for the magnitude and $J - 2$ PDF histogram values (there are $J - 1$ of these values but one is determined by the PDF normalization condition). Alternatively, rather than perform matrix inversion, one can solve an overdetermined system, where $K > J + 1$ by methods such as least-squares [[Press et al. 1996](#)].

The PDF is separable

Sometimes $f(\mathbf{x}, \boldsymbol{\theta})$ can be approximated by a finite series of which the terms are products of separable terms, i.e., one term contains $\mathbf{x}$ and the other $\boldsymbol{\theta}$. In such cases, the PDF can be written as

$$f(\mathbf{x}, \boldsymbol{\theta}) \simeq \sum_{m=0}^M h_m(\boldsymbol{\theta}) g_m(\mathbf{x}), \quad (7.29)$$

where the functions $g_j(\mathbf{x})$ and $h_j(\boldsymbol{\theta})$ are known and $\boldsymbol{\theta}$ has J unknown components. Generally, the larger M the better is the approximation.

In such cases, a general inverse problem for $\boldsymbol{\theta}$ can be written as

$$R(\mathbf{y}_k, \boldsymbol{\theta}) = \int_V z(\mathbf{x}, \mathbf{y}_k) \sum_{m=0}^M h_m(\boldsymbol{\theta}) g_m(\mathbf{x}) d\mathbf{x}. \quad (7.30)$$

A symbolic Monte Carlo solution could, in general, take the form

$$\bar{R}_k = \frac{1}{N} \sum_{i=1}^N z(\mathbf{x}_i, \mathbf{y}_k) \frac{f(\mathbf{x}_i, \boldsymbol{\theta})}{f^*(\mathbf{x}_i)} = \sum_{m=0}^M h_m(\boldsymbol{\theta}) S_{km}, \quad k = 1, 2, \dots, K, \quad (7.31)$$

where

$$S_{km} = \frac{1}{N} \sum_{i=1}^N z(\mathbf{x}_i, \mathbf{y}_k) \frac{g_m(\mathbf{x}_i)}{f^*(\mathbf{x}_i)}, \quad k = 1, 2, \dots, K, \quad (7.32)$$

and where $f^*(x)$ is the PDF used to sample values of x_i .

Upon setting Eq. (7.31) to measured or specified responses $\mathcal{R}_k$, a set of K equations in the J unknown values of θ_j is obtained, namely,

$$\mathcal{R}_k = \sum_{m=0}^M h_m(\theta) S_{km}, \quad k = 1, 2, \dots, K. \quad (7.33)$$

To avoid over- and underdetermined sets of equations usually K is chosen to equal J . However, even in this case nonlinear numerical techniques generally must be used to obtain a solution.

A PDF of one random variable

A PDF with a single random variable x can always be cast into a separable PDF by using a truncated Taylor series expansion about some point x_o , namely,

$$f(x, \theta) \simeq \sum_{m=0}^M \frac{(x - x_o)^m}{m!} \left. \frac{\partial^m f(x, \theta)}{\partial x^m} \right|_{x=x_o}. \quad (7.34)$$

This has the form of Eq. (7.29) with

$$h_m(\theta) = \left. \frac{\partial^m f(x, \theta)}{\partial x^m} \right|_{x=x_o} \quad \text{and} \quad g_m(x) = \frac{x^m}{m!}. \quad (7.35)$$

Exponential PDF

To illustrate the above separable PDF technique consider the exponential PDF $f(x, \theta) = \theta e^{-\theta x}$ for $x \in [0, \infty)$. Exponential PDFs are encountered in many physical processes. For instance, this PDF, which is nonlinear in the parameter θ , applies to many physical phenomena, such as radioactive decay, neutral particle transport, and growth and decay of biological systems.

With an M term Taylor series expansion of Eq. (7.34) about $x_o = 0$, this PDF can be approximated as

$$f(x, \theta) \simeq \sum_{m=0}^M h_m(\theta) g_m(x), \quad (7.36)$$

where

$$h_m(\theta) = \frac{d^m}{dx^m} (\theta s^{-\theta x}) \Big|_{x=0} = (-1)^m \theta^{m+1} \quad \text{and} \quad g_m(x) = \frac{x^m}{m!}. \quad (7.37)$$

Now consider a model of the form

$$R(y, \theta) = \int_0^\infty z(x, y) e^{-\theta x} dx, \quad (7.38)$$

where z is a kernel for which the integral exists. Assume $\mathcal{R}_1$ is known or can be determined for some value y_1 . How can SMC be used to obtain the unknown parameter θ in a single simulation? One way to proceed is to rewrite the above equation as

$$R_1(\theta) = \frac{1}{\theta} \int_0^\infty z(x, y_1) \theta e^{-\theta x} dx. \quad (7.39)$$

Because θ is unknown, it is not possible to sample from this PDF directly, but it is possible to sample from $f^*(x) = \theta_o e^{-\theta_o x}$, where θ_o is known. Doing so, an estimate of $\bar{R}_1(\theta)$ can be constructed as

$$\bar{R}_1(\theta) = \frac{1}{\theta} \sum_{m=0}^M h_m(\theta) \left[\frac{1}{N} \sum_{i=1}^N z(x_i, y_1) \frac{g_m(x_i)}{f^*(x_i)} \right], \quad (7.40)$$

where the x_i are sampled from $f^*(x)$. Substitution for f^* , h_m , and g_m then gives an equation from which θ can be calculated, namely,

$$\mathcal{R}_1 \simeq \bar{R}_k = \frac{1}{\theta_o N} \sum_{m=0}^M \theta^m \left[\sum_{i=1}^N \frac{z(x_i, y_1) x_i^m}{m! e^{-\theta_o x_i}} \right]. \quad (7.41)$$

This equation is a polynomial equation of degree M for the unknown θ and generally has M solutions. Only solutions that are real and positive are of interest. If more than one real positive solution is found (the nonuniqueness of inverse problems), then the one that is most physically realistic is chosen.

Example 7.6 A Specific Exponential PDF

Consider the particular case

$$R(\theta) = \int_0^\infty x^2 e^{-\theta x} dx,$$

where it can be argued on physical grounds that $0.5 \leq \theta \leq 1.5$. This integral has a known value for a given θ of $R(\theta) = 2/\theta^3$. However, for purposes of illustration, this fact is ignored and one is asked to use SMC to determine the value of θ corresponding to a given measured

value $\mathcal{R} = 3.90625$. A five-term expansion of the exponential function can be used with $\theta_o = 1$. Because

$$\frac{f(x_i, \theta)}{f^*(x_i)} = \theta e^{-(\theta-1)x_i},$$

the Taylor series expansion of this exponential can be written as

$$e^{-(\theta-1)x_i} \simeq 1 + \sum_{j=1}^5 \frac{(-(\theta-1)x_i)^j}{j!}.$$

Then, the SMC solution can be written as

$$\begin{aligned} 3.90625 = \frac{1}{N} \left\{ \sum_{i=1}^N x_i^2 - (\theta-1) \sum_{i=1}^N x_i^3 + \frac{(\theta-1)^2}{2} \sum_{i=1}^N x_i^4 - \frac{(\theta-1)^3}{6} \sum_{i=1}^N x_i^5 \right. \\ \left. + \frac{(\theta-1)^4}{24} \sum_{i=1}^N x_i^6 - \frac{(\theta-1)^5}{120} \sum_{i=1}^N x_i^7 \right\}. \end{aligned} \quad (7.42)$$

A single simulation allows estimations of the six summations on the right side of the above equation. This equation for θ can then be solved by any of several numerical techniques. For example, $N = 10^6$ histories were used to estimate the six summations, and then a simple numerical search with θ varying from 0.5 to 1.5 in increments of 0.01 was conducted. Linear interpolation between the two closest points yielded $\theta = 0.798$ (the correct value is $\theta = 0.800$). Note that a five-term expansion was sufficient because $\theta - 1$ is small. In general, the larger the difference between θ and θ_o , the more terms may be required.

The PDF is linear in the unknown parameters

The simplest linear PDF defined on a finite interval $[a, b]$ is the uniform PDF. Such a PDF has no adjustable parameters because its value is constrained to be $f(x) = 1/(b-a)$. More generally, the specific form of Eq. (7.29) in which $h_j(\theta) = \theta_j$, $j = 0, 1, 2, \dots, J$ leads to a PDF of the form

$$f(\mathbf{x}, \theta) = \sum_{j=0}^J \theta_j g_j(\mathbf{x}), \quad \mathbf{x} \in V, \quad (7.43)$$

which is linear in the parameters θ_j , $j = 0, 1, 2, \dots, J$. The normalization condition can be used to relate one of the unknown parameters, say θ_0 , to the others, yielding

$$\theta_0 = \sum_{j=1}^J \theta_j \int_V g_j(\mathbf{x}) d\mathbf{x} \bigg/ \int_V g_0(\mathbf{x}) d\mathbf{x}.$$

Thus, in principle, only $K = J$ responses are needed from a response model of the form of Eq. (7.4) in order to construct a symbolic Monte Carlo model to estimate the $J + 1$ unknown parameters.

For such a PDF, Eq. (7.4) assumes the form

$$\mathcal{R}_k \simeq R(\mathbf{y}_k, \boldsymbol{\theta}) = \int_V z(\mathbf{x}_i, \mathbf{y}_k) \sum_{j=0}^J \theta_j g_j(\mathbf{x}) d\mathbf{x}, \quad k = 1, 2, \dots, K, \quad (7.44)$$

which is an inverse problem for $\boldsymbol{\theta}$ if at least one $\mathcal{R}_k$ is known. The general SMC estimate of R can be written as

$$\begin{aligned} \bar{R}_k &= \frac{1}{N} \sum_{i=1}^N z(\mathbf{x}_i, \mathbf{y}_k) \frac{f(\mathbf{x}, \boldsymbol{\theta})}{f^*(\mathbf{x}_i)} \\ &= \sum_{j=0}^J \theta_j \mathcal{S}_{kj}, \quad k = 1, 2, \dots, K, \end{aligned} \quad (7.45)$$

where

$$\mathcal{S}_{kj} = \frac{1}{N} \sum_{i=1}^N z(\mathbf{x}_i, \mathbf{y}_k) \frac{g_j(\mathbf{x}_i)}{f^*(\mathbf{x}_i)}, \quad k = 1, 2, \dots, K. \quad (7.46)$$

A one-dimensional linear PDF

For instance, consider the linear PDF, $f(x) = \alpha + \beta x$, for which the normalization condition $\int_a^b (\alpha + \beta x) dx = 1$ can be used to show that

$$\alpha = \frac{1}{b-a} - \beta \frac{b+a}{2}. \quad (7.47)$$

Then the integral

$$R(y_k, \alpha, \beta) = \int_a^b z(x, y_k) (\alpha + \beta x) dx, \quad k = 1, 2, \dots, K \quad (7.48)$$

poses an inverse problem for α and β if $\mathcal{R}_k$ is known for at least one value of k . An SMC solution, from Eq. (7.9), is given by

$$\begin{aligned} \mathcal{R}_k &= \frac{1}{N} \sum_{i=1}^N z(x_i, y_k) \left[\frac{f(x_i)}{f^*(x_i)} \right], \\ &= \frac{1}{b-a} \mathcal{S}_{1k} + \beta \mathcal{S}_{2k} - \beta \frac{b+a}{2} \mathcal{S}_{1k}, \quad k = 1, 2, \dots, K, \end{aligned} \quad (7.49)$$

where

$$S_{1k} = \frac{1}{N} \sum_{i=1}^N z(x_i, y_k) \frac{1}{f^*(x_i)} \quad \text{and} \quad S_{2k} = \frac{1}{N} \sum_{i=1}^N z(x_i, y_k) \frac{x_i}{f^*(x_i)},$$

and the x_i are sampled from the PDF $f^*(x)$ of one's choice. The solution for β , for arbitrary k , can be written as

$$\beta_k = \left\{ \mathcal{R}_k - \frac{1}{b-a} S_{1k} \right\} / \left\{ S_{2k} - \frac{b+a}{2} S_{1k} \right\}, \quad k = 1, 2, \dots, K. \quad (7.50)$$

Clearly, only one response is needed to obtain an estimate of β but if $K > 1$, then β can be estimated as an average of K estimates, i.e.,

$$\beta = \frac{1}{K} \sum_{k=1}^K \beta_k. \quad (7.51)$$

An estimate of $\sigma(\beta_k)$, due only to uncertainties in the responses, is easily obtained by differentiating Eq. (7.50) with respect to $\mathcal{R}_k$, which yields

$$\frac{\partial \beta_k}{\partial \mathcal{R}_k} = \frac{1}{S_{2k} - (b+a/2)S_{1k}}, \quad k = 1, 2, \dots, K,$$

and then using Eq. (7.14) to calculate

$$\sigma(\beta_k) \simeq \sum_{k=1}^K \left(\frac{\partial \beta_k}{\partial \mathcal{R}_k} \right)^2 \sigma^2(\mathcal{R}_k) = \sum_{k=1}^K \left(S_{2k} - \frac{b+a}{2} S_{1k} \right)^{-2} \sigma^2(\mathcal{R}_k), \quad (7.52)$$

$$k = 1, 2, \dots, K.$$

Example 7.7 A Linear PDF

Consider the specific case where $z(x, y_k) = e^{-\lambda x/y_k}$ for $K = 2$, where $a = 0, b = 5, \lambda = 0.1, y_1 = 0.5$, and $y_2 = 2.5$. Quadrature can be used to show that $\mathcal{R}_1 = 0.619166$ and $\mathcal{R}_2 = 0.902527$ (for $\alpha = 0.1750$ and $\beta = 0.0100$), but assume measurements yield $\mathcal{R}_1 = 0.6210 \pm 0.0031$ and $\mathcal{R}_2 = 0.9018 \pm 0.0045$. Also assume it is known that $\beta \geq 0$, i.e., the PDF is nondecreasing as x increases. Then, it is straightforward to show that $0 \leq \beta \leq 0.08$ in order for $f(x) = \alpha + \beta x$ to be a PDF. A single SMC simulation is run, sampling from the uniform PDF $f^*(x) = 1/(b-a) = 1/5$ to provide estimates of β and $\sigma(\beta)$. The results in Table 7.4 are obtained for three values of N . The corresponding value of α can be obtained from Eq. (7.47). Of course, the $\sigma(\beta)$ are underestimated because the contribution from the number of histories is ignored. Also, it is noted that $\sigma(\beta)$ is somewhat dependent on the value of β but not on the value of N .

Table 7.4 The parameter β obtained for different values of N . The correct value is $\beta = 0.0100$.

N	β	$\sigma(\beta)$	N	β	$\sigma(\beta)$	N	β	$\sigma(\beta)$
10^3	0.0109	0.0086	10^5	0.0103	0.0121	10^7	0.0103	0.0122

7.3.4 Unknown Parameter in Domain of $\mathbf{x}$

Sometimes the domain V of the random variable $\mathbf{x}$ may be unknown. For such problems the best one can do is to use the iterative method to try different domains until agreement with the $\mathcal{R}_k$ is obtained. To illustrate a systematic approach to this iterative procedure [Example 7.8](#) is given.

Example 7.8 A Buffon Needle Inverse Problem

Consider the problem of finding the grid spacing D , for Buffon's needle problem, when $P_{cut}(= \mathcal{R})$ and needle length L are known. This is a somewhat contrived inverse problem because if the needle length could be measured then surely the grid could also be measured. Nevertheless, the problem is a valid inverse problem and is illustrative of an unknown domain. In Chapter 1 it was shown that

$$P_{cut} = \frac{2}{\pi D} \int_0^U \cos^{-1}\left(\frac{x}{L}\right) dx, \quad (7.53)$$

where

$$U = \begin{cases} D, & D < L \\ L, & D \geq L \end{cases}.$$

It is not clear how to evaluate the integral because D is unknown. One way to proceed is to rewrite [Eq. \(7.53\)](#) in the form

$$P_{cut} = \frac{2L}{\pi D} \int_0^U [H(x) - H(D - x)] \cos^{-1}\left(\frac{x_i}{L}\right) \frac{1}{L} dx, \quad (7.54)$$

where H is the unit-step function. Then, in principle, one can sample uniformly over the interval $[0, L]$, because $1/L$ is a properly normalized PDF over this interval, to obtain the result

$$P_{cut} \simeq \frac{2L}{\pi DN} \sum_{i=1}^N h_i \cos^{-1}\left(\frac{x}{L}\right),$$

where

$$h_i = \begin{cases} 1, & \text{if } x_i \leq D \\ 0, & \text{if } x_i > D \end{cases}.$$

This may not seem promising, because D is unknown and so it is unknown when h_i is unity or zero.

One can, however, proceed by dividing the interval $[0, L]$ into J contiguous subintervals of equal size $\Delta = L/J$. Then Eq. (7.54) can be rewritten as

$$P_{cut} \simeq \frac{2L}{\pi D} \sum_{j=1}^J \delta_j S_j, \quad (7.55)$$

where

$$S_j = \frac{1}{N_j} \sum_{i=1}^{N_j} \cos^{-1} \left(\frac{x_{ij}}{L} \right), \quad (7.56)$$

N_j is the number of samples in the j th interval, x_{ij} is the i th sample in the j th interval, and

$$\delta_j = \begin{cases} 1, & \text{if } D \leq j\Delta \\ 0, & \text{if } D > j\Delta \end{cases}.$$

A possible IMC solution proceeds as follows. Conduct a single Monte Carlo simulation sampling x uniformly over $[0, L]$ and score the S_j , $j = 1, 2, \dots, J$. Then, start at $k = 1$, set $\bar{D}_k = \bar{D}_1 = \Delta/2$ (the midpoint of the first subinterval), and calculate the right side of Eq. (7.56), setting $\delta_j = 0$ for all $j > k = 1$. Incrementing k , one can perform similar operations for each $\bar{D}_k = \bar{D}_{k-1} + \Delta$, where δ_j is set to 0 for all $j > k$. Because the calculated P_{cut} is larger than the correct value if $\bar{D}_k < D$, the incrementing of k is stopped when the calculated P_{cut} is less than the given P_{cut} . The estimated value of D is interpolated as

$$\bar{D} = \bar{D}_k - \frac{P_{cut} - P_k}{\bar{P}_{k-1} - \bar{P}_k} \Delta,$$

where $\bar{P}_k$ is the right side of Eq. (7.55) when the value $\bar{D}_k$ is used, and $\bar{P}_{k-1}$ is the right side of Eq. (7.55) when the value $\bar{D}_{k-1}$ is used. Thus, the SMC procedure requires only a single simulation to form the S_j of Eq. (7.56) and an iterative algebraic solution of Eq. (7.55).

It is clear that the values of P_{cut} range from 0 (for infinite grid spacing) to 1 (for zero grid spacing), because it is a probability. For any case in which $D \geq L$, all the δ_j are unity and

$$P_{cut} = \frac{2}{\pi D} \int_0^L \cos^{-1} \left(\frac{x}{L} \right) dx.$$

For any case in which $D < L$, some of the δ_j are zero (for large enough J) and hence some of the computation is “wasted,” because some of the S_j are unused in the inversion process. It is also clear that the accuracy of the process depends on the accuracy with which L and P_{cut} are known, and the precision of estimated grid spacing can be improved by increasing N and J .

The SMC process outlined here was implemented using $N = 10^6$ histories and $J = 10^3$. The results shown in Table 7.5 were obtained.

Table 7.5 Recovery of grid spacing D by SMC.

L	Actual D	P_{cut}	Estimated D
4.00	1.00	0.920000	1.010
4.00	3.00	0.747487	2.998
4.00	4.00	0.636620	3.998
4.00	5.00	0.509296	4.996
4.50	10.00	0.254648	9.993

7.4 Inverse Monte Carlo by Simulation

The previous discussion considered IMC problems when the response can be written in the form of the integral of Eq. (7.4). However, there are many inverse problems (perhaps the majority) that cannot be put into the form of an integral. In such problems, multiple simulations are performed by varying the problem parameters until the difference between measured responses and the simulation results is minimized as discussed in Section 7.1.3. Such an IMC approach is a “brute force” approach and may be computationally expensive if each simulation requires a large computational effort. The use of SMC methods may be difficult in many simulation problems and requires that the tallies be analytically written in terms of the unknown parameters. It is also true that construction of an SMC model may be time-consuming and is “worth the effort” primarily in cases where it saves considerable computation time. The simulation approach to SMC has been employed in various radiation transport problems [e.g., Floyd et al. 1986; Michael 1991; and Dunn and Womersley 1994]. This approach is illustrated next by considering a simpler hypothetical case.

7.4.1 A Simple Two-Dimensional World

To illustrate the simulation approach to SMC, consider a two-dimensional world in which light scatters only at angles that are multiples of $\pi/2$. In this world, a rectangular medium is encountered that is five units wide and ten units tall. Light particles are projected at the center of the left wall and detectors are mounted on all four sides that count all of the particles that exit each of the sides. It is stipulated that the medium

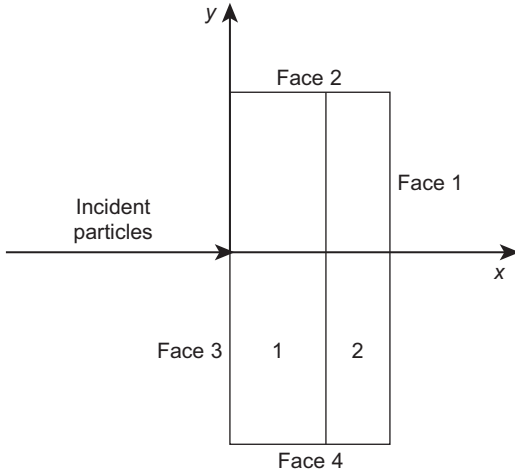


Figure 7.3 A two-dimensional scattering medium consisting of two materials, with particles entering normally from the left. The particles exiting each face are counted.

$\mathcal{R}_j, j = 1, \dots, 4$. The problem is to determine the thicknesses of each of the two materials that compose the medium given the four measured responses. This inverse problem is not posed in integral form. Nevertheless, an SMC solution can be obtained.

A possible approach is outlined below. Specify the particle direction of travel by a variable id , such that

$$id = \begin{cases} 1, & \text{if moving in the } x \text{ direction toward face 1} \\ 2, & \text{if moving in the } y \text{ direction toward face 2} \\ 3, & \text{if moving in the } -x \text{ direction toward face 3} \\ 4, & \text{if moving in the } -y \text{ direction toward face 4} \end{cases}$$

Then, construct a simulation model that proceeds as follows:

1. Introduce particles at $x = 0$ and $y = 0$ moving in the x -direction ($id = 1$).
2. First, presume the medium is all material 1. Thus, set $k = 1$ and $t_k = 5$, where t_k is the thickness of material k .
3. Sample a distance d to the next interaction. This requires sampling d from the PDF $f(d) = \mu_k e^{-\mu_k d}$, which leads to $d = -\ln(\rho)/\mu_k$, where ρ is sampled uniformly from $(0, 1)$.
4. Move the particle to a new position. This is accomplished by moving the particle d units in the direction indicated by id . This yields a new position (x, y) .
5. Check to see if the particle has exited the medium. For instance, if $x > 5$, the particle has exited the right face and the detector response for the right detector is incremented by

is composed of at most two materials, of which the scattering and absorption probabilities are known. The probability of an interaction per unit differential distance in material 1 is μ_1 and in material 2 is μ_2 . In the k th material, a photon is scattered left with probability p_l^k , is scattered backward with probability p_b^k , is scattered right with probability p_r^k , and is absorbed with probability p_a^k , for $k = 1$ and 2 .

An experiment is run in which one photon per second is introduced normally on the center of the front face, as shown in Fig. 7.3, and the detectors record the number of particles exiting each of the four sides per second

unity. A similar procedure is used for the other three faces. After scoring the appropriate response, start a new history. If the particle is still in the medium, proceed to the next step.

6. Sample a new direction by picking a pseudorandom number ρ . If $\rho \leq p_l^k$, then the particle scatters to the left and set $id = id + 1$. If $p_l^k < \rho \leq p_l^k + p_b^k$, then the particle scatters backward and set $id = id + 2$. If $p_l^k + p_b^k \leq \rho < p_l^k + p_b^k + p_r^k$, then the particle scatters to the right and set $id = id + 3$. Otherwise, the particle is absorbed and this history stops. It is necessary to renormalize id to values from 1 to 4 and so if $id > 4$, then $id = id \bmod 4$.
7. After all histories have been completed, save the four responses $\bar{R}_j = (\text{number leaving face } j) / (\text{number particles tracked}) = (\text{number escaping through face } j \text{ per second})$, $j = 1, 2, 3, 4$.
8. Next, assume the medium is all material 2, so that $k = 2$ and $t_k = 5$, where t_k is now the thickness of material 2. Again, start particles at $x = 0$ and $y = 0$ with $id = 1$. Repeat steps 3 to 6 and output the responses for this case.
9. Now assume the medium is part material 1 and part material 2. Start with some thickness for t_1 and let $t_2 = 5 - t_1$. Again, introduce particles at $x = 0$ and $y = 0$ in direction $id = 1$. Repeat steps 3 to 6, except this time, note crossings between the two materials and sample distances using the appropriate μ_k and sample new directions according to the probabilities for the medium in which the interactions occur.
10. Step through some set of discrete thicknesses of material 1, with corresponding thicknesses of material 2, and score the responses.
11. Finally, compare values of the objective function

$$\Omega(t_i) = \sum_{j=1}^4 |\mathcal{R}_j - \bar{R}_j^i|,$$

where $\bar{R}_j^i$ is the response of the j th detector for the i th simulation. The minimum value of Ω^i gives the best estimate of the thicknesses of the two materials.

This procedure may seem a little involved, but in fact is easily programmed and implemented.

Example 7.9 A Simplified Photon Scattering Inverse Problem

Assume a medium for which $t_1 = 3$ and $t_2 = 2$. Given that $\mu_1 = 0.5$ (in appropriate units) and $\mu_2 = 1.0$ (in the same units), and $p_l^1 = 0.25$, $p_b^1 = 0.25$, $p_r^1 = 0.25$, $p_a^1 = 0.25$, $p_l^2 = 0.25$, $p_b^2 = 0.5$, $p_r^2 = 0.25$, and $p_a^2 = 0$, one can run a forward Monte Carlo model to estimate the four responses. For 10^6 histories, the following responses were obtained: $\mathcal{R}_1 = 0.2520 \pm 0.0004 \text{ s}^{-1}$, $\mathcal{R}_2 = 0.0554 \pm 0.0002 \text{ s}^{-1}$, $\mathcal{R}_3 = 0.2240 \pm 0.0004 \text{ s}^{-1}$, and $\mathcal{R}_4 = 0.0551 \pm 0.0002 \text{ s}^{-1}$. The results of conducting the symbolic Monte Carlo procedure outlined here given these data are shown in Fig. 7.4. It is obvious that Ω is minimized near the correct value of $t_1 = 3$ and thus the SMC procedure can be used to solve inverse problems that are in the form of simulation models and not expressed explicitly as integral equations.

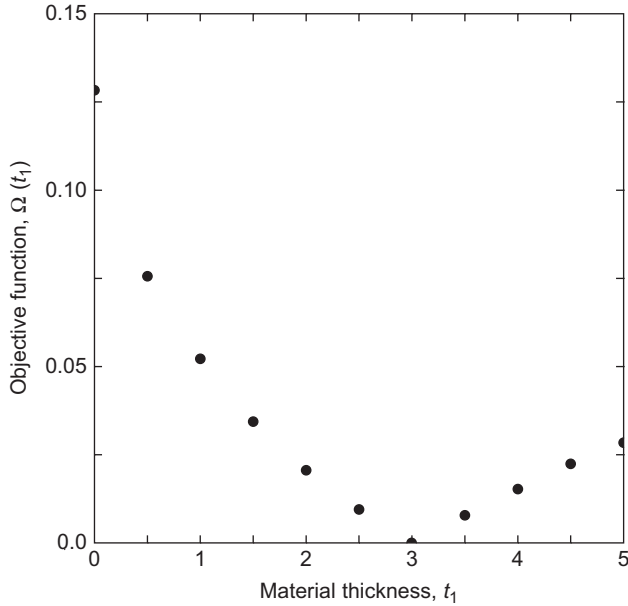


Figure 7.4 The figure of merit is minimized at the correct thickness of material 1, i.e. 3 units.

7.5 General Applications of IMC

Iterative IMC has been used in countless inverse problem applications, some published and many others not publicly shared. For instance, [Mosegaard and Tarantola \[1995\]](#) consider IMC methods applied to geophysics inverse problems. As another example, the paper by [Hammer et al. \[1995\]](#) describes a detailed iterative IMC model for determining properties of ocular fundus tissues from transmittance and reflectance measurements. The symbolic version of IMC has seen less application but has been shown to be useful. The examples given in this chapter have been rather generic and simple. In general, for applications of SMC, one must fully understand the process being simulated. The original paper by [Dunn \[1981\]](#) in which SMC was formally introduced (although it then was just called inverse Monte Carlo) considered a problem in remote sensing as an illustrative application. This was followed by applications to radiative transfer [[Dunn 1983](#); [Subramaniam and Menguc 1991](#)], to medical imaging [[Floyd et al. 1985](#); [Floyd et al. 1986](#)], to energy-dispersive X-ray fluorescence [[Yacout and Dunn 1987](#); [Michael 1991](#)], and to design of particle beam modifiers [[Dunn et al. 1987](#); [Dunn and Womersley 1994](#)].

7.6 Summary

Inverse problems are a generic class of problems for which outcomes are either measured or specified and the values of parameters that could lead to the outcomes are

sought. The existence and uniqueness of solutions to inverse problems are usually unknown. As a practical matter, however, inverse problems often are approached by presuming at least one solution exists and then attempting to find one or more solutions. If more than one solution is found, other considerations often must be used to select the most plausible solution. The term inverse Monte Carlo or IMC is sometimes used to describe the use of Monte Carlo methods to attempt to solve inverse problems.

The most direct form of IMC is the approach of specifying values of the unknown parameters, performing direct Monte Carlo simulations to estimate the outcomes, comparing the estimated outcomes to the given ones, and iterating until the estimated and given outcomes converge to an acceptable extent. This approach requires that a full direct Monte Carlo simulation be run for each set of candidate parameter values. Multiple simulations may not be a significant problem if each simulation is simple but can become problematic if the simulations are complex (requiring long run times) or highly multidimensional. Also, this “brute force” approach to IMC requires some sort of parameter update procedure be implemented. The “art” in using this approach is in how the parameter values are updated from iteration to iteration.

Another form of IMC, called symbolic Monte Carlo or SMC, is motivated by the desire to eliminate the need to perform multiple simulations. If the PDF, from which the Monte Carlo samples are being drawn, is dependent on some or all of the unknown parameters, then one can proceed by sampling from an assumed PDF that is independent of the unknown parameters. The histories must be weighted, in the importance sampling manner, by the ratio of actual to assumed PDFs. The actual PDF may contain the unknown parameters and, thus, the weight factors contain these parameters. Equating the Monte Carlo estimators to the measured or specified outcomes produces a system of equations that, in principle, can be solved for the unknown parameter values.

In IMC as in direct Monte Carlo, it is important to quantify the uncertainties in the estimates obtained. This is less straightforward in IMC than in direct Monte Carlo, but can be accomplished using an approximate procedure involving the uncertainties in the measured outcomes, the uncertainties in the direct Monte Carlo estimates, and the sensitivities of the responses to the unknown parameters.

Bibliography

- BEVINGTON, P.R. 1969, *Data Reduction and Error Analysis for the Physical Sciences*, McGraw-Hill, New York.
- DUNN, W.L. 1981, “Inverse Monte Carlo Analysis,” *J. Comput. Phys.*, **41**(1), 154–166.
- DUNN, W.L. 1983, “Inverse Monte Carlo Solutions for Radiative Transfer in Inhomogeneous Media,” *J. Quant. Spectrosc. Ra.*, **29**(1), 19–26.
- DUNN, W.L. 2006, “The Inverse Monte Carlo Formalism for Solving Inverse Radiation Problems” *T. Am. Nucl. Soc.*, **45**, 532–533.
- DUNN, W.L. AND J.K. SHULTIS 2009, “Monte Carlo Methods for Design and Analysis of Radiation Detectors,” *Radiat. Phys. Chem.*, **78**, 852–858.
- DUNN, W.L. AND W.J. WOMERSLEY 1994, *Solutions of Inverse and Optimization Problems in High Energy and Nuclear Physics Using Inverse Monte Carlo*, MC93 International

- Conference Monte Carlo Simulation in High Energy and Nuclear Physics, P. Dragovitsch, S.L. Linn, and M. Burbank, eds., pp. 78–87, World Scientific, Singapore.
- DUNN, W.L., V.C. BOFFI, AND F. O'FOGLUDHA 1987, "Applications of the Inverse Monte Carlo Method in Photon Beam Physics," *Nucl. Instrum. Meth. A*, **255**, 147–151.
- FLOYD, C.E., R.J. JASZCZAK, AND R.E. COLEMAN 1985, "Inverse Monte Carlo: A Unified Reconstruction Algorithm for SPECT," *IEEE T. Nucl. Sci.*, **NS-32**, 779–785.
- FLOYD, C.E., R.J. JASZCZAK, K.L. GREER, AND R.E. COLEMAN 1986, "Inverse Monte Carlo as a Unified Reconstruction Algorithm for ECT," *J. Nucl. Med.*, **27**(10), 1577–1585.
- HAMMER, M., A. ROGGAN, AND G. MULLER 1995, "Optical Properties of Ocular Fundus Tissues—an *in Vitro* Study using the Double-Integrating-Sphere Technique and Inverse Monte Carlo Simulation," *Phys. Med. Biol.*, **40**, 963–978.
- MICHAEL, M. 1991, "A Complete Inverse Monte Carlo Model for Energy-Dispersive X-ray Fluorescence Analysis," *Nucl. Instrum. Meth. A*, **301**(3), 523–542.
- MOSEGAARD, K. AND M. SAMBRIDGE 2002, "Monte Carlo Analysis of Inverse Problems," *Inverse Probl.*, **18**, R29–R54.
- MOSEGAARD, K. AND A. TARANTOLA 1995, "Monte Carlo Sampling of Solutions to Inverse Problems," *J. Geophys. Res.*, **100**(B7), 12,431–12,447.
- PRESS, W.H., S.A. TEUKOLSKY, W.T. VETTERLING, AND B.P. FLANNERY 1996, *Numerical Recipes in Fortran 77*, 2nd ed., Cambridge University Press, New York.
- SUBRAMANIAM, S. AND M.P. MENGUC 1991, "Solution of the Inverse Radiation Problem for Inhomogeneous and Anisotropically Scattering Media Using a Monte Carlo Technique," *Int. J. Heat Mass Tran.*, **34**(1), 253–266.
- YACOUT, A.M. AND W.L. DUNN 1987, "Application of the Inverse Monte Carlo Method to Energy-dispersive X-ray Fluorescence," *Adv. X-Ray Anal.*, **30**, 113–120.

Problems

1. It can be shown that, for a particular value of θ , $0 < \theta \leq 2$, the definite integral

$$R = \int_0^{\pi/2} \sin^{\theta}(x) dx$$

assumes the value $\mathcal{R} = 0.814183$. Use iterative IMC to estimate the value of θ to two decimal places.

2. The definite integral

$$R(y) = \int_0^5 e^{-\theta xy} dx$$

assumes the value 1.567 for $y = 2$. Use iterative IMC to estimate the value of θ , $0 < \theta \leq 2$ to two decimal places.

3. Consider the model

$$R(y) = \int_a^b (\theta_1 + \theta_2 xy) \frac{1}{b-a} dx.$$

Assume that, for $a = 0$ and $b = 1$, the model gives $R(1) = 2$ and $R(3) = 5$. Use SMC to find θ_1 and θ_2 .

4. Consider the model

$$R(y) = \int_a^b (\theta_1 + \theta_2 x^2 y) dx.$$

Assume that, for $a = 0$ and $b = 5$, the model gives $R(1) = 156.6667$ and $R(4) = 656.6667$. Use SMC to find θ_1 and θ_2 .

5. Consider the definite integral

$$R = \int_0^\infty e^{-bx} dx,$$

which can be rewritten as

$$R = \frac{1}{b} \int_0^\infty b e^{-bx} dx.$$

Employ the linearization of the PDF method and sample from $f^*(x) = e^{-x}$, $0 \leq x < \infty$ to find the value of the parameter b if $I = 0.8333$.

6. A process involving a stochastic variable, which is exponentially distributed on the interval $[0, \infty)$, is known to have a linear kernel and expected response of the form

$$R(y) = \int_0^\infty (\theta_1 y + \theta_2 x) 4e^{-4x} dx.$$

Assume the values $\mathcal{R}_1 = R(y_1 = 1) = 2.0$ and $\mathcal{R}_2 = R(y_2 = 3) = 10.0$ are given and the parameters θ_1 and θ_2 are unknown. Use SMC to estimate the values of θ_1 and θ_2 .

7. Use a five-term Taylor series expansion of the exponential function to find the value of b if

$$R(b) = \int_0^\infty x^2 e^{-bx} dx = 1.9223.$$

8. Generalize the example in [Section 7.4](#) to the case of isotropic scattering in the two-dimensional world, i.e., probability of scattering through an angle θ is uniform in $[0, 2\pi]$. The total probability of scattering in the two media are $p_s^1 = 0.8$ and $p_s^2 = 1.0$ and the absorption probabilities are $p_a^1 = 0.2$ and $p_a^2 = 0.0$.

8 Linear Operator Equations

*The equations gave Susan great fears,
and solving them drove her to tears.
But try as she might
the answers weren't right
until she tried dancing on spheres.*

One interesting application of Monte Carlo methods is to solve numerically equations involving linear operators. Such equations include linear algebraic equations, Fredholm integral equations, and linear differential equations that describe boundary value problems. It should be noted, however, that most of these equations usually can be solved much more efficiently and accurately with standard numerical analysis techniques than they can with Monte Carlo calculations. Nevertheless, there are situations where only approximate values of the solution are needed or where standard methods perform badly, such as when the problem is very large or very intricate. For example, multidimensional integral equations or boundary value problems with complicated boundaries are often better treated using Monte Carlo methods.

In this chapter, some basic Monte Carlo methods are introduced for treating linear operator equations. For more complex methods and for details of the methods presented here, the reader is referred to [Hammersley and Handscomb \[1964\]](#), [Rubinstein \[1981\]](#), [Sabelfeld \[1991\]](#), and [Mikhailov \[1995\]](#).

8.1 Linear Algebraic Equations

Most sets of linear algebraic equations can be solved using Monte Carlo, but such equations are almost always more effectively solved using standard techniques of numerical analysis. Nevertheless, Monte Carlo methods for linear algebraic equations are introduced because analogous methods can be applied to integral equations, for which standard numerical methods often are not as effective, particularly for highly multidimensional integral equations.

Consider the set of n linear algebraic equations $\mathbf{Ax} = \mathbf{f}$, which can be rewritten as

$$\mathbf{x} = \mathbf{f} + \mathbf{Bx}, \quad (8.1)$$

where the $n \times n$ matrix $\mathbf{B} \equiv \mathbf{I} - \mathbf{A}$ and has elements b_{ij} . Equations of this type can often be solved by the following iteration scheme:

$$\mathbf{x}^{(k+1)} = \mathbf{f} + \mathbf{Bx}^{(k)}, \quad k = 0, 1, 2, \dots, \quad \text{with } \mathbf{x}^{(0)} = \mathbf{f}, \quad (8.2)$$

$$= (\mathbf{I} + \mathbf{B} + \mathbf{B}^2 + \dots + \mathbf{B}^k)\mathbf{f}. \quad (8.3)$$

If the sequence $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}^{(3)}, \dots$ converges, then the converged vector $\mathbf{x}$ satisfies Eq. (8.2), which is equivalent to Eq. (8.1), so that the converged $\mathbf{x}$ is the desired solution. It can be shown [Wachspress 1966] that a necessary condition for convergence is that the spectral radius $\mu(\mathbf{B})$ of the matrix $\mathbf{B}$ be less than unity, i.e.,

$$\mu(\mathbf{B}) \equiv \max_i |\lambda_i(\mathbf{B})| < 1, \quad (8.4)$$

where $\lambda_i(\mathbf{B})$ are the eigenvalues of $\mathbf{B}$. The smaller the spectral radius, the more rapid is the convergence. From Eq. (8.3) the i th component of $\mathbf{x}^{(k+1)}$ is

$$\begin{aligned} x_i^{(k+1)} &= f_i + \sum_{j_1} b_{ij_1} f_{j_1} + \sum_{j_1 j_2} b_{ij_1} b_{j_1 j_2} f_{j_2} + \dots + \sum_{j_1 j_2 \dots j_k} b_{ij_1} b_{j_1 j_2} \dots b_{j_{k-1} j_k} f_{j_k} \\ &= \sum_{m=0}^k \sum_{j_1=1}^n \sum_{j_2=1}^n \dots \sum_{j_m=1}^n b_{ij_1} b_{j_1 j_2} \dots b_{j_{m-1} j_m} f_{j_m}, \quad \text{with } f_{j_0} = f_i. \end{aligned} \quad (8.5)$$

Often, it is not the solution $\mathbf{x}$ that is of interest but, rather, some inner product,

$$(\mathbf{h}, \mathbf{x}) \equiv \sum_{i=1}^n h_i x_i, \quad (8.6)$$

where $\mathbf{h}$ is a specified weighting vector. In particular, if $h_i = \delta_{ij}$, then $(\mathbf{h}, \mathbf{x}) = x_j$, the solution for the j th component of $\mathbf{x}$.

8.1.1 Solution of Linear Equations by Random Walks

To solve Eq. (8.1) by Monte Carlo techniques, an $n \times n$ transition matrix $\mathbf{P}$ and an n -component starting vector $\mathbf{p}$ are constructed of which the non-negative elements have the properties [Rubinstein 1981]

$$\sum_{j=1}^n P_{ij} = 1, \quad i = 1, \dots, n \quad \text{and} \quad \sum_{i=1}^n p_i = 1. \quad (8.7)$$

Moreover, $p_i > 0$ if $f_i \neq 0$ or $p_i = 0$ if $f_i = 0$ and $P_{ij} > 0$ if $b_{ij} \neq 0$ or $P_{ij} = 0$ if $b_{ij} = 0$. The $\mathbf{P}$ and $\mathbf{p}$ can be chosen more or less arbitrarily (subject to the above constraints), although better results are obtained if the elements are approximately proportional to the magnitudes of the elements in $\mathbf{B}$ and $\mathbf{f}$, respectively.

With these auxiliary quantities a series of *random walks* or Markov chains is made through a set of states consisting of the integers 1 through n . Here, p_i determines the probability of starting a random walk at state i and P_{ij} determines the probability, at each step in the random walk, of moving from state i to state j . An integer k is initially chosen, corresponding to the value of k in Eq. (8.2). Thus, each random walk consists of a sequence of $k + 1$ states $i_0 \rightarrow i_1 \rightarrow i_2 \rightarrow \dots \rightarrow i_k$.

At each step in a random walk, the quantity

$$U_m = \frac{b_{i_0 i_1} b_{i_1 i_2} \cdots b_{i_{m-1} i_m}}{P_{i_0 i_1} P_{i_1 i_2} \cdots P_{i_{m-1} i_m}} \quad (8.8)$$

is accumulated, which can be constructed recursively during the random walk as

$$U_m = U_{m-1} \frac{b_{i_{m-1} i_m}}{P_{i_{m-1} i_m}}, \quad \text{with } U_0 = 1. \quad (8.9)$$

Also, at each step in the random walk, the sum $\sum_m U_m f_{i_m}$ is constructed, so that, at the end of the random walk, a value of the random variable

$$Z_k(\mathbf{h}) = \frac{h_{i_0}}{P_{i_0}} \sum_{m=0}^k U_m f_{i_m} \quad (8.10)$$

is obtained. A large number N of random walks is performed, each producing a value of $Z_k^{(s)}(\mathbf{h})$, $s = 1, \dots, N$. The desired inner product is then estimated as

$$(\mathbf{h}, \mathbf{x}^{(k+1)}) \simeq \frac{1}{N} \sum_{s=1}^N Z_k^{(s)}(\mathbf{h}). \quad (8.11)$$

Proof of Eq. (8.11)

The path $i_0 \rightarrow i_1 \rightarrow i_2 \rightarrow \cdots \rightarrow i_k$ has a probability of being realized of $p_{i_0} P_{i_0 i_1} P_{i_1 i_2} \cdots P_{i_{k-1} i_k}$. Thus, the expected value of the random variable $Z_k(\mathbf{h})$ is

$$\langle Z_k(\mathbf{h}) \rangle = \sum_{i_0=1}^n \sum_{i_1=1}^n \cdots \sum_{i_k=1}^n Z_k(\mathbf{h}) p_{i_0} P_{i_0 i_1} P_{i_1 i_2} \cdots P_{i_{k-1} i_k}. \quad (8.12)$$

Substitution of Eqs. (8.10) and (8.8) into this expression yields

$$\begin{aligned} \langle Z_k(\mathbf{h}) \rangle &= \sum_{i_0=1}^n \cdots \sum_{i_k=1}^n h_{i_0} \sum_{m=0}^k \frac{b_{i_0 i_1} b_{i_1 i_2} \cdots b_{i_{m-1} i_m}}{P_{i_0 i_1} P_{i_1 i_2} \cdots P_{i_{m-1} i_m}} f_{i_m} P_{i_0 i_1} P_{i_1 i_2} \cdots P_{i_{k-1} i_k}, \\ &= \sum_{i_0=1}^n \cdots \sum_{i_k=1}^n h_{i_0} \sum_{m=0}^k b_{i_0 i_1} b_{i_1 i_2} \cdots b_{i_{m-1} i_m} f_{i_m} P_{i_m i_{m+1}} \cdots P_{i_{k-1} i_k}. \end{aligned} \quad (8.13)$$

Now, use the fact that $\sum_{j=1}^n P_{ij} = 1$ to reduce Eq. (8.13) as follows:

$$\begin{aligned}
 \langle Z_k(\mathbf{h}) \rangle &= \sum_{i_0=1}^n h_{i_0} f_{i_0} + \sum_{i_0=1}^n \sum_{i_1=1}^n h_{i_0} b_{i_0 i_1} f_{i_1} + \cdots + \sum_{i_0=1}^n \cdots \sum_{i_k=1}^n h_{i_0} b_{i_0 i_1} \cdots b_{i_{k-1} i_k} f_{i_k} \\
 &= \sum_{m=0}^k \sum_{i_0=1}^n \cdots \sum_{i_m=1}^n h_{i_0} b_{i_0 i_1} b_{i_1 i_2} \cdots b_{i_{m-1} i_m} f_{i_m} \\
 &= \sum_{m=0}^k \sum_{i_0=1}^n h_{i_0} [\mathbf{B}^m \mathbf{f}]_{i_0} = \left(\mathbf{h}, \sum_{m=0}^k \mathbf{B}^m \mathbf{f} \right) = (\mathbf{h}, \mathbf{x}^{(k+1)}), \tag{8.14}
 \end{aligned}$$

where the last line follows from Eq. (8.5). Thus, Eq. (8.11) is proven.

Limiting cases

The limit of Eq. (8.11) then gives

$$\lim_{k \rightarrow \infty} \langle Z_k(\mathbf{h}) \rangle = \lim_{k \rightarrow \infty} (\mathbf{h}, \mathbf{x}^{(k+1)}) = (\mathbf{h}, \mathbf{x}). \tag{8.15}$$

Hence, *provided* the sequence $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}^{(3)}, \dots$ converges and the random-walk path $i_0 \rightarrow i_1 \rightarrow \cdots \rightarrow i_k \rightarrow \cdots$ is infinitely long, an unbiased estimator of $(\mathbf{h}, \mathbf{x})$ is obtained. In practice, good estimates of $(\mathbf{h}, \mathbf{x})$ can be obtained from Eq. (8.11) for small values of k . The faster the iterative solution converges, i.e., the smaller the spectral radius $\mu(\mathbf{B})$, the smaller can be k .

Three features of the above Monte Carlo procedure for evaluation $(\mathbf{h}, \mathbf{x}^{(k+1)})$ should be noted. The same random paths can be used to evaluate this inner product for different values of $\mathbf{f}$. Also, if the $\mathbf{B}$ and $\mathbf{P}$ matrices are such that $\mathbf{B} = \alpha \mathbf{P}$, where $0 < \alpha < 1$, then Eq. (8.9) reduces to $U_m = \alpha^m$ so that

$$Z_k(\mathbf{h}) = \frac{h_{i_0}}{P_{i_0}} \sum_{m=0}^k \alpha^m f_{i_m}. \tag{8.16}$$

Finally, if $h_i = \delta_{ij}$ (the Kronecker delta function), the random walks always start at $i_0 = j$ with $h_{i_0}/p_{i_0} = 1$ so that

$$\lim_{k \rightarrow \infty} \langle Z_k(\mathbf{h}) \rangle = \lim_{k \rightarrow \infty} x_j^{(k+1)} = x_j, \tag{8.17}$$

the j th component of $\mathbf{x}$. An example of the above method is shown in Example 8.1.

Example 8.1 Solution of Linear Algebraic Equations $\mathbf{x} = \mathbf{f} + \mathbf{B}\mathbf{x}$

As an example of the Monte Carlo method described in [Section 8.1.1](#), consider the 4×4 set of linear algebraic equations

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \frac{1}{10} \begin{pmatrix} -7 \\ 18 \\ 33 \\ 28 \end{pmatrix} + \frac{1}{10} \begin{pmatrix} 4 & 1 & 1 & 2 \\ 1 & -5 & 1 & 2 \\ 1 & 2 & -4 & 1 \\ 1 & 2 & 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}.$$

The exact solution of this equations is $\mathbf{x} = (1 \ 2 \ 3 \ 4)$ and for $\mathbf{h} = (1 \ 1 \ 1 \ 1)$ the exact value of $(\mathbf{h}, \mathbf{x}) = 10$. For this Monte Carlo method, the starting probability vector was taken as $\mathbf{p} = (0.25 \ 0.25 \ 0.25 \ 0.25)$ and the transition matrix $\mathbf{P}$ was chosen as

$$\mathbf{P} = \begin{pmatrix} 0.55 & 0.15 & 0.15 & 0.15 \\ 0.15 & 0.55 & 0.15 & 0.15 \\ 0.20 & 0.20 & 0.40 & 0.20 \\ 0.25 & 0.25 & 0.25 & 0.25 \end{pmatrix}.$$

The results of the Monte Carlo solution are shown in [Fig. 8.1](#). The left figure shows how the error in $(\mathbf{h}, \mathbf{x}^{(k)})$ decreases as the number of histories N increases for the case of 25 steps in the random walk. As seen from the right figure, the maximum number of steps k used to terminate each random walk for this example should be about 10 or greater to avoid errors from random walks that are too short. However, using very long random walks affords little, if any, increase in accuracy while wasting computing time.

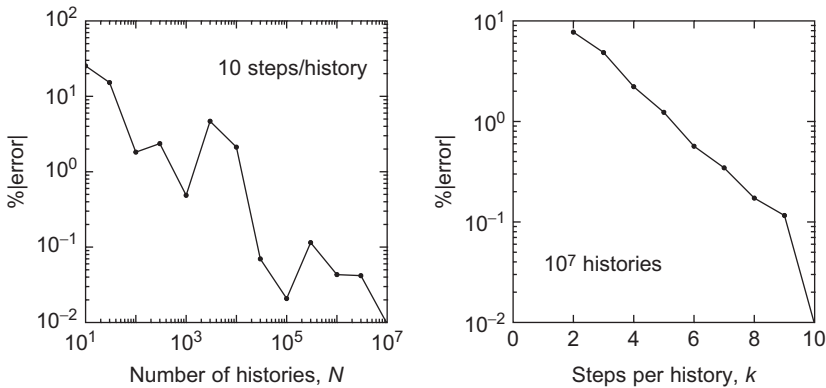


Figure 8.1 The percent error in $(\mathbf{h}, \mathbf{x})$ as a function of the number of histories N (left) and for the number of steps k in each history (right).

8.1.2 Solution of the Adjoint Linear Equations by Random Walks

Associated with the set of linear equations of [Eq. \(8.1\)](#) is the set of equations

$$\mathbf{x}^\dagger = \mathbf{h} + \mathbf{B}^T \mathbf{x}^\dagger, \quad (8.18)$$

where $\mathbf{B}^T$ is the transpose or *adjoint* of $\mathbf{B}$. Again, this adjoint set of equations can be solved by an iterative solution

$$\mathbf{x}^{\dagger(k+1)} = \mathbf{h} + \mathbf{B}^T \mathbf{x}^{\dagger(k)}, \quad k = 1, 2, \dots \quad \text{with} \quad \mathbf{x}^{\dagger(0)} = \mathbf{0}. \quad (8.19)$$

Because $\mathbf{B}$ and $\mathbf{B}^T$ have the same eigenvalues, $\mu(\mathbf{B})$ equals $\mu(\mathbf{B}^T)$, and this iterative solution converges whenever that of Eq. (8.2) converges.

Multiplication of Eq. (8.1) by the transpose of $\mathbf{x}^\dagger$ yields¹

$$(\mathbf{x}^\dagger, \mathbf{x}) = (\mathbf{x}^\dagger, \mathbf{f}) + (\mathbf{x}^\dagger, \mathbf{B}\mathbf{x}). \quad (8.20)$$

Similarly, from Eq. (8.18), it is found that

$$(\mathbf{x}, \mathbf{x}^\dagger) = (\mathbf{x}, \mathbf{h}) + (\mathbf{x}, \mathbf{B}^T \mathbf{x}^\dagger). \quad (8.21)$$

It is readily shown that $(\mathbf{x}, \mathbf{x}^\dagger) = (\mathbf{x}^\dagger, \mathbf{x})$ and that $(\mathbf{x}^\dagger, \mathbf{B}\mathbf{x}) = (\mathbf{x}, \mathbf{B}^T \mathbf{x}^\dagger)$. Subtraction of Eq. (8.20) from Eq. (8.21) then yields

$$(\mathbf{h}, \mathbf{x}) = (\mathbf{x}^\dagger, \mathbf{f}). \quad (8.22)$$

Thus, by solving Eq. (8.18) for $\mathbf{x}^\dagger$ by a random-walk algorithm, another independent estimator of the inner product $(\mathbf{h}, \mathbf{x})$ can be obtained.

To find $(\mathbf{x}^\dagger, \mathbf{f})$ using Markov chains, a state-transition matrix $\mathbf{P}^\dagger$ and initial state probability vector $\mathbf{p}^\dagger$ are defined such that [Rubinstein 1981]

$$\sum_{j=1}^n P_{ij}^\dagger = 1, \quad i = 1, \dots, n \quad \text{and} \quad \sum_{i=1}^n p_i^\dagger = 1, \quad (8.23)$$

where $P_{ij}^\dagger > 0$ if $b_{ij}^T = b_{ji} \neq 0$ or $P_{ij}^\dagger = 0$ if $b_{ij}^T = b_{ji} = 0$ and $p_i^\dagger > 0$ if $h_i \neq 0$ or $p_i^\dagger = 0$ if $h_i = 0$. Then, to estimate $(\mathbf{x}^\dagger, \mathbf{f})$, first select an integer k (again corresponding to the k in Eq. (8.3)) and simulate a large number N of random walks, each of k steps. The initial state i_0 is selected with probability $p_{i_0}^\dagger$. For each random walk, calculate the random variable

$$Z_k^\dagger(\mathbf{f}) = \frac{f_{i_0}}{p_{i_0}^\dagger} \sum_{m=1}^k U_m^\dagger h_{i_m}, \quad (8.24)$$

¹ The *inner product* of two n -component real vectors $\mathbf{x}$ and $\mathbf{y}$ is defined as

$$(\mathbf{x}, \mathbf{y}) \equiv \mathbf{x}^T \cdot \mathbf{y} = \sum_{i=1}^n x_i y_i.$$

where $U_m^\dagger$ is accumulated at each step of the random walk according to

$$U_m^\dagger = U_{m-1}^\dagger \frac{b_{i_{m-1}i_m}^T}{P_{i_{m-1}i_m}^\dagger} = U_{m-1}^\dagger \frac{b_{i_m i_{m-1}}}{P_{i_{m-1}i_m}^\dagger}, \quad \text{with } U_0^\dagger = 1. \quad (8.25)$$

The inner product $(\mathbf{x}^{\dagger(k+1)}, \mathbf{f})$ then is estimated as

$$(\mathbf{x}^{\dagger(k+1)}, \mathbf{f}) \simeq \frac{1}{N} \sum_{s=1}^N Z_k^{\dagger(s)}(\mathbf{f}), \quad (8.26)$$

where $Z_k^{\dagger(s)}(\mathbf{f})$ is the value of Eq. (8.24) for random walk s . Finally, as $k \rightarrow \infty$,

$$\lim_{k \rightarrow \infty} \langle Z_k^\dagger(\mathbf{f}) \rangle = \lim_{k \rightarrow \infty} (\mathbf{f}, \mathbf{x}^{\dagger(k+1)}) = (\mathbf{f}, \mathbf{x}^\dagger) = (\mathbf{h}, \mathbf{x}). \quad (8.27)$$

Notice that the same random walk can be used to evaluate different $\mathbf{h}$ vectors by simply changing the h_{i_m} in Eq. (8.24). In particular, if $h_i = \delta_{ij}$, $i = 1, 2, \dots, n$, then

$$x_j^{(k+1)} = \langle Z_k^\dagger(\mathbf{f}) \rangle = \langle (f_{i_0}/P_{i_0}^\dagger) \sum_{m=1}^k U_m^\dagger \delta_{i_m j} \rangle, \quad j = 1, 2, \dots, n. \quad (8.28)$$

Thus, the adjoint method is more suitable for finding the general shape of $\mathbf{x}$ because the same random walk can be used to build estimates of all the x_j , whereas the direct method of Section 8.1.1 is best used for a single $\mathbf{f}$ vector.

8.1.3 Solution of Linear Equations by Finite Random Walks

The difficulty with the previous two methods for solving linear algebraic equations is that the Markov chain is stopped after k steps in the random walk. The value of k is chosen rather arbitrarily and one hopes that the result for $(\mathbf{h}, \mathbf{x}^{(k+1)}) \simeq (\mathbf{h}, \mathbf{x})$. To verify this, estimates based on different values of k need to be performed to ensure convergence has been achieved.

There is another approach that avoids this difficulty and produces estimates of $(\mathbf{h}, \mathbf{x})$ directly. In fact, this method, which introduces *absorption states* into the random walk, predates the previous methods. This alternative method was first proposed by Neumann and Ulam in the 1940s and was later published and expanded by Forsythe and Liebler [1950].

As before, a vector $\mathbf{p}$ with $\sum_{i=1}^n p_i = 1$ is used to define the starting state of each random walk. But in this method, the state-transition matrix $\mathbf{P}$, which has elements

$P_{ij} \geq 0$, such that $P_{ij} > 0$ if $b_{ij} \neq 0$, has rows that do not always sum to unity. The *kill vector*

$$g_i^{kill} \equiv 1 - \sum_{j=1}^n P_{ij} \geq 0, \quad i = 1, \dots, n \quad (8.29)$$

is used to terminate each random walk. A random walk is begun at state i_0 with probability p_{i_0} . At each step in the walk, the particle goes from state i_{j-1} to state i_j with probability P_{ij} or the walk is terminated with probability $g_{i_{j-1}}^{kill}$. Thus, each random walk is finite in length $i_0 \rightarrow i_1 \rightarrow \dots \rightarrow i_k$, where the number of steps k may be different for each random walk. For each random walk calculate the random variable

$$Z_k(\mathbf{h}) = \frac{h_{i_0}}{p_{i_0}} U_k \frac{f_{i_k}}{g_{i_k}^{kill}}, \quad (8.30)$$

where U_k is given by Eq. (8.8). The inner product $(\mathbf{h}, \mathbf{x})$ is then estimated by averaging Z_k over N random walks, i.e.,

$$(\mathbf{h}, \mathbf{x}) \simeq \frac{1}{N} \sum_{s=1}^N Z^{(s)}(\mathbf{h}). \quad (8.31)$$

Proof of Eq. (8.31)

The probability of obtaining the path $i_0 \rightarrow i_1 \rightarrow \dots \rightarrow i_k$ is $p_{i_0} P_{i_0 i_1} P_{i_1 i_2} \dots P_{i_{k-1} i_k} g_{i_k}^{kill}$. Hence, the expected value of $Z_k(\mathbf{h})$ is

$$\langle Z_k(\mathbf{h}) \rangle = \sum_{k=0}^{\infty} \sum_{i_0=1}^n \dots \sum_{i_k=1}^n Z_k(\mathbf{h}) p_{i_0} P_{i_0 i_1} P_{i_1 i_2} \dots P_{i_{k-1} i_k} g_{i_k}^{kill}. \quad (8.32)$$

Substitution of Eqs. (8.30) and (8.8) into this result then gives

$$\begin{aligned} \langle Z_k(\mathbf{h}) \rangle &= \sum_{k=0}^{\infty} \sum_{i_0=1}^n \dots \sum_{i_k=1}^n h_{i_0} b_{i_0 i_1} b_{i_1 i_2} \dots b_{i_{k-1} i_k} f_{i_k}, \\ &= \sum_{i_0=1}^n h_{i_0} [f_{i_0} + (\mathbf{B}\mathbf{f})_{i_0} + (\mathbf{B}^2\mathbf{f})_{i_0} + (\mathbf{B}^3\mathbf{f})_{i_0} + \dots] \\ &= \sum_{i_0=1}^n h_{i_0} x_{i_0} = (\mathbf{h}, \mathbf{x}), \end{aligned} \quad (8.33)$$

where the last line follows from Eq. (8.3) and, thus, Eq. (8.31) is justified.

Use considerations

A final note about Monte Carlo solutions of linear algebraic equations. Classic numerical methods [Press et al. 1996] are almost always more computationally efficient. Only if the matrix size is very large ($\gtrsim 10^4$) and only if a rough approximation is sufficient should Monte Carlo methods be considered. Also, if $(\mathbf{h}, \mathbf{x})$ is to be estimated for many different $\mathbf{h}$ and $\mathbf{f}$, the same random walks can be used to obtain estimates of the different inner products.

8.2 Linear Integral Equations

In contrast to the limited utility of using Monte Carlo methods for linear algebraic equations, Monte Carlo is often the best method for numerically solving linear integral equations, particularly if the integrals are multidimensional. Moreover, many important differential equations can be formally converted into integral equations and so, in principle, Monte Carlo can again be used. In fact, one of the first applications of Monte Carlo analysis was to solve the integral form of the Boltzmann equation for neutral particle transport (see Chapter 10).

Consider the Fredholm integral equation

$$y(\mathbf{x}) = f(\mathbf{x}) + \int_V K(\mathbf{x}', \mathbf{x}) y(\mathbf{x}') d\mathbf{x}', \quad \mathbf{x} \in V. \quad (8.34)$$

For the one-dimensional case,

$$y(x) = f(x) + \int_a^b K(x', x) y(x') dx'. \quad (8.35)$$

This equation can be approximated by a set of linear algebraic equations in several ways. Application of some M -point deterministic quadrature formula immediately gives

$$y(x_i) = f(x_i) + \sum_{j=1}^M w_j K(x_j, x_i) y(x_j), \quad i = 1, \dots, M, \quad (8.36)$$

where the x_j and w_j are the quadrature abscissas and weights. Equation (8.36) is a set of M linear algebraic equations that can be written in the standard form $\mathbf{y} = \mathbf{f} + \mathbf{A}\mathbf{y}$, where $y_i \equiv y(\mathbf{x}_i)$.

Alternatively, a complete set of functions $\phi_1(x), \phi_2(x), \dots$ can be used to expand $f(x)$ as

$$f(x) = f_1 \phi_1(x) + f_2 \phi_2(x) + \dots \quad (8.37)$$

If the expansions

$$\int_a^b K(x', x) \phi_j(x') dx' = a_{1j} \phi_1(x) + a_{2j} \phi_2(x) + \cdots \quad (8.38)$$

are known, then if $y(x)$ is approximated by an M -term expansion as

$$y(x) \simeq y_1 \phi_1(x) + y_2 \phi_2(x) + \cdots + y_M \phi_M(x), \quad (8.39)$$

then the unknown expansion coefficients y_j are the solution of the M linear equations $\mathbf{y} = \mathbf{f} + \mathbf{A}\mathbf{y}$.

For multidimensional problems, the region V can be partitioned into M contiguous (mutually disjoint) subdomains V_i , $i = 1, \dots, M$ so that $V = \cup_{i=1}^M V_i$. The subdomains are chosen so that $f(\mathbf{x})$ and $K(\mathbf{x}', \mathbf{x})$ are approximately constant in each subdomain, i.e., $f(\mathbf{x}) \simeq f_i$ for $\mathbf{x} \in V_i$ and $K(\mathbf{x}', \mathbf{x}) \simeq a_{ij}$ for $\mathbf{x} \in V_i$ and $\mathbf{x}' \in V_j$. Equation (8.34) can, thus, be approximated, for $\mathbf{x} \in V_i$, as

$$\begin{aligned} y_i &= f_i + \sum_{j=1}^M \int_{V_j} K(\mathbf{x}', \mathbf{x}) y(\mathbf{x}') d\mathbf{x}' \\ &\simeq f_i + \sum_{j=1}^M a_{ij} \int_{V_j} y(\mathbf{x}') d\mathbf{x}' \simeq f_i + \sum_{j=1}^M (V_j a_{ij}) y_j, \end{aligned} \quad (8.40)$$

or, with $b_{ij} = V_j a_{ij}$,

$$y_i \simeq f_i + \sum_{j=1}^M b_{ij} y_j, \quad i = 1, 2, \dots, M, \quad (8.41)$$

a set of M linear algebraic equations.

However, as the dimensionality of the integral in Eq. (8.34) increases, converting the integral equation to linear algebraic equations becomes increasingly complex, resulting in very large $\mathbf{A}$ matrices. For dimensions greater than 3 or 4, Monte Carlo techniques become increasingly attractive.

8.2.1 Monte Carlo Solution of a Simple Integral Equation

In this section, Monte Carlo techniques are combined with a Neumann expansion to solve an integral equation of which the kernel and inhomogeneous term have special properties in order to show the general idea behind the method. Later, a more robust procedure is introduced to handle equations of which the kernels do not have these special properties.

Suppose the value of some inner product

$$H = (h(\mathbf{x}), y(\mathbf{x})) \equiv \int_V h(\mathbf{x}) f(\mathbf{x}) d\mathbf{x}$$

is sought where $h(\mathbf{x})$ is specified and $y(\mathbf{x})$ is the solution of the integral equation

$$y(\mathbf{x}) = \int_V K(\mathbf{x}', \mathbf{x}) y(\mathbf{x}') d\mathbf{x}' + f(\mathbf{x}), \quad \mathbf{x} \in V, \quad (8.42)$$

where $f(\mathbf{x}) \geq 0$ and $K(\mathbf{x}', \mathbf{x}) \geq 0$ for $\mathbf{x}', \mathbf{x} \in V$. For simplicity, it is assumed that the source term and kernel are normalized, such that

$$\int_V f(\mathbf{x}) d\mathbf{x} = 1 \quad \text{and} \quad \int_V K(\mathbf{x}', \mathbf{x}) d\mathbf{x} = 1. \quad (8.43)$$

To gain insight into what sampling scheme should be used, first solve the integral equation for $y(\mathbf{x})$ using the Neumann expansion method and then evaluate H . Thus, a solution of the form

$$y(\mathbf{x}) = \sum_{n=0}^{\infty} y_n(\mathbf{x}) \quad (8.44)$$

is sought. Here,

$$y_0(\mathbf{x}) = f(\mathbf{x}), \quad y_1(\mathbf{x}) = \int_V K(\mathbf{x}', \mathbf{x}) y_0(\mathbf{x}') d\mathbf{x}', \quad y_2(\mathbf{x}) = \int_V K(\mathbf{x}', \mathbf{x}) y_1(\mathbf{x}') d\mathbf{x}', \dots,$$

or, in general,

$$y_n(\mathbf{x}) = \int_V K(\mathbf{x}', \mathbf{x}) y_{n-1}(\mathbf{x}') d\mathbf{x}', \quad n = 1, 2, 3, \dots \quad (8.45)$$

It is assumed that the kernel is sufficiently well behaved that all these integrals exist and the series converges. With this *formal* solution for $y(\mathbf{x})$, the functional H is given by

$$H = \int_V d\mathbf{x} h(\mathbf{x}) \sum_{n=0}^{\infty} y_n(\mathbf{x}). \quad (8.46)$$

Notice, that, because of the non-negativity of $K(\mathbf{x}', \mathbf{x})$, the $y_n(\mathbf{x})$ are non-negative and that, because of the normalization of Eq. (8.43), the $y_n(\mathbf{x})$ have unit areas. Thus, $f(\mathbf{x})$ is a PDF characterizing a source, $K(\mathbf{x}', \mathbf{x})$ is a PDF describing the transition of an event going from $\mathbf{x}'$ to $\mathbf{x}$, and $y_n(\mathbf{x})$ is the PDF of $\mathbf{x}$ after n events or transitions. Further, H can be interpreted as the sum of the expected values of H after each event, namely,

$$H = \sum_{n=0}^{\infty} \int_V h(\mathbf{x}) y_n(\mathbf{x}) d\mathbf{x} \equiv \sum_{n=0}^{\infty} \langle h(\mathbf{x}) \rangle_n.$$

But how do you estimate H by using random sampling techniques? Each history consists of a series of points in phase space $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n, \dots$, where $\mathbf{x}_0$ is randomly sampled from $f(\mathbf{x})$, $\mathbf{x}_1$ is sampled from $K(\mathbf{x}_0, \mathbf{x})$, $\mathbf{x}_2$ is sampled from $K(\mathbf{x}_1, \mathbf{x})$, $\dots$, $\mathbf{x}_n$ is sampled from $K(\mathbf{x}_{n-1}, \mathbf{x})$, and so on. Thus, after N histories, the contributions to H from event n is obtained by averaging over all histories, and the expected value of H is then found by summing over all events, i.e.,

$$H \simeq \sum_{n=0}^{\infty} \left[\frac{1}{N} \sum_{s=1}^N h(\mathbf{x}_n^s) \right].$$

In practice, the s th history is truncated after M_s events when $h(\mathbf{x}_{M_s})$ becomes negligibly small. Hence, the estimate of H is

$$H \simeq \frac{1}{N} \sum_{s=1}^N \sum_{n=0}^{M_s} h(\mathbf{x}_n^s). \quad (8.47)$$

This Monte Carlo approach is now generalized to treat integral equations of which the kernels and inhomogeneous terms do not have the special properties of Eq. (8.43).

8.2.2 A More General Procedure for Integral Equations

Equation (8.34) can be written as $y(\mathbf{x}) = f(\mathbf{x}) + \mathcal{K}y(\mathbf{x})$ where the integral operator $\mathcal{K}$ is defined as

$$\mathcal{K}\psi(\mathbf{x}) \equiv \int_V K(\mathbf{x}', \mathbf{x}) \psi(\mathbf{x}') d\mathbf{x}'. \quad (8.48)$$

Rather than use a Neumann expansion, as was done in the previous section, an iterative solution, analogous to the iterative solution of linear algebraic equations given by Eq. (8.44), is used, namely

$$y^{(k+1)}(\mathbf{x}) = f(\mathbf{x}) + \mathcal{K}y^{(k)}(\mathbf{x}) = \sum_{m=0}^k \mathcal{K}^m f(\mathbf{x}), \quad (8.49)$$

with $\mathcal{K}^0 = 1$ and $y^{(0)} = 0$. If this Neumann series $y^{(0)}, y^{(1)}, y^{(2)}, \dots$ converges, then²

$$y(\mathbf{x}) = \lim_{k \rightarrow \infty} y^{(k)}(\mathbf{x}) = \sum_{m=0}^{\infty} \mathcal{K}^m f(\mathbf{x}). \quad (8.50)$$

² Here, it is assumed that $f(\mathbf{x}) \in L_2(V)$ (i.e., f is square integrable) and $K(\mathbf{x}', \mathbf{x}) \in L_2(V \times V)$. A sufficient (but not a necessary) condition for Eq. (8.50) to converge is

$$\sup_{\mathbf{x} \in V} \int_V |K(\mathbf{x}', \mathbf{x})| d\mathbf{x}' < 1.$$

Suppose what is sought is, not $y(\mathbf{x})$, but the inner product

$$H \equiv (h(\mathbf{x}), y(\mathbf{x})) \equiv \int_V h(\mathbf{x})y(\mathbf{x}) d\mathbf{x}. \quad (8.51)$$

This inner product can be estimated by a procedure analogous to that used in [Section 8.1.1](#) by using a *continuous* random walk. A PDF function $p(\mathbf{x})$ is devised to sample the starting position $\mathbf{x}_0$ of each random walk in V . Next, devise a transition function $P(\mathbf{x}', \mathbf{x})$, such that $P(\mathbf{x}', \mathbf{x}) > 0$ if $K(\mathbf{x}', \mathbf{x}) \neq 0$ and

$$\int_V P(\mathbf{x}', \mathbf{x}) d\mathbf{x} = 1. \quad (8.52)$$

Ideally, $p(\mathbf{x})$ and $P(\mathbf{x}', \mathbf{x})$ should be proportional to the magnitudes of $f(\mathbf{x})$ and $K(\mathbf{x}', \mathbf{x})$, respectively. But uniform distributions often give satisfactory results. The inner product (h, y) can be estimated by using the following procedure:

1. Choose a value of k corresponding to k in [Eq. \(8.49\)](#).
2. Select a starting location $\mathbf{x}_0$ in V by sampling from $p(\mathbf{x})$.
3. Create a continuous random walk $\mathbf{x}_0 \rightarrow \mathbf{x}_1 \rightarrow \dots \rightarrow \mathbf{x}_k$, where the probability of moving from state $\mathbf{x}_i$ to another state or location in phase space $\mathbf{x}$ is determined by sampling from $P(\mathbf{x}_i, \mathbf{x})$.
4. For each random walk $s = 1, \dots, N$, the random variable

$$Z_k^{(s)}(h) \equiv \frac{h(\mathbf{x}_0)}{p(\mathbf{x}_0)} \sum_{m=0}^k U_m f(\mathbf{x}_m) \quad (8.53)$$

is calculated. Here, U_m is

$$U_m = \frac{K(\mathbf{x}_0, \mathbf{x}_1)}{P(\mathbf{x}_0, \mathbf{x}_1)} \frac{K(\mathbf{x}_1, \mathbf{x}_2)}{P(\mathbf{x}_1, \mathbf{x}_2)} \dots \frac{K(\mathbf{x}_{m-1}, \mathbf{x}_m)}{P(\mathbf{x}_{m-1}, \mathbf{x}_m)}, \quad \text{with } U_0 = 1.$$

5. After N such random walks, (h, y) is estimated as

$$H = (h(\mathbf{x}), y(\mathbf{x})) \simeq \frac{1}{N} \sum_{s=1}^N Z_k^{(s)}(h).$$

Should the solution $y(\mathbf{x}_0)$ be sought, set $h(\mathbf{x}) = p(\mathbf{x}) = \delta(\mathbf{x} - \mathbf{x}_0)$ so that every random walk begins at $\mathbf{x}_0$ in phase space V . Also, in [Eq. \(8.53\)](#), the term $h(\mathbf{x}_0)/p(\mathbf{x}_0)$ is set to unity. An example of the this procedure is shown in [Example 8.2](#).

Example 8.2 Monte Carlo Solution of an Integral Equation

Consider the integral equation

$$y(x) = x + \lambda \int_0^1 e^{x-x'} y(x') dx'.$$

The analytical solution of this equation, for any real value of $\lambda \neq 1$, is

$$y(x) = x + \frac{\lambda}{1-\lambda} \left[1 - \frac{2}{e} \right] e^x,$$

which is easily verified by substitution into the integral equation.

It can also be shown that $\lambda = 1$ is the only eigenvalue with the corresponding eigenfunction (unnormalized) $y(x) = e^x$. Thus, the spectral radius of the integral operator $\mu(\mathcal{K}) = 1$ and the Neumann series of Eq. (8.49) converges for $\lambda < 1$. Thus, the Monte Carlo solution of this integral equation can be realized only for $\lambda < 1$. Moreover, as $\lambda \rightarrow 1$, the length of the random walk must become ever larger to obtain an estimate of $(h(x), y^{(k)}(x))$ that approximates the true value of the inner product $(h(x), y(x))$, as can be seen from the Monte Carlo results of Fig. 8.2. In this example, $h(x) = 1$ so that the analytical value of the inner product is

$$\begin{aligned} (h, y) &= \int_0^1 h(x)y(x) dx = \int_0^1 \left[x + \frac{\lambda(1-2e^{-1})}{1-\lambda} e^x \right] dx \\ &= \frac{1}{2} + \frac{\lambda}{1-\lambda} \left[e + \frac{2}{e} - 3 \right]. \end{aligned}$$

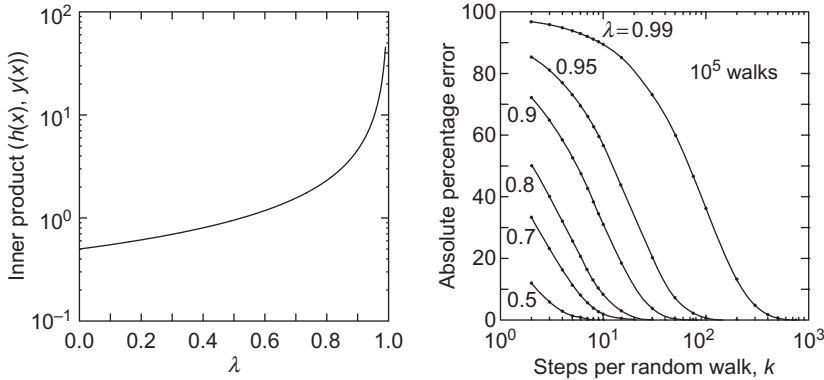


Figure 8.2 The inner product $(h(x), y(x))$ (left) and the percent error for the length of the continuous random walk for different values of λ (right). In these results, $h(x) = 1$. For simplicity, the PDFs used were $p(x) = 1$ and $P(x', x) = 1$.

8.3 Linear Differential Equations

Here, two important differential equations are considered. First, the Poisson equation

$$\nabla^2 u(\mathbf{r}) = -q(\mathbf{r}), \quad \mathbf{r} \in D, \quad (8.54)$$

which reduces to the Laplace equation when $q=0$, is routinely encountered, for example, in steady-state heat conduction problems, electric potential calculations, and absorption-free diffusion analyses. For diffusion problems with absorption, the Helmholtz equation

$$\nabla^2 u(\mathbf{r}) - \beta^2 u(\mathbf{r}) = -q(\mathbf{r}), \quad \mathbf{r} \in D \quad (8.55)$$

is needed. Both equations are usually called boundary value problems because conditions on $u(\mathbf{r})$ on the boundary Γ of D are needed in order to obtain a unique solution.

Analytic solutions of these equations generally can be obtained only for homogeneous media with simple boundaries such as planes, spheres, or cylinders. For heterogeneous media with irregular boundaries, numerical methods must be employed. Even though most boundary value problems described by differential equations are usually far more efficiently evaluated using standard numerical techniques designed for these equations, the Monte Carlo approach is discussed here both for completeness and for application to cases with irregular boundaries where the standard methods often have difficulty.

8.3.1 Monte Carlo Solutions of Linear Differential Equations

The application of Monte Carlo techniques to evaluate solutions of various important partial differential equations received much early attention [Curtiss 1949; Muller 1956]. Even today, work continues on developing new Monte Carlo methods for particular differential equations [Hwang and Mascagni 2001; Mikhailov 1995]. For the differential equations considered here it is assumed that $u(\mathbf{r})$ is specified on the boundary Γ of D (a Dirichlet boundary condition), although other types of boundary conditions can be accommodated [Haji-Sheikh and Sparrow 1967].

The Monte Carlo solutions of linear differential equations can be evaluated by using three different approaches. The first is to use some form of discretization such as finite differences or finite elements to *approximate* the differential equation by a set of linear algebraic equations. These algebraic equations can then be treated using the methods discussed in Section 8.1. Such an approach, which uses random walks over the finite discretization grid, produces results that not only have the statistical uncertainties associated with Monte Carlo estimation but also have truncation errors arising from the discretization process itself. An example of this approach is given in Example 2.4 in which the one-dimensional heat conduction equation is solved by random walks over a discrete grid.

The second approach is to replace the differential equation by an integral equation. This can always be done for a second-order Hermitian differential operator by finding the Green's functions for the specified geometry and boundary conditions. Once an equivalent integral equation is found, the Monte Carlo methods of Section 8.2 can be used to estimate the desired solution.

The third approach is to perform a random walk through the domain D until the boundary Γ is reached. From the value at the boundary, which terminates each random walk, and from the value of the inhomogeneous terms at each stage in the random walk,

an estimate of $\mathbf{r}_0$ can be obtained where $\mathbf{r}_0$ is the starting location of the random walks. The random walks can be performed with each step being a predetermined length (so-called *grid-based* Markov chains) or can be a continuous variable (so-called *continuous Markov chains*). In the sections that follow, this grid-free or continuous Monte Carlo method is explored.

8.3.2 Continuous Monte Carlo for Laplace's Equation

Let us begin the exploration of continuous Monte Carlo by considering the simple two-dimensional heat conduction equation without any internal heat source, i.e.,

$$\frac{\partial^2 u(x, y)}{\partial x^2} + \frac{\partial^2 u(x, y)}{\partial y^2} = 0, \quad (x, y) \in D, \quad \text{with } u(x, y)|_{\Gamma} = \psi(x, y). \quad (8.56)$$

This can be rewritten in polar coordinates, with the origin at an arbitrary position $\mathbf{r}_0 = (x_0, y_0)$, as

$$\frac{\partial^2 u(r, \theta)}{\partial r^2} + \frac{1}{r} \frac{\partial u(r, \theta)}{\partial r} + \frac{1}{r^2} \frac{\partial^2 u(r, \theta)}{\partial \theta^2} = 0. \quad (8.57)$$

Form of solution

To gain insight into how $u(r, \theta)$ varies around $\mathbf{r}_0$, examine the form of the solution to Eq. (8.57). Assume the solution is separable, i.e., $u(r, \theta) = R(r)\Theta(\theta)$, and substitute this into Eq. (8.57) to obtain

$$\frac{r^2}{R} \frac{d^2 R}{dr^2} + \frac{r}{R} \frac{dR}{dr} + \frac{1}{\Theta} \frac{d^2 \Theta}{d\theta^2} = 0. \quad (8.58)$$

The first two terms are a function only of r and the third term is a function only of θ so that the third term must equal a constant (α^2 say) and the first two terms must, therefore, sum to $-\alpha^2$. Thus,

$$\frac{d^2 \Theta}{d\theta^2} + \alpha^2 \Theta(\theta) = 0 \quad (8.59)$$

and

$$\frac{d^2 R}{dr^2} + \frac{1}{r} \frac{dR}{dr} - \frac{\alpha^2}{r^2} R(r) = 0. \quad (8.60)$$

The general solution of Eq. (8.59) is

$$\Theta(\theta) = A \cos \alpha \theta + B \sin \alpha \theta, \quad (8.61)$$

where A and B are arbitrary constants. The cyclic requirement that $\Theta(\theta) = \Theta(\theta + 2\pi)$ requires that α be an integer $\ell = 0, 1, 2, \dots$. The solution of the radial equation, Eq. (8.60), is

$$R(r) = \begin{cases} A \ln r + B, & \ell = 0 \\ Ar^{-\ell} + Br^{\ell}, & \ell > 0 \end{cases}. \quad (8.62)$$

Because $u(r, \theta)$ must be finite as $r \rightarrow 0$, the arbitrary constant A in this result must vanish. Hence, the most general, physically realistic solution of Eq. (8.57) must have the form

$$u(r, \theta) \equiv R(r)\Theta(\theta) = a_0 + \sum_{\ell=1}^{\infty} r^{\ell} [a_{\ell} \cos \ell\theta + b_{\ell} \sin \ell\theta], \quad (8.63)$$

where a_{ℓ} and b_{ℓ} are arbitrary constants of which the values are determined by the boundary conditions imposed on Eq. (8.56).

An important property of the solution

An important property of this solution can be observed by the integration of Eq. (8.63) over all $\theta \in [0, 2\pi]$ to obtain

$$\int_0^{2\pi} u(r, \theta) d\theta = 2\pi a_0. \quad (8.64)$$

Also, the solution at the origin $u(0, \theta) \equiv u(\mathbf{r}_0)$ is independent of θ so that, from Eq. (8.63), $u(\mathbf{r}_0) = a_0$. Combining this result with Eq. (8.64) gives

$$u(\mathbf{r}_0) = \frac{1}{2\pi} \int_0^{2\pi} u(r, \theta) d\theta. \quad (8.65)$$

In words, Eq. (8.65) says that *the value of u at the center of a circle of radius r equals the average of u on the circle.*

Monte Carlo estimation of $u(\mathbf{r}_0)$

Suppose we seek the value of u at some point $\mathbf{r}_0$. Begin by drawing a circle that is centered at $\mathbf{r}_0$ and that just touches the nearest boundary (see Fig. 8.3). Then pick a point $\mathbf{r}_1$ uniformly sampled on this circle. The value of $u(\mathbf{r}_1)$ is then one particle's estimate of $u(\mathbf{r}_0)$. Averaged over many particles, each starting from $\mathbf{r}_0$, Eq. (8.65) shows that the average $\overline{u(\mathbf{r}_1)} \simeq u(\mathbf{r}_0)$.

However, $u(\mathbf{r}_1)$ is generally unknown *unless* $\mathbf{r}_1$ lies on the boundary Γ , in which case $u(\mathbf{r}_1) = \psi(\mathbf{r}_1)$, the specified boundary value. For $\mathbf{r}_1$ not on Γ , $u(\mathbf{r}_1)$ is estimated

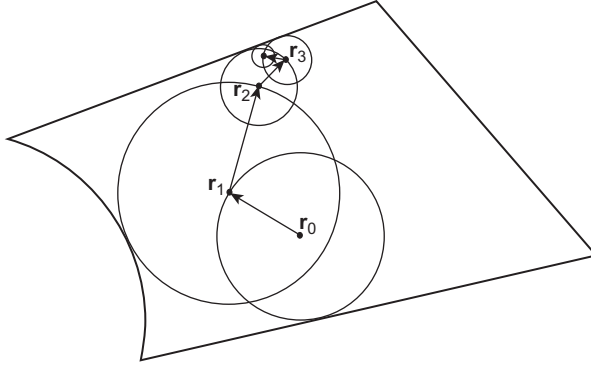


Figure 8.3 The continuous random walk used to estimate $u(\mathbf{r}_0)$. The walking-on-circles continues until the particle is within some small distance ϵ from the boundary.

in the same manner used to estimate $u(\mathbf{r}_0)$, namely, a point $\mathbf{r}_2$ is uniformly sampled on the largest circle wholly in D and centered at $\mathbf{r}_1$ (the *maximal* circle) and $u(\mathbf{r}_2)$ is the particle's estimate of $u(\mathbf{r}_1)$. If $\mathbf{r}_2$ does not lie on the boundary, the random walk continues on successive circles until a point $\mathbf{r}_n$ "on" the boundary is found and the history is terminated. The score or tally for this walk is then recorded as $\psi(\mathbf{r}_n)$. This procedure is repeated for a large number N of particles, all starting from $\mathbf{r}_0$. The estimate of $u(\mathbf{r}_0)$ is then found by averaging all the scores, i.e.,

$$u(\mathbf{r}_0) = \frac{1}{N} \sum_{i=1}^N \psi(\mathbf{r}_{n_i}), \quad (8.66)$$

where $\mathbf{r}_{n_i}$ is the terminating position on the boundary for the i th history.

The only difficulty with this scheme is that randomly selected points on a circle have zero probability of actually being on the boundary. Moreover, as a particle approaches the boundary, the circles get smaller and smaller, requiring more and more computing effort. One way to avoid this is to use a small parameter $\epsilon > 0$, such that if $\mathbf{r}_n$ is within ϵ of the closest boundary point $\mathbf{r}^*$, the walk is terminated and $u(\mathbf{r}_n)$ is set to $\psi(\mathbf{r}^*)$.

The above *walking-on-circles* procedure is exact in the interior. However, only approximate near the boundary because it allows a particle to get only arbitrarily close to the boundary but never actually reach it. Clearly, as ϵ decreases, this approximation error decreases. Although standard numerical techniques for solving the Laplace equation are usually far more efficient, this Monte Carlo procedure can be used to good advantage when the boundary Γ assumes a complicated shape, such as that in Fig. 8.3.

8.3.3 Generalization to Three Dimensions

The previous continuous Monte Carlo procedure for evaluating the solution to the Laplace equation with a Dirichlet boundary condition can be extended directly to arbitrary dimension K [Mikhailov 1995]. In particular, for three dimensions, points are

uniformly sampled on the surfaces of maximal spheres. For a sphere centered at $\mathbf{r}$, the radius of the sphere is

$$r = \min[d_1(\mathbf{r}), d_2(\mathbf{r}), \dots, d_M(\mathbf{r})], \quad (8.67)$$

where the $\{d_i(\mathbf{r})\}$ are the shortest distances from $\mathbf{r}$ to the M elementary surfaces that form the boundary Γ of D and on which the value of u is specified. This method of generating a random walk between points on spheres is known as the *walking-on-spheres* (WOS) algorithm or, perhaps more poetically, the *dancing-on-bubbles* algorithm.

To see the generalization to three dimensions, consider the Laplace equation in three-dimensional spherical coordinates (r, θ, ϕ) centered at some arbitrary point $\mathbf{r}_0 \in D$, namely,

$$\frac{1}{r^2} \frac{\partial}{\partial r} \left[r^2 \frac{\partial u}{\partial r} \right] + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left[\sin \theta \frac{\partial u}{\partial \theta} \right] + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 u}{\partial \phi^2} = 0, \quad (8.68)$$

with $u(r, \theta, \phi)|_{\Gamma} = \psi(\mathbf{r})$ prescribed on the surface. The most general nonsingular solution of this equation is [Margenau and Murphy 1956]

$$u(r, \theta, \phi) = a_0 + \sum_{\ell=1}^{\infty} \sum_{m=1}^{\ell} r^{\ell} P_{\ell}^m(\cos \theta) [a_{\ell} \cos \ell \phi + b_{\ell} \sin \ell \phi], \quad (8.69)$$

where the functions $P_{\ell}^m(\cos \theta)$ are the associated Legendre functions and a_{ℓ} and b_{ℓ} are arbitrary constants. Integration of the result over all directions gives

$$\int_{4\pi} u(\mathbf{r}) d\Omega = \int_0^{\pi} d(\cos \theta) \int_0^{2\pi} d\phi u(r, \theta, \phi) = 4\pi a_0. \quad (8.70)$$

Because $u(0, \theta, \phi) \equiv u(\mathbf{r}_0) = a_0$ and is independent of θ and ϕ , Eq. (8.70) then requires

$$u(\mathbf{r}_0) = \frac{1}{4\pi} \int_{4\pi} u(r, \theta, \phi) d\Omega, \quad (8.71)$$

i.e., the average value of u over the surface of a sphere equals the value of u at the center of the sphere, exactly analogous to Eq. (8.65).

To estimate $u(\mathbf{r}_0)$ by the WOS algorithm, start each particle at $\mathbf{r}_0$ and pick a point $\mathbf{r}_1$ uniformly sampled on the surface of the largest sphere contained in D . Point $\mathbf{r}_1$ is then taken as the center of the next maximal sphere in D and a point $\mathbf{r}_2$ is uniformly sampled on this sphere. The n th particle's walking-on-spheres terminates when a point $\mathbf{r}_{n_i}$ is selected that is within ϵ of the surface Γ , the nearest surface point being $\mathbf{r}_n^*$. The

value of $\psi(\mathbf{r}_n^*)$ is then scored as the tally for the particle. Finally, after N histories, the estimate of $u(\mathbf{r}_0)$ is taken as

$$u(\mathbf{r}_0) \simeq \frac{1}{N} \sum_{i=1}^N \psi(\mathbf{r}_n^*). \quad (8.72)$$

8.3.4 Continuous Monte Carlo for Poisson's Equation

The inhomogeneous form of Laplace's equation, also known as Poisson's equation,

$$\frac{1}{r^2} \frac{\partial}{\partial r} \left[r^2 \frac{\partial u}{\partial r} \right] + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left[\sin \theta \frac{\partial u}{\partial \theta} \right] + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 u}{\partial \phi^2} = -q(\mathbf{r}), \quad (8.73)$$

can also be solved by the WOS Monte Carlo algorithm. First considered is the special case in which $q(\mathbf{r})$ is a constant, followed later by the more general case.

Case: q is a constant

It is readily verified that a particular solution of Eq. (8.73), with $q(\mathbf{r}) = q_o$, is $u_p = -r^2 q_o / 6$ so that the most general nonsingular solution is the solution of the homogeneous equation, Eq. (8.69), plus the particular solution, i.e.,

$$u(r, \theta, \phi) = a_0 + \sum_{\ell=1}^{\infty} \sum_{m=1}^{\ell} r^{\ell} P_{\ell}^m(\cos \theta) [a_{\ell} \cos \ell \phi + b_{\ell} \sin \ell \phi] - q_o \frac{r^2}{6}. \quad (8.74)$$

As before, $u(0, \theta, \phi) \equiv u(\mathbf{r}_0) = a_0$, and integration of Eq. (8.74) over all directions yields $\int_{4\pi} u(\mathbf{r}) d\Omega = 4\pi [a_0 - q_o(r^2/6)]$. Combining these two results yields

$$u(\mathbf{r}_0) = \frac{1}{4\pi} \int_{4\pi} u(\mathbf{r}) d\Omega + Q(r), \quad (8.75)$$

where $Q(r) = q_o r^2 / 6$. This important result says that the value of u at the center of a sphere of radius r equals the average u over the surface *plus* the function $Q(r)$.

Thus, the WOS algorithm again can be used with a slight modification, namely, that at each step in a particle's random walk, the quantity $q_o r_j^2 / 6$ (where r_j is the radius of the j th sphere in the walk) is added to the particle's tally. When the i th particle on the n_i th sphere is within ϵ of Γ , the walk is, as before, terminated and the value of $\psi(\mathbf{r}_{n_i}^*)$ is also added to the tally. After N random walks, the estimate of $u(\mathbf{r}_0)$ is given by

$$u(\mathbf{r}_0) \simeq \frac{1}{N} \sum_{i=1}^N \left[q_o \sum_{j=1}^{n_i} \frac{r_j^2}{6} + \psi(\mathbf{r}_{n_i}^*) \right]. \quad (8.76)$$

In [Example 8.3](#), an example of this WOS algorithm is given. The estimate of [Eq. \(8.76\)](#) can also be used with the two-dimensional Poisson's equation with $Q(r) = q_0 r^2/4$ [[Booth 1981](#)].

Case: $q(\mathbf{r})$ is variable

For the case that $q(\mathbf{r})$ varies throughout D , [Eq. \(8.75\)](#) still is valid but $Q(\mathbf{r})$ is now given by [[Mikhailov 1995](#)]

$$Q(r) = \int_{|\mathbf{r}'| < r} G(0, \mathbf{r}') q(\mathbf{r}') dV', \quad (8.77)$$

where the integration is over all differential volume elements dV' about $\mathbf{r}'$ in a sphere of radius r centered at the origin. Here, $G(\mathbf{r}, \mathbf{r}')$ is the Green's function for a sphere and gives the contribution to $u(\mathbf{r})$ caused by a unit strength source at $\mathbf{r}'$. For the case of a sphere of radius r , the Green's function is

$$G(\mathbf{r}, \mathbf{r}') = \frac{1}{4\pi} \left[\frac{1}{|\mathbf{r} - \mathbf{r}'|} - \frac{1}{r} \right]. \quad (8.78)$$

Note that $G(0, r')$ is spherically symmetric about $\mathbf{r}' = 0$, i.e., there is no θ' or ϕ' dependence, so that dV' in [Eq. \(8.77\)](#) equals $4\pi r'^2 dr'$. Substitution of [Eq. \(8.78\)](#) into [Eq. \(8.77\)](#) yields

$$Q(r) = \frac{1}{4\pi} \int_{|\mathbf{r}'| < r} \left[\frac{1}{|\mathbf{r}'|} - \frac{1}{r} \right] q(\mathbf{r}') dV'. \quad (8.79)$$

Example 8.3 Monte Carlo Solution of Poisson's Equation

Consider Poisson's Equation, [Eq. \(8.73\)](#), with $q = q_0$, a constant. The domain D is a cube with sides of length L parallel to the axes and with the origin in the center of the cube, i.e.,

$$D = \{x, y, z : -L/2 \leq x, y, z \leq L/2\}.$$

On the surface, $u = \psi_0$, a constant. For this problem, the analytic solution can be expressed as

$$u(x, y, z) = \psi_0 + \sum_{\ell=1}^{\infty} \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \frac{64q_0 L^2}{\ell m n \pi^5} \frac{(-1)^{(\ell+m+n-3)/2}}{\ell^2 + m^2 + n^2} \cos \gamma_\ell x \cos \gamma_m y \cos \gamma_n z,$$

where $\sum'$ indicates the summations are over only the odd integers, and $\gamma_i \equiv i\pi/L$. Of interest is the largest value of u in D , which occurs at $\mathbf{r}_0 = (0, 0, 0)$. In this example, the problem parameters are chosen as $L = 10$, $q_0 = 10$, and $\psi = 100$ at all points on the surface. For this particular case, the exact value of $u(0, 0, 0) \simeq 156.21283 \dots$

The WOS algorithm produced the results for this problem as shown in [Fig. 8.4](#). In the left figure, the percent error decreases roughly as $1/N^2$, where N is the number of histories for

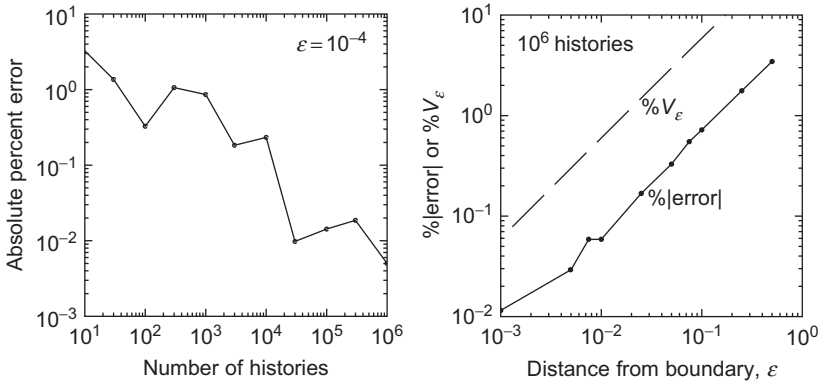


Figure 8.4 The percent error (left) as a function of the number of random walks and (right) as a function of the ϵ parameter.

a very small value of $\epsilon = 0.0001$. Because the Monte Carlo algorithm is only approximate in the ϵ layer, the value of ϵ determines the accuracy of the “converged” Monte Carlo result. In the right graph of Fig. 8.4, the converged accuracy is seen to decrease proportionally to the fraction of the cube’s volume in the ϵ layer, i.e., as $\%V_\epsilon = 100(L^3 - V_\epsilon)/L^3$ decreases, where

$$V_\epsilon = L^3 - (L - 2\epsilon)^3.$$

Rather than evaluate this integral numerically at each step in the random walk, Monte Carlo quadrature can be used. Pick a point $\mathbf{r}'_i$ inside the sphere of radius r from a PDF proportional to $G(0, \mathbf{r}')$. Hence,

$$\mathcal{Q}(r) \simeq q(\mathbf{r}'_i) \frac{1}{4\pi} \int_{|\mathbf{r}'| < r} \left[\frac{1}{|\mathbf{r}'|} - \frac{1}{r} \right] dV' = q(\mathbf{r}'_i) \int_0^r \left[\frac{1}{r'} - \frac{1}{r} \right] r'^2 dr' = q(\mathbf{r}'_i) \frac{r^2}{6}. \quad (8.80)$$

When this approximation is averaged over many random walks, the proper value for Eq. (8.79) is obtained. The random position $\mathbf{r}'_i$, for the i th history, is taken as $\mathbf{r}'_i = r'_i \boldsymbol{\Omega}_i$, where the direction vector $\boldsymbol{\Omega}_i$ is taken from an isotropic distribution and r'_i is sampled from a PDF proportional to

$$\left[\frac{1}{r'} - \frac{1}{r} \right] r'^2 \propto r'(r - r').$$

From Eq. (8.80) it can be seen that at the j th step of the WOS algorithm the value of the particle’s tally is increased by q evaluated at some selected point $\mathbf{r}_j + \mathbf{r}'_j$ in the sphere of radius r_j , centered at $\mathbf{r}_j$, and multiplied by $r_j^2/6$. Thus, the tally for the i th

history, which terminates on step n_i , is computed as

$$T_i = \sum_{j=1}^{n_i} q(\mathbf{r}'_j + \mathbf{r}_j) \frac{r_j^2}{6} + \psi(\mathbf{r}_{n_i}^*), \quad (8.81)$$

where $\mathbf{r}_{n_i}^*$ is the point on the boundary closest to $\mathbf{r}_{n_i}$ that is within ϵ of the boundary.

Finally, when this tally is averaged over many histories N , the estimate of $u(\mathbf{r}_0)$ is obtained, namely,

$$u(\mathbf{r}_0) \simeq \frac{1}{N} \sum_{i=1}^N \left\{ \sum_{j=1}^{n_i} q(\mathbf{r}'_j + \mathbf{r}_j) \frac{r_j^2}{6} + \psi(\mathbf{r}_{n_i}^*) \right\}. \quad (8.82)$$

As would be expected, this result reduces to Eq. (8.76) when $q(\mathbf{r}) \rightarrow q_0$.

8.3.5 The Two-dimensional Helmholtz Equation

Diffusion processes with absorption are governed by the Helmholtz equation

$$\nabla^2 u(\mathbf{r}) - \beta^2 u(\mathbf{r}) = -q(\mathbf{r}), \quad \mathbf{r} \in D, \quad \text{with } u(\mathbf{r})|_{\Gamma} = \psi(\mathbf{r}), \quad (8.83)$$

with $\beta^2 > 0$. In the following subsections, we have relied on Booth [1981], Hwang and Mascagni [2001], and Mikhailov [1995] who present many more results than presented here.

Solution of the homogeneous equation

To see how the previous Monte Carlo methods must be altered for this equation, first consider the homogeneous form of Eq. (8.83) in two dimensions. In polar coordinates with respect to an arbitrary origin at $\mathbf{r}_0$, Eq. (8.83) can be written as

$$\frac{\partial^2 u(r, \theta)}{\partial r^2} + \frac{1}{r} \frac{\partial u(r, \theta)}{\partial r} + \frac{1}{r^2} \frac{\partial^2 u(r, \theta)}{\partial \theta^2} - \beta^2 u(r, \theta) = 0. \quad (8.84)$$

As was done in Section 8.3.2, first separate the dependent variable by assuming $u(r, \theta) = R(r)\Theta(\theta)$. This separation results in two ordinary differential equations (with a separation constant α), namely,

$$\frac{d^2 \Theta}{d\theta^2} + \alpha^2 \Theta(\theta) = 0, \quad (8.85)$$

$$\frac{d^2 R}{dr^2} + \frac{1}{r} \frac{dR}{dr} + \left(-\beta^2 - \frac{\alpha^2}{r^2} \right) R = 0. \quad (8.86)$$

The general solution of the angular equation, Eq. (8.85), subject to the requirement that Θ be cyclic in 2π is

$$\Theta(\theta) = A \cos \ell \theta + B \sin \ell \theta, \quad (8.87)$$

where $\alpha \equiv \ell = 0, 1, 2, \dots$. The radial equation, Eq. (8.86) with α^2 replaced by ℓ^2 , is the modified Bessel's equation of which the two independent solutions are $I_n(\beta r)$ and $K_n(\beta r)$, the modified Bessel functions of order n . Because $u(r, \theta)$ must be finite at $r = 0$, the singular $K_n(\beta r)$ functions are discarded. Thus, the most general physically realistic solution of Eq. (8.84) has the form

$$u(r, \theta) \equiv R\Theta = a_0 I_0(\beta r) + \sum_{\ell=1}^{\infty} I_{\ell}(\beta r) [a_{\ell} \cos \ell \theta + b_{\ell} \sin \ell \theta], \quad (8.88)$$

where a_{ℓ} and b_{ℓ} are constants.

Integration of this solution over θ yields

$$\int_0^{2\pi} u(r, \theta) d\theta = 2\pi a_0 I_0(\beta r). \quad (8.89)$$

The solution at the origin $u(0, \theta) = u(\mathbf{r}_0)$ is seen from Eq. (8.88) to be simply a_0 because $I_0(0) = 1$ and $I_{\ell}(0) = 0$, $\ell = 1, 2, 3, \dots$. From these two results, it follows that

$$u(\mathbf{r}_0) = \frac{1}{I_0(\beta r)} \frac{1}{2\pi} \int_0^{2\pi} u(r, \theta) d\theta. \quad (8.90)$$

This result shows that the solution of the Helmholtz equation at any point $\mathbf{r}_0$ in D equals the average of u over the circumference of a circle of radius r (that is contained in D and centered at $\mathbf{r}_0$) multiplied by the factor $1/I_0(\beta r)$.

To evaluate $u(\mathbf{r}_0)$ by continuous Monte Carlo, a random walking-on-circles procedure can again be used. At each step j in the random walk, the particle's weight W is modified by multiplying its previous weight W_{j-1} by $1/I_0(\beta r_j)$, i.e.,

$$W_j = W_{j-1}/I_0(\beta r_j), \quad j = 1, 2, 3, \dots, n-1, \quad \text{with } W_0 = 1, \quad (8.91)$$

where r_j are the radii of the maximal circles used in the random walking-on-circles. As before, the i th random walk is terminated when $\mathbf{r}_{n_i}$ is within ϵ of the surface Γ , and the score of the walk is set to $W_{n_i}^{(i)} \psi(\mathbf{r}_{n_i}^*)$, where $\mathbf{r}_{n_i}^*$ is the nearest point on Γ from $\mathbf{r}_{n_i}$. Finally, after N random walks the value of $u(\mathbf{r}_0)$ is estimated as

$$u(\mathbf{r}_0) \simeq \frac{1}{N} \sum_{i=1}^N W_{n_i}^{(i)} \psi(\mathbf{r}_{n_i}^*). \quad (8.92)$$

Solution for a constant source term

To obtain the form of the solution to

$$\frac{\partial^2 u(r, \theta)}{\partial r^2} + \frac{1}{r} \frac{\partial u(r, \theta)}{\partial r} + \frac{1}{r^2} \frac{\partial^2 u(r, \theta)}{\partial \theta^2} - \beta^2 u(r, \theta) = -q_o, \quad (8.93)$$

observe that $u_p = q_o/\beta^2$ is a particular solution. Addition of u_p to the most general physically realistic solution of the corresponding homogeneous equation gives the general solution of Eq. (8.93), namely,

$$u(r, \theta) \equiv R\Theta = a_0 I_0(\beta r) + \sum_{\ell=1}^{\infty} I_{\ell}(\beta r) [a_{\ell} \cos \ell \theta + b_{\ell} \sin \ell \theta] + q_o/\beta^2. \quad (8.94)$$

To see how $u(\mathbf{r})$ at the center of a circle is related to the value on the circumference, integrate Eq. (8.94) over θ to obtain

$$\int_0^{2\pi} u(r, \theta) d\theta = 2\pi a_0 I_0(\beta r) + 2\pi q_o/\beta^2. \quad (8.95)$$

The value at the circle center $u(0, \theta) = u(\mathbf{r}_0)$ is independent of θ and, from Eq. (8.94), has the value

$$u(\mathbf{r}_0) = a_0 + q_o/\beta^2. \quad (8.96)$$

Combining these two results to eliminate a_0 gives the important result

$$u(\mathbf{r}_0) = \frac{1}{I_0(\beta r)} \frac{1}{2\pi} \int_0^{2\pi} u(r, \theta) d\theta + Q(r), \quad (8.97)$$

where

$$Q(r) \equiv \frac{q_o}{\beta^2} \left(1 - \frac{1}{I_0(\beta r)} \right). \quad (8.98)$$

From this result it is seen that u at the center of a circle of radius r equals $Q(r)$ plus the average of $u(r, \theta)/I_0(\beta r)$ on the circumference. Thus, the previous random walking-on-circles procedure is altered to score the particle's weight times $Q(\mathbf{r})$ at each step, in addition to scoring the surface contribution at the end of the walk. After N random walks the value of $u(\mathbf{r}_0)$ is then estimated as

$$u(\mathbf{r}_0) \simeq \frac{1}{N} \sum_{i=1}^N \left\{ \sum_{j=1}^{n_i} W_j^{(i)} Q(r_j) + W_{n_i}^{(i)} \psi(r_{n_i}^*) \right\}. \quad (8.99)$$

In Example 8.4, this walking-on-circles procedure is used to solve a particular diffusion problem.

Example 8.4 Monte Carlo Solution of a Diffusion Problem

Consider the two-dimensional Helmholtz equation on a square with constant source q_o and boundary value ψ_o , namely,

$$\frac{\partial^2 u(x, y)}{\partial x^2} + \frac{\partial^2 u(x, y)}{\partial y^2} - \beta^2 u = -q_o, \quad \frac{-L}{2} \leq x, y \leq \frac{L}{2}, \quad \text{with } u(x, y)|_{\Gamma} = \psi_o.$$

For this problem, the analytic solution can be expressed as

$$u(x, y) = \psi_o + \sum_{n=1}^{\infty} \frac{S_n}{\gamma_n^2} \left[1 - \frac{\cosh \gamma_n x}{\cosh \gamma_n L/2} \right] \cos \frac{n\pi}{L} y,$$

where $\sum'$ indicates the summation is over only the odd integers. Here,

$$\gamma_n = \left(\frac{n\pi}{L} \right)^2 + \beta^2 \quad \text{and} \quad S_n = \frac{4}{n\pi} [q_o - \beta^2 \psi_o] (-1)^{(n-1)/2}.$$

For the case $L = 10$, $q_o = \psi_o = 100$, and $\beta^2 = 0.2$ the numerical value of $u(0, 0)$ at the center of the square is computed to be $u_{\text{exact}}(0, 0) \simeq 371.62142$.

The walking-on-circles algorithm, discussed previously, produced the results for $u(0, 0)$ as shown in Fig. 8.5. In the left figure, the percent error decreases roughly as $1/N^2$, where N is the number of histories for a very small value of $\epsilon = 0.00001$. Because the Monte Carlo algorithm is only approximate in the ϵ layer, the value of ϵ determines the accuracy of the “converged” Monte Carlo result (here, assumed, to be obtained after 10^7 histories). In the right graph of Fig. 8.5, the converged accuracy is seen to decrease as the fraction of the square’s area in the ϵ layer decreases, i.e., as $\%A_\epsilon = 100(L^2 - A_\epsilon)/L^2$ decreases, where $A_\epsilon = L^2 - (L - 2\epsilon)^2$. Although initially the decrease in the error is directly proportional to $\%A_\epsilon$, it decreases less rapidly as ϵ becomes very small because the small boundary layer has less effect on u at the center than does the internal source q_o .

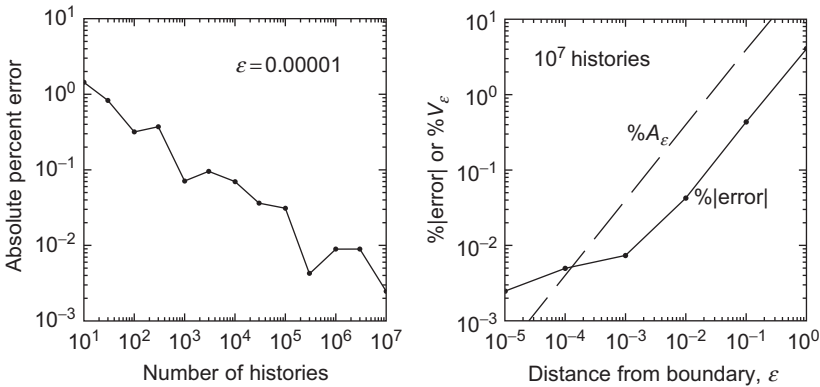


Figure 8.5 The percent error (left) as a function of the number of random walks and (right) as a function of the ϵ parameter.

Case for negative β^2

Sometimes for a multiplying medium, for example neutrons diffusing in a medium containing fissionable material, the constant β^2 in Eq. (8.83) may become negative. If $-\beta^2 < \lambda^*$, where λ^* is the smallest eigenvalue of the system, a unique solution for $u(\mathbf{r})$ still exists. Let $\gamma^2 = -\beta^2 > 0$. Then repeating the above analysis produces the same result except $I_0(\beta r)$ is replaced by $J_0(\gamma r)$, the regular Bessel function of the first kind (see problem 9).

8.3.6 The Three-dimensional Helmholtz Equation

The three-dimensional diffusion (Helmholtz) equation, Eq. (8.83) with $\beta^2 > 0$, can also be solved by the walking-on-spheres algorithm. As shown by Mikhailov [1995], the weight of the particle as it leaves the j th sphere $j = 0, 1, 2, 3, \dots$ with radius r_j , for $\beta^2 > 0$, is given by

$$W_j = \frac{\beta r_j}{\sinh(\beta r_j)} W_{j-1}, \quad j = 1, 2, 3, \dots, n-1, \quad \text{with } W_0 = 1. \quad (8.100)$$

Here, the quantity $(\beta r_j)/\sinh(\beta r_j)$ is the probability that the particle, beginning the diffusion process from the center of the sphere, escapes absorption and reaches the surface (for the first time; see problem 5). At each step in the random walk, the tally for the particle is augmented by a source contribution $W_j Q_j$, where

$$Q_j = \frac{1}{4\pi} \int_{V_j} \frac{\sinh(\beta[r_j - |\mathbf{r}_j - \mathbf{r}'|])}{|\mathbf{r}_j - \mathbf{r}'| \sinh(\beta r_j)} q(\mathbf{r}') dV' \quad (8.101)$$

and the integration is over the volume of the j th sphere. This integral could be evaluated by standard numerical techniques, but, more aptly, Monte Carlo methods can be used. A point $\mathbf{r}'_j$ is picked randomly and uniformly in the j th sphere. Then evaluation of the integrand of Eq. (8.101) at $\mathbf{r}'_j$ multiplied by the sphere's volume $V_j = (4/3)\pi r_j^3$ gives the source contribution for the j th step as

$$Q_j = \frac{r_j^3}{3} \frac{\sinh(\beta[r_j - |\mathbf{r}_j - \mathbf{r}'_j|])}{|\mathbf{r}_j - \mathbf{r}'_j| \sinh(\beta r_j)} q(\mathbf{r}'_j). \quad (8.102)$$

This quantity, averaged over many particle walks, then produces a Monte Carlo estimate of Eq. (8.101). For the case that $q(\mathbf{r}) = q_o$, a constant, the integral in Eq. (8.101) can be evaluated analytically as (see problem 7)

$$Q_j = \frac{q_o}{\beta^2} \left[1 - \frac{\beta r_j}{\sinh(\beta r_j)} \right]. \quad (8.103)$$

The tally for the i th random walk, which reaches a point within ϵ of the boundary on step n_i , is

$$T_i = \sum_{j=1}^{n_i} W_j Q_j + W_{n_i} \psi(\mathbf{r}_{n_i}^*), \quad (8.104)$$

where $\mathbf{r}_{n_i}^*$ is the nearest point on Γ from $\mathbf{r}_{n_i}$. Thus, after N random walks, the estimate of $u(\mathbf{r}_0)$ is

$$u(\mathbf{r}_0) = \frac{1}{N} \sum_{i=1}^N T_i. \quad (8.105)$$

The Monte Carlo WOS algorithm for treating the Helmholtz equation can be extended to cases where, not only is $q(\mathbf{r})$ variable, but β^2 , the parameter that determines how readily a diffusing particle is absorbed (or multiplied), varies with position in the medium [Mikhailov 1995]. The WOS method can also be extended to an arbitrary number of dimensions in which the random walks are performed on the surfaces of hyperspheres [Booth 1981; Mikhailov 1995].

8.4 Eigenvalue Problems

Monte Carlo techniques can also be used to find eigenvalues and eigenfunctions of many linear operators. The eigenvalue problem for the three linear operator equations considered in this chapter are as follows:

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}, \quad (8.106)$$

$$y(\mathbf{x}) = \lambda \int_V K(\mathbf{x}', \mathbf{x}) y(\mathbf{x}') d\mathbf{x}' \equiv \lambda \mathcal{K}y(\mathbf{x}), \quad (8.107)$$

and

$$\nabla^2 u(\mathbf{x}) - \beta^2 u(\mathbf{x}) + \lambda u(\mathbf{x}) = 0. \quad (8.108)$$

These homogeneous equations generally have only the trivial null solution. However, for certain values of λ (the *eigenvalues*) there may exist nontrivial solutions (the *eigenfunctions* or *eigenvectors*). Monte Carlo techniques usually can find only the extreme eigenvector and associated eigenfunction. Of most interest are *Hermitian* (self-adjoint) operators because their eigenvalues are real and the eigenfunctions form a complete set of functions that can be used to expand other functions. The familiar Fourier series is one such example. A matrix $\mathbf{A}$ is Hermitian if it is symmetric, and the integral operator is Hermitian if $K(\mathbf{x}, \mathbf{x}') = K(\mathbf{x}', \mathbf{x})$. Equation (8.108) has a Hermitian differential operator.

Consider the matrix eigenvalue problem of Eq. (8.106). For almost any $\mathbf{x}_0$, the largest eigenvalue of $\mathbf{A}$ is given by

$$\lambda_{\max} = \lim_{k \rightarrow \infty} \frac{(\mathbf{x}_0, \mathbf{A}^{k+1} \mathbf{x}_0)}{(\mathbf{x}_0, \mathbf{A}^k \mathbf{x}_0)}. \quad (8.109)$$

The evaluation of the inner products is readily performed using the Monte Carlo techniques described in Section 8.1. The eigenvector associated with $\lambda_{\max}$ is $\mathbf{x}_{\max} = \lim_{k \rightarrow \infty} \mathbf{A}^k \mathbf{x}_0$.

Similarly, the largest eigenvector of Eq. (8.107) can be found from

$$\lambda_{\max} = \lim_{k \rightarrow \infty} \frac{(f(\mathbf{x}), \mathcal{K}^{k+1} f(\mathbf{x}))}{(f(\mathbf{x}), \mathcal{K}^k f(\mathbf{x}))}, \quad (8.110)$$

for almost any arbitrary $f(\mathbf{x})$. The eigenfunction associated with this eigenvalue is given by $y_{\max}(\mathbf{x}) = \lim_{k \rightarrow \infty} \mathcal{K}^k f(\mathbf{x})$.

For additional techniques to estimate eigenvalues, the reader should consult Hammersley and Handscomb [1964].

8.5 Summary

In this chapter, Monte Carlo was applied to a variety of linear equations: algebraic, integral, and differential. Linear algebraic equations are almost always better solved by standard numerical methods. However, multidimensional integral equations and differential equations with irregular boundaries are often best treated by Monte Carlo techniques. Because most problems in the real world can be modeled by linear equations, it is, thus, not surprising that Monte Carlo analysis can be applied to a large variety of problems. Although the analyses presented in this chapter are very analytical in nature and do not immediately suggest simulations at an underlying level, an attempt has been made to demonstrate how the Monte Carlo algorithms proposed are, fundamentally, random walks through phase space, that, with the accumulation of random variates, allow quantification of the solution to these linear equations.

Bibliography

- BOOTH T.E. 1981, "Exact Monte Carlo Solution of Elliptic Partial Differential Equations," *J. Comput. Phys.*, **39**, 396–404.
- CURTISS, J.H. 1949, "Sampling Methods Applied to Differential and Difference Equations," *Proc. Seminar on Scientific Computing*, IBM Corp., New York.
- FORSYTHE, G.E. AND R.A. LIEBLER 1950, "Matrix Inversion by a Monte Carlo Method," *Math. Tables Other Aids Comput.*, **4**, 127–129.
- HAJI-SHEIKH, A. AND E.M. SPARROW 1967, "The Solution of Heat Conduction Problems by Probability Methods," *J. Heat Transf.*, **89**, 121–131.
- HAMMERSLEY, J.M. AND D.C. HANDSCOMB 1964, *Monte Carlo Methods*, Wiley, New York.

- HWANG, C. AND M. MASCAGNI 2001, "Efficient Modified 'Walk on Spheres' Algorithm for the Linearized Poisson-Boltzmann Equation," *Appl. Phys. Lett.*, **78**(6), 787–789.
- MARGENAU, H. AND G.M. MURPHY 1956, *The Mathematics of Physics and Chemistry*, Van Nostrand, Princeton, NJ.
- MIKHAILOV, G.A. 1995, *New Monte Carlo Methods with Estimating Derivatives*, VSP, Utrecht, The Netherlands.
- MULLER, M.E. 1956, "Some Continuous Monte Carlo Methods for the Dirichlet Problem," *Ann. Math. Stat.*, **27**(3), 569–589.
- PRESS, W.H., S.A. TEUKOLSKY, W.T. VETTERLING, AND B.P. FLANNERY 1996, *Numerical Recipes*, 2nd ed., Cambridge University Press, Cambridge.
- RUBINSTEIN, R.V. 1981, *Simulation and the Monte Carlo Method*, Wiley, New York.
- SABELFELD K.K. 1991, *Monte Carlo Methods in Boundary Value Problems*, Springer-Verlag, Berlin.
- WACHSPRESS, E.L. 1966, *Iterative Solution of Elliptic Systems*, Prentice-Hall, Englewood Cliffs, NJ.

Problems

1. Write a program to solve the linear algebraic equations of [Example 8.1](#) using the adjoint random-walk method of [Section 8.1.2](#). Explore how the accuracy of the Monte Carlo solution varies with the number of walks used and with the length k of each walk.
2. Write a program to solve the linear algebraic equations of [Example 8.1](#) using random walks with absorption states as described in [Section 8.1.3](#). Use the following transition matrix:

$$\mathbf{P} = \begin{pmatrix} 0.60 & 0.10 & 0.10 & 0.10 \\ 0.10 & 0.60 & 0.10 & 0.10 \\ 0.05 & 0.10 & 0.60 & 0.05 \\ 0.20 & 0.20 & 0.20 & 0.20 \end{pmatrix}.$$

Explore how the accuracy of your estimate of $(\mathbf{h}, \mathbf{x})$ varies with the number of histories. Also try different transition matrices to determine how increasing the strength of the absorption influences how rapidly the Monte Carlo converges.

3. Consider the integral equation in [Example 8.2](#) with $\lambda = 0.95$. Write a program to evaluate $y(0.5)$ by the continuous random-walk method discussed in [Section 8.2.2](#). Compare your estimates for different N to the exact value of $y(0.5)$. Also determine the effect of using different PDFs $p(x)$ and $P(x', x)$ used to generate the random walks.
4. Show that the probability that a diffusing particle born at the center of a diffusing sphere of radius R reaches the surface is

$$P_{esc} = \frac{\beta R}{\sinh(\beta R)},$$

where $\beta^2 = \Sigma_a/D$ and Σ_a is the absorption coefficient (the probability of absorption per unit differential distance of travel) and D is the diffusion coefficient.

HINT: Solve the diffusion equation with spherical symmetry:

$$\frac{1}{r^2} \frac{d}{dr} \left(r^2 \frac{d\phi(r)}{dr} \right) - \beta^2 \phi(r) = -q(r), \quad 0 \leq r \leq R, \quad \phi(R) = 0, \quad q(r) = \delta(r).$$

First, show that the solution of this equation for $\phi(r)$ is

$$\phi(r) = \frac{1}{8\pi Dr \sinh \beta R} \left[e^{\beta(R-r)} - e^{-\beta(R-r)} \right].$$

Then the probability the particle is absorbed before it reaches the surface is $P_{abs} = \int \Sigma_a \phi(r) dV = \Sigma_a \int_0^R \phi(r) 4\pi r^2 dr$. Finally, the probability of reaching the surface is $P_{esc} = 1 - P_{abs}$.

5. For diffusion problems, governed by the Helmholtz equation, in which the diffusing particle can interact with the medium and produce new particles, such as occurs for neutrons diffusing in a fissionable medium, the parameter β^2 in Eq. (8.83) may become negative. Using methods analogous to those used in problem 4, show that the probable number of particles reaching the surface, per source particle emitted at $r = 0$, is

$$\frac{\beta R}{\sin(\beta R)}, \quad \text{where} \quad \beta = \sqrt{|\beta^2|}.$$

6. With the use of L'Hopital's rule to evaluate indeterminate ratios, show that as $\beta \rightarrow 0$ the constant source contribution at each step of the WOS algorithm of Eq. (8.103) reduces to

$$\lim_{\beta \rightarrow 0} \frac{q_o}{\beta^2} \left[1 - \frac{\beta d_i}{\sinh(\beta d_i)} \right] = \frac{q_o d_i^2}{6},$$

which is consistent with Eq. (8.80).

7. For the case that $q(\mathbf{r}) = q_o$, a constant, show that Eq. (8.101) reduces to Eq. (8.103).
8. Consider the three-dimensional diffusion equation for a cube

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} - \beta^2 u = -q_o, \quad -\frac{L}{2} \leq x, y, z \leq \frac{L}{2},$$

where the value on the boundary is $u|_{\Gamma} = u_o$, a constant. Show that the solution can be expressed as

$$u(x, y, z) = u_o + \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \frac{S_{nm}}{\gamma_{nm}^2} \left[1 - \frac{\cosh \gamma_{nm} x}{\cosh \gamma_{nm} L/2} \right] \cos\left(\frac{n\pi}{L} y\right) \cos\left(\frac{m\pi}{L} z\right).$$

Here,

$$\gamma_{nm}^2 = \left(\frac{n\pi}{L}\right)^2 + \left(\frac{m\pi}{L}\right)^2 + \beta^2$$

and

$$S_{nm} = \frac{16}{nm\pi^2} [q_o - \beta^2 u_o] (-1)^{(n+m-2)/2}.$$

Write a Monte Carlo WOS program that evaluates $u(0, 0, 0)$ at the center of the cube for $L = 10$, $u_o = 100$, $\beta^2 = 0.2$, and $q_o = 100$. Construct graphs that show how the error decreases with the number of random walks N and how the converged error varies with ϵ .

9. Perform an analysis of the two-dimensional Helmholtz equation with $\beta^2 < 0$ similar to that of Section 8.3.5.

9 The Fundamentals of Neutral Particle Transport

*Particles that migrate through matter
do so without any clatter.
Now some, very meek,
eventually leak;
the rest are absorbed or they scatter.*

So far the basis of Monte Carlo has been explored in several contexts, and many of the tricks that make up the tool box of Monte Carlo have been identified. It is now time to begin the serious applications of Monte Carlo and its tools to important problems that are far more complex than the simple example problems shown to this point. Monte Carlo calculations are used in virtually every discipline involved with quantitative analyses. A book the size of this one cannot consider every application of Monte Carlo. In this and the following chapter, the application of Monte Carlo to the problem of radiation transport is considered, not because it is the most important application, but because we, the authors, are most familiar with this particular problem.

Particle transport refers to the diffusion or transport of small particles, such as protons, neutrons, photons, electrons, ions, or even neutral atoms, within some host medium. What is the host medium? It can be a uniform medium such as air or water, it could be a star, it could be the earth's atmosphere, it could be a lead brick, it could be the human body (or a part of it), or it could be just about anything. A special case is a vacuum, which can be thought of as a medium with no constituents (other than the particles that are streaming through it). The point is that small particles, like neutrons in a reactor or X-rays in the body, interact with the host medium (unless it is a vacuum, in which case they simply stream in straight lines away from their point of birth). The problem is to calculate information about the particle field, for example, the number of particles in a certain volume or the energy transferred from the particles to the host medium.

However, before Monte Carlo is applied to radiation transport, it is necessary to introduce some generic aspects of the radiation field and how the radiation particles interact with the host medium. In this generic discussion, the term *radiation* refers to the particles that are migrating through the host medium. This is a slightly loose use of the term (because, for instance, neutral atoms are not typically thought of as radiation), but it is a shorthand notation that serves well as the generic case is considered. Some specific particle types are considered in later subsections. The term *host medium* or simply *medium* refers to the background space through which the particles are transported, whether it be a single uniform medium, a heterogeneous medium, or

even a medium with one or more regions of vacuum. The discussion that follows is restricted to neutral particles, i.e., neutrons or photons, for simplicity and due to space limitations.

9.1 Description of the Radiation Field

To study particle transport a few quantities must first be introduced. The “strength” of the radiation field must be quantified as must also the interactions of the radiation with the host medium. Here are summarized those quantities with which the reader must be aware to understand how to apply Monte Carlo to particle transport problems. The discussion is somewhat abbreviated but should provide the required essentials. For a fuller discussion of terms and concepts, the interested reader is referred to excellent texts such as [Duderstadt and Martin \[1979\]](#), [Lewis and Miller \[1984\]](#), and [Bell and Glasstone \[1970\]](#). Those who are already familiar with this material may skip to later sections.

The first thing to realize is that particle transport attempts to quantify the expected behavior of a large number of particles. The laws that govern particle transport are largely probabilistic in nature. There is always a granularity to the radiation field. Even under steady-state conditions, the number of particles in a small volume fluctuates from one instant of time to another. This stochastic nature of a radiation field arises from both the random emission of radiation by sources and the random manner in which radiation interacts with the medium before reaching the small test volume. Thus, the quantification of a radiation field can be made only in terms of ensemble averages of large numbers of particles. These averages are just expected values and hence are ripe for evaluation by Monte Carlo.

9.1.1 Directions and Solid Angles

The directional properties of radiation fields are almost universally described using spherical polar coordinates as illustrated in [Fig. 9.1](#). The direction vector $\mathbf{\Omega}$ is a unit vector, given in terms of the orthogonal Cartesian unit vectors $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ by

$$\mathbf{\Omega} = \mathbf{i}u + \mathbf{j}v + \mathbf{k}w = \mathbf{i} \sin \theta \cos \psi + \mathbf{j} \sin \theta \sin \psi + \mathbf{k} \cos \theta. \quad (9.1)$$

By increasing θ by $d\theta$ and ψ by $d\psi$, an area $dA = \sin \theta d\theta d\psi$ is swept out on a unit sphere. The solid angle, contained in the range of directions through some area on a sphere, is defined as the area divided by the square of the sphere’s radius. Thus, the differential solid angle associated with differential area dA is

$$d\Omega = \sin \theta d\theta d\psi. \quad (9.2)$$

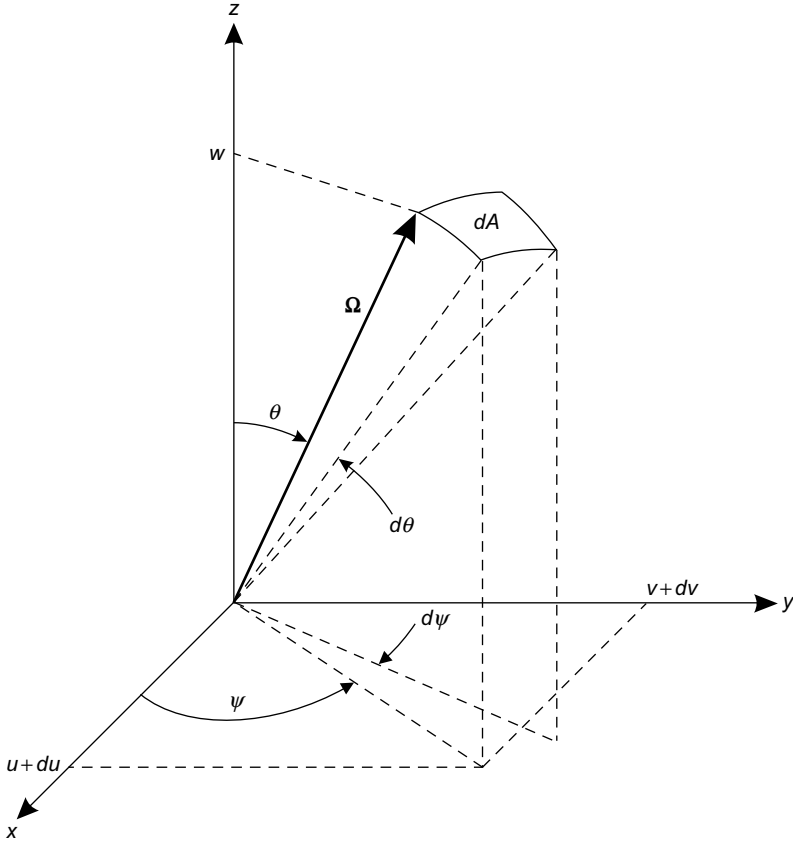


Figure 9.1 Spherical polar coordinate system for specification of the unit direction vector Ω , polar angle θ , azimuthal angle ψ , and associated direction cosines (u, v, w) .

From its definition, the solid angle is a dimensionless quantity. However, to avoid confusion when referring to distribution functions of direction, units of *steradians*, with the abbreviation sr, are assigned to the solid angle.

A frequent simplification in notation is realized by using $\omega = \cos \theta$ as an independent variable instead of the angle θ . With this variable, $\sin \theta d\theta \equiv -d\omega$. An immediate benefit of this new variable is a more straightforward integration over direction and the avoidance of the common error in which the differential solid angle is incorrectly stated as $d\theta d\psi$ instead of $\sin \theta d\theta d\psi$. The benefit is also evident when the solid angle subtended by “all possible directions” is computed, namely,

$$\Omega = \int_0^\pi d\theta \sin \theta \int_0^{2\pi} d\psi = \int_{-1}^1 d\omega \int_0^{2\pi} d\psi = \int_{4\pi} d\Omega = 4\pi. \quad (9.3)$$

9.1.2 Particle Density

The expected number of radiation particles per unit volume, i.e., the particle density, can be used to describe a radiation field at some point $\mathbf{r}$ of interest. Generally, the particles in the unit volume have a distribution of energies E , and the density will vary in time if the sources of the radiation are time varying. The *particle density spectrum* $n(\mathbf{r}, E, t)$ is defined such that

$$n(\mathbf{r}, E, t) dV dE = \text{the expected number of particles in a differential volume } dV \text{ at position } \mathbf{r} \text{ that have energies in } dE \text{ about } E \text{ at time } t. \quad (9.4)$$

Unfortunately, there is no exact equation of which the solution yields $n(\mathbf{r}, E, t)$. However, if the particle density is generalized somewhat, a quantity can be found that does satisfy an exact equation. To this end the *angular particle density spectrum* $n(\mathbf{r}, E, \boldsymbol{\Omega}, t)$ is introduced¹ such that

$$n(\mathbf{r}, E, \boldsymbol{\Omega}, t) dV dE d\Omega = \text{the expected number of particles in a differential volume } dV \text{ at position } \mathbf{r} \text{ traveling in a range of directions in } d\Omega \text{ about } \boldsymbol{\Omega} \text{ with energies in } dE \text{ about } E \text{ at time } t. \quad (9.5)$$

If a particle has velocity $\mathbf{v}$, then the unit vector describing its direction of travel is simply $\boldsymbol{\Omega} = \mathbf{v}/|\mathbf{v}|$. From its definition, $n(\mathbf{r}, E, \boldsymbol{\Omega}, t)$ has units of, for example, $\text{cm}^{-3} \text{ MeV}^{-1} \text{ sr}^{-1}$.

Once $n(\mathbf{r}, E, \boldsymbol{\Omega}, t)$ has been found, it provides a very detailed description of the radiation field (a function of seven independent variables), but usually it is too detailed to be of direct use. Simpler and less detailed descriptions of the radiation field can, however, be readily obtained by integrating the angular density spectrum. For example, the particle *angular density* is

$$n(\mathbf{r}, \boldsymbol{\Omega}, t) = \int_0^\infty n(\mathbf{r}, E, \boldsymbol{\Omega}, t) dE \quad [\text{cm}^{-3} \text{ sr}^{-1}]. \quad (9.6)$$

Similarly, the *density spectrum* is

$$n(\mathbf{r}, E, t) = \int_{4\pi} n(\mathbf{r}, E, \boldsymbol{\Omega}) d\Omega \quad [\text{cm}^{-3} \text{ MeV}^{-1}], \quad (9.7)$$

and the *particle density* is

$$n(\mathbf{r}, t) = \int_0^\infty \int_{4\pi} n(\mathbf{r}, E, \boldsymbol{\Omega}, t) d\Omega dE \quad [\text{cm}^{-3}]. \quad (9.8)$$

¹ The particle density can be further disaggregated to include other properties of the particles besides their direction of travel—for example, the particles' polarizations or spin states, excitation states for composite particles, etc. However, in this chapter such refinements are not needed.

9.1.3 Flux Density

Particle density is a straightforward concept. It is a measure of how many particles occupy a unit volume at any time t , just as we might think of the people density as the number of people per unit volume in a high-rise office building on different floors at some time of day. However, as will be seen, a more useful measure of a radiation field is the particle *angular flux density spectrum* defined as

$$\phi(\mathbf{r}, E, \mathbf{\Omega}, t) \equiv v(E)n(\mathbf{r}, E, \mathbf{\Omega}, t) \quad [\text{cm}^{-2} \text{ MeV}^{-1} \text{ sr}^{-1} \text{ s}^{-1}], \quad (9.9)$$

where $v(E)$ is the particle speed. For photons, $v = c$, the speed of light, while for non-relativistic neutrons with mass m , $v = \sqrt{2E/m}$. Similarly, one has the *angular flux density*

$$\phi(\mathbf{r}, \mathbf{\Omega}, t) = \int_0^\infty v(E) n(\mathbf{r}, E, \mathbf{\Omega}, t) dE \quad [\text{cm}^{-2} \text{ sr}^{-1} \text{ s}^{-1}], \quad (9.10)$$

the *flux density spectrum*

$$\phi(\mathbf{r}, E, t) = \int_{4\pi} v(E)n(\mathbf{r}, E, \mathbf{\Omega}) d\Omega = v(E)n(\mathbf{r}, E, t) \quad [\text{cm}^{-2} \text{ MeV}^{-1} \text{ s}^{-1}], \quad (9.11)$$

and the *scalar flux density* or just the *flux density*

$$\phi(\mathbf{r}, t) = \int_0^\infty \int_{4\pi} v(E) n(\mathbf{r}, E, \mathbf{\Omega}, t) d\Omega dE \quad [\text{cm}^{-2} \text{ s}^{-1}]. \quad (9.12)$$

Because $v(E)$ is the distance traveled in a unit time by a particle of energy E , it follows that $\phi(\mathbf{r}, t)$ can be interpreted as the total distance traveled in a unit time by all particles within a unit volume about $\mathbf{r}$ at time t .

To visualize this interpretation, consider the small sphere centered at $\mathbf{r}$ and of volume ΔV (see Fig. 9.2(b)). Let $\sum_i s_i$ be the sum of all the path-length segments traversed within the sphere by all particles entering the sphere in a unit time at time t . Then the flux density is

$$\phi(\mathbf{r}, t) = \lim_{\Delta V \rightarrow 0} \left[\frac{\sum_i s_i}{\Delta V} \right]. \quad (9.13)$$

Similar interpretations hold for the angular flux density spectrum, angular flux density, and flux density spectrum. This interpretation of flux density as being the particle

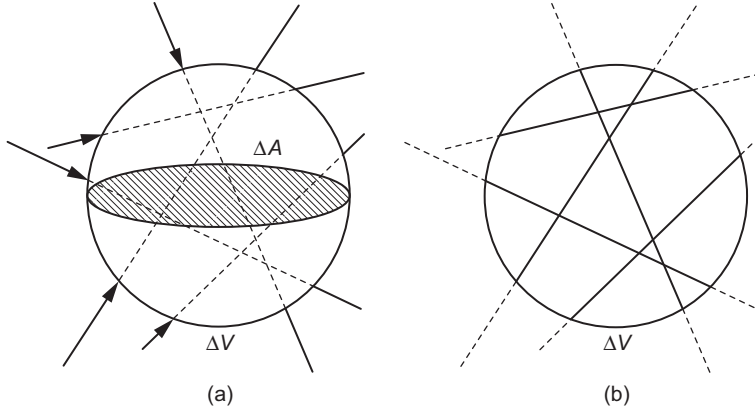


Figure 9.2 Two interpretations of the flux density can be made by consideration of a sphere with volume ΔV and cross-sectional area ΔA . In (a), the number of particles passing through the surface *into* the sphere is of interest. In (b), the paths traveled within the sphere by particles passing through the sphere are of interest.

path length per unit time is important in the next section when the rate of radiation interactions with the medium is discussed.

An alternative interpretation of the concept of flux density is made in terms of the number of particles ΔN_p entering a small sphere of cross-sectional area ΔA in a unit time at time t (see Fig. 9.2(a)). The flux density is then

$$\phi(\mathbf{r}, t) = \lim_{\Delta A \rightarrow 0} \left[\frac{\Delta N_p}{\Delta A} \right]. \quad (9.14)$$

In problem 1 at the end of this chapter, the reader is asked to show that this alternative interpretation of the flux density is equivalent to the path length per unit volume concept.

9.1.4 Fluence

Related to flux density is a quantity called *fluence*, which is defined as the integral of the flux density over some time interval $[t_o, t]$, namely,

$$\Phi(\mathbf{r}, t) = \int_{t_o}^t \phi(\mathbf{r}, t') dt'. \quad (9.15)$$

Conversely, the flux density is the time rate of change of the fluence, i.e.,

$$\phi(\mathbf{r}, t) = \frac{d\Phi(\mathbf{r}, t)}{dt}. \quad (9.16)$$

9.1.5 Current Vector

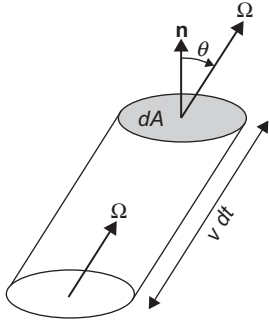


Figure 9.3 Flow through a surface.

Sometimes the flow through a surface is needed in transport calculations. Consider the small surface dA with unit normal $\mathbf{n}$ shown in Fig. 9.3. The surface is located at position $\mathbf{r}$ and the density at time t of particles with energies in dE about E and moving in directions $d\Omega$ about Ω is $n(\mathbf{r}, E, \Omega, t) dE d\Omega$. The number of these particles that cross dA in the following time interval dt is the number of such particles in the slant cylinder of axial length $v dt$ shown in Fig. 9.3. The volume of this cylinder is $(v dt)(dA \cos \theta) = \mathbf{n} \cdot \mathbf{v} dA = v \mathbf{n} \cdot \Omega dA$. Hence, the number of particles crossing dA in a time of dt is $n(\mathbf{r}, E, \Omega, t) v dt \mathbf{n} \cdot \Omega dA dE = \mathbf{n} \cdot \Omega \phi(\mathbf{r}, E, \Omega, t) dA dt dE$.

The *net* number of particles with energies in dE about E crossing dA per unit time in the direction of positive $\mathbf{n}$ regardless of their direction of travel is thus

$$\left[\int_{4\pi} \mathbf{n} \cdot \Omega \phi(\mathbf{r}, E, \Omega, t) d\Omega \right] dE dA = \left[\int_{4\pi} \mathbf{n} \cdot \mathbf{j}(\mathbf{r}, E, \Omega, t) d\Omega \right] dE dA,$$

where the *current vector spectrum* is defined as

$$\mathbf{j}(\mathbf{r}, E, \Omega, t) \equiv \Omega \phi(\mathbf{r}, E, \Omega, t). \quad (9.17)$$

Then the current vector is

$$\mathbf{j}(\mathbf{r}, E, t) \equiv \int_{4\pi} \Omega \phi(\mathbf{r}, E, \Omega, t) d\Omega. \quad (9.18)$$

More generally, the net number of particles crossing a unit area of surface with unit normal $\mathbf{n}$ in the positive direction (the direction of $\mathbf{n}$) per unit energy per unit time is

$$j_n(\mathbf{r}, E, t) = \mathbf{n} \cdot \mathbf{j}(\mathbf{r}, E, t) = \int_{4\pi} \mathbf{n} \cdot \Omega \phi(\mathbf{r}, E, \Omega, t) d\Omega. \quad (9.19)$$

Finally, the *partial currents* are sometimes needed. These quantities are the total flow of particles, per unit energy, through a unit area in the positive direction and in the negative direction, i.e., $j_n(\mathbf{r}, E, t) = j_n^+(\mathbf{r}, E, t) - j_n^-(\mathbf{r}, E, t)$, where the partial currents are given by

$$j_n^+ = \int_{\mathbf{n} \cdot \Omega > 0} \mathbf{n} \cdot \Omega \phi(\mathbf{r}, E, \Omega, t) d\Omega \quad \text{and} \quad j_n^- = \int_{\mathbf{n} \cdot \Omega < 0} |\mathbf{n} \cdot \Omega| \phi(\mathbf{r}, E, \Omega, t) d\Omega. \quad (9.20)$$

The current vector in particle transport is analogous to the heat flux in heat transfer, both being vectors that quantify flow (of particles in the first case and heat energy in the second) per unit area per unit time. Further, the direction of the vector at a point in space and time identifies the direction of the *net* flow there and then.

9.2 Radiation Interactions with the Medium

The quantification of how radiation interacts and is affected by the host medium depends not only on the strength of the radiation field but also on the propensity of the particles to interact with the atoms of the medium. The manner in which radiation interacts with the medium depends on the type of particle (neutral or charged) and the type of atoms composing the medium. For example, electrons interact simultaneously with the electrons of all the nearby ambient atoms and, consequently, their paths through a medium are tortuous, exhibiting countless small angle deflects. Heavy ions plow through the electron clouds of the ambient atoms in nearly straight-line paths leaving in their wake thousands of secondary free electrons and ions. By contrast, neutral particles such as photons and neutrons stream in straight lines through matter and pass countless atoms before undergoing some sort of interaction.

To cover transport problems for all these types of radiation is well beyond the scope of this text. To demonstrate, however, how radiation interacts with the host medium, the discussion below is limited to neutral subatomic particles such as neutrons and photons. Unlike charged particles, neutral particles do not interact with the ambient charged electrons and nuclei through the long-range Coulombic force and, hence, they travel in straight-line trajectories punctuated by “point” interactions with a single electron or nucleus.

9.2.1 Interaction Coefficient Macroscopic Cross Section

The interaction of radiation with the host medium is a statistical (random or stochastic) process. There is no way to predict how a single radiation particle will interact or how far it will travel before interacting. However, the average interaction behavior of a large number of particles can be estimated.

Consider the situation of Fig. 9.4 in which a large number N of neutral particles are incident on a thin slab of the host medium with thickness Δx . While traversing the slab of thickness Δx , a small number ΔN interact, while most particles simply stream through without interaction. The probability that any one of the incident particles interacts in the slab is $\Delta N/N$. If one were to measure $\Delta N/N$ for different thicknesses Δx and plot the ratio, a plot similar to that of Fig. 9.5 would be obtained. The probability of interaction per unit distance of travel, $(\Delta N/N)/\Delta x$, would appear to vary smoothly for large Δx . But as Δx becomes smaller and smaller, the statistical fluctuations in the number of interactions in the slab would become apparent and the measured interaction probability per unit distance of travel would have larger and larger statistical fluctuations.

If the experiment were repeated a large number of times, and the results of each experiment were averaged, the statistical fluctuations at small values of Δx would decrease and approach the dashed line in Fig. 9.5. This dashed line could also be obtained by extrapolating the interaction probability per unit travel distance from those values without large statistical fluctuation obtained at large Δx . The statistically averaged interaction probability per unit travel distance, in the limit of infinitely small Δx ,

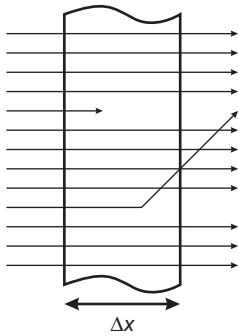


Figure 9.4 N particles incident on a thin slab in which ΔN interact while traversing the slab.

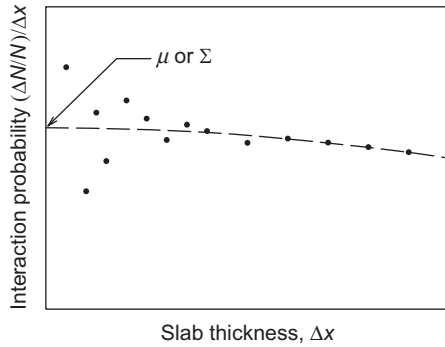


Figure 9.5 Measured interaction probabilities $\Delta N/N$ divided by the slab thickness.

approaches a constant μ , i.e., μ is defined as

$$\mu \equiv \lim_{\Delta x \rightarrow 0} \frac{(\Delta N/N)}{\Delta x}. \quad (9.21)$$

The *linear interaction coefficient* μ is a property of the host medium for a specified incident particle with a specified energy E . The interaction coefficient can be specialized to specific types of interaction, designated by a subscript on μ , i.e., μ_i , where, for example, $i = s$ for a scatter, $i = a$ for an absorption, and so forth. Thus, in the limit of small path lengths, μ_i is the probability, per unit differential path length, that a particle undergoes an i th type of interaction.

That μ_i is constant for a given material and for a given type of interaction implies that the probability of interaction, per unit differential path length, is independent of the path length traveled prior to the interaction. In this book, when the interaction coefficient is referred to as the *probability per unit path length* of an interaction, it is understood that this is true only in the limit of very small path lengths.

The constant μ_i is called the *linear interaction coefficient* for reaction i . For each type of reaction, there is a corresponding linear coefficient. For example, μ_a is the *linear absorption coefficient*, μ_s the *linear scattering coefficient*, and so on. Although this nomenclature is widely used to describe photon interactions, μ_i is often referred to as the *macroscopic cross section* for reactions of type i and is usually given the symbol Σ_i when describing neutron interactions. However, for unity of presentation, the symbol μ_i is used here for all types of radiation.

The probability, per unit path length, that a neutral particle undergoes some sort of reaction, μ_t , is the sum of the probabilities, per unit path length of travel, for each type of possible reaction, i.e.,

$$\mu_t(E) = \sum_i \mu_i(E). \quad (9.22)$$

Often, the subscript t is dropped and the symbol μ is used to represent the total interaction coefficient. Since these coefficients generally depend on the particle's kinetic energy E , this dependence has been shown explicitly. The total interaction probability per unit path length, μ_t , is fundamental in describing how indirectly ionizing radiation interacts with matter and is usually called the *linear attenuation coefficient*. It is perhaps more appropriate to use the words *total linear interaction coefficient* because many interactions do not “attenuate” the particle in the sense of an absorption interaction.

9.2.2 Attenuation of Uncollided Radiation

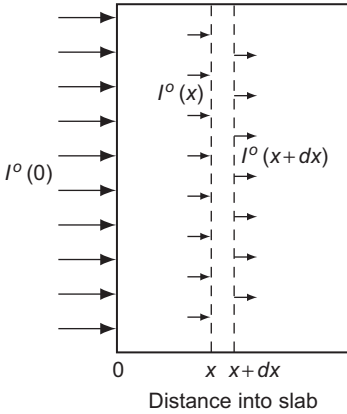


Figure 9.6 Uniform illumination of a slab by radiation.

To see how radiation is attenuated as it travels through matter, consider a plane parallel beam of neutral particles of intensity $I^o \text{ cm}^{-2} \text{ s}^{-1}$ incident normally on the surface of a slab (see Fig. 9.6). As the particles travel into the slab some interact with the atoms of the slab, thereby reducing the beam's intensity. In particular, the intensity $I^o(x)$ of *uncollided* particles at depth x into the slab is of interest. At distance x into the slab, some uncollided particles undergo interactions for the first time as they try to traverse the next Δx of distance, thereby reducing the uncollided beam intensity at x , $I^o(x)$, to a lesser value $I^o(x + \Delta x)$ at $x + \Delta x$. The probability an uncollided particle interacts as it traverses Δx is

$$P(\Delta x) = \frac{I^o(x) - I^o(x + \Delta x)}{I^o(x)}.$$

In the limit $\Delta x \rightarrow 0$, we have from Eq. (9.21)

$$\mu_t = \lim_{\Delta x \rightarrow 0} \frac{P(\Delta x)}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{I^o(x) - I^o(x + \Delta x)}{\Delta x} \frac{1}{I^o(x)} \equiv -\frac{dI^o(x)}{dx} \frac{1}{I^o(x)}$$

or

$$\frac{dI^o(x)}{dx} = -\mu_t I^o(x). \quad (9.23)$$

The solution for the uncollided intensity is

$$I^o(x) = I^o(0)e^{-\mu_t x}. \quad (9.24)$$

From this result, it is seen that uncollided neutral-particle radiation is *exponentially* attenuated as it passes through a medium.

Also from this result, the probability $P(x)$ a particle interacts somewhere along a path of length x is

$$P(x) = 1 - \frac{I^o(x)}{I^o(0)} = 1 - e^{-\mu_t x}. \quad (9.25)$$

Conversely, the probability $\bar{P}(x)$ a particle does *not* interact while traveling a distance x is

$$\bar{P}(x) = 1 - P(x) = e^{-\mu_t x}. \quad (9.26)$$

Notice that as $x \rightarrow dx$, $P(dx) \rightarrow \mu_t dx$, a result in agreement with the interpretation that μ_t is the probability, per unit differential path length of travel, the particle interacts.

9.2.3 Average Travel Distance Before an Interaction

Although where a particle interacts cannot be predicted, the probability distribution for how far a neutral particle travels before interacting can be predicted from the previous results. Let $p(x)dx$ be the probability that a particle interacts for the first time between x and $x + dx$. Then

$$\begin{aligned} p(x)dx &= \text{Prob}\{\text{particle travels a distance } x \text{ without interaction}\} \\ &\quad \times \text{Prob}\{\text{particle interacts in the next } dx\} \\ &= \{\bar{P}(x)\} \{P(dx)\} = \{e^{-\mu_t x}\} \{\mu_t dx\} = \mu_t e^{-\mu_t x} dx. \end{aligned} \quad (9.27)$$

Note that $\int_0^\infty p(x) dx = 1$, as it must be for a probability distribution function.

With this probability distribution the average distance $\bar{x}$ traveled by a neutral particle to the site of its first interaction can be found. The average value of x is simply

$$\bar{x} = \int_0^\infty x p(x) dx = \mu_t \int_0^\infty x e^{-\mu_t x} dx = \frac{1}{\mu_t}. \quad (9.28)$$

This average travel distance before an interaction, $1/\mu_t$, is called the *mean-free-path length*.

9.2.4 Scattering Interaction Coefficients

A radiation field generally has two components, namely, the uncollided particles, which are attenuated exponentially as they pass through the host medium, and secondary particles that usually are a result of scattering interactions.² To describe the scattering of radiation the scattering interaction coefficient must be generalized.

² Other reactions producing secondary particles include stimulated photon emission and neutron-induced fission.

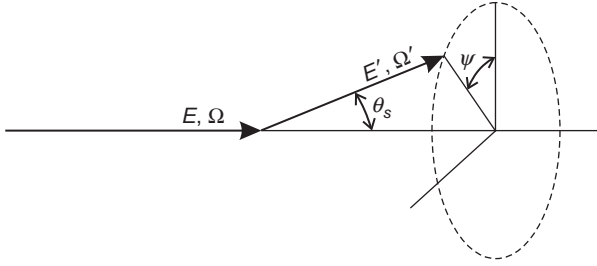


Figure 9.7 A particle moving in direction Ω scatters into a new direction Ω' . The probability of such a scatter generally depends on the scattering angle $\theta_s = \cos^{-1} \Omega \cdot \Omega'$, but in an isotropic medium all azimuthal angles about the incident direction are equally probable.

In a scattering interaction, a particle with energy E and direction Ω scatters into a new direction Ω' and has a new energy E' . The *doubly differential scattering coefficient* is defined such that $\mu_s(E \rightarrow E', \Omega \rightarrow \Omega') dE' d\Omega'$ is the probability, per unit differential path length of travel, of a scattering interaction in which the incident particle of energy E and direction Ω emerges from the interaction with an energy in dE' about E' and with a direction in $d\Omega'$ about Ω' . In an isotropic medium the probability of interaction at some point is independent of the particle direction and, for the scattering event shown in Fig. 9.7, the scattering probability depends on only the initial and final energies and the scattering angle $\theta_s = \cos^{-1} \Omega \cdot \Omega'$, or, equivalently, on $\omega_s \equiv \cos \theta_s$.³ All changes in the azimuthal angle ψ about the incident direction are equally likely. Thus, the doubly differential scattering coefficient is usually written as $\mu_s(E \rightarrow E', \omega_s)$ or $\mu_s(E \rightarrow E', \Omega \cdot \Omega')$. In this form, $\mu_s(E \rightarrow E', \omega_s)$ has units such as $\text{cm}^{-1} \text{MeV}^{-1} \text{sr}^{-1}$.

In any scattering interaction, the constraints of energy and linear momentum conservation require that, if any two of the variables E , E' , and ω_s are specified, the third must assume a specific value. Thus, the scattering angle is a function of the initial and final energies, or the final energy is a function of the initial energy and scattering angle. For example, when a neutron scatters elastically from a nucleus with atomic mass A the cosine of the scattering angle is given by [Shultis and Faw 2008]

$$\omega_s = S(E, E') \equiv \frac{1}{2} \left[(A + 1) \sqrt{\frac{E}{E'}} - (A - 1) \sqrt{\frac{E'}{E}} \right].$$

Other types of photon and neutrons scattering have different expressions for $S(E, E')$ as derived by Shultis and Faw [2000].

³ This dependence is true for isotropic media and is assumed to be the case unless otherwise noted specifically. For crystalline and other anisotropic media, μ_s generally depends on the incident radiation direction Ω and the exit radiation direction Ω' .

No matter how a neutron or photon scatters, the differential scattering cross section, thus, has the form

$$\mu_s(E \rightarrow E', \omega_s) = \mu_s(E \rightarrow E')\delta(\omega_s - S(E, E')). \quad (9.29)$$

Integration over E' or ω_s eliminates the delta function and produces a *singly differential scattering coefficient*. Often, it is of interest to deal only with, say, the energy dependence or the angular distribution of scattered radiation. In this case, one or the other of the following forms of the singly differential scattering coefficient may be used:

$$\mu_s(E, E') \equiv \int_{4\pi} d\Omega \mu_s(E \rightarrow E', \omega_s) \quad (9.30)$$

or

$$\mu_s(E, \theta_s) \equiv \int_0^\infty dE' \mu_s(E \rightarrow E', \omega_s). \quad (9.31)$$

Here, $\mu_s(E, E') dE'$ is the probability per unit path length for scattering into dE' about E' without regard to scattering angle, and $\mu_s(E, \omega_s) d\Omega$ is the like probability for scattering into direction range $d\Omega$ without regard to the energy of the scattered radiation.

Also of interest is

$$\mu_s(E) \equiv \int_0^\infty dE' \mu_s(E \rightarrow E'), \quad (9.32)$$

which is just the total linear interaction coefficient for the scattering of incident radiation of energy E without regard to energy loss or angle of scattering.

9.2.5 Microscopic Cross Sections

The linear coefficient $\mu_i(E)$ depends on the type and energy E of the incident particle, the type of interaction i , and the interacting medium. An important property of the medium that determines μ_i is the density of target atoms or electrons. It is reasonable to expect that μ_i is proportional to the “target” atom density N in the material, i.e.,

$$\mu_i = \sigma_i N = \sigma_i \frac{\rho N_a}{A}, \quad (9.33)$$

where σ_i is a constant of proportionality independent of N . Here, ρ is the mass density of the medium, N_a is Avogadro’s constant (6.022×10^{23} particles mol^{-1}), and A is the atomic mass of the medium.

The proportionality constant σ_i is called the *microscopic cross section* for reaction i . From Eq. (9.33) it is seen to have dimensions of area. Often, σ_i is interpreted as the *effective* cross-sectional area presented by the target atom to the incident particle for a given type of interaction. Indeed, many times σ_i has dimensions comparable to those

expected from the physical size of the nucleus or atom. However, this widely used interpretation of the microscopic cross section, while easy to grasp, leads to philosophical difficulties when it is observed that σ_i generally varies with the energy of the incident particle and, for a crystalline material, the particle direction. An alternative interpretation that σ_i is the interaction probability per unit differential path length of travel, normalized to one target atom per unit volume, avoids such conceptual difficulties while emphasizing the statistical nature of the interaction process. Microscopic cross sections are usually expressed in units of *barns*, where the barn equals 10^{-24} cm^2 .

Data on cross sections and linear interaction coefficients, especially for photons, are frequently expressed as the ratio of μ_i to the medium mass density ρ . This ratio is called the *mass interaction coefficient* for reaction i . Division of Eq. (9.33) by ρ yields

$$\frac{\mu_i}{\rho} = \frac{\sigma_i N}{\rho} = \frac{N_a}{A} \sigma_i. \quad (9.34)$$

From this result, it is seen that μ_i/ρ is an intrinsic property of the interacting medium—independent of its mass density. For photons μ_i/ρ , data are more frequently tabulated, while for neutrons σ_i , values are the preferred tabulated quantities.

The linear and mass interaction coefficients in compounds or homogeneous mixtures for interactions of type i are, respectively,

$$\mu_i = \sum_j \mu_i^j = \sum_j N^j \sigma_i^j \quad (9.35)$$

and

$$\frac{\mu_i}{\rho} = \sum_j w_j \left(\frac{\mu_i}{\rho} \right)^j. \quad (9.36)$$

Here, the subscript i refers to the type of interaction, the superscript j to the j th component of the material, and w_j is the mass fraction of component j . In Eq. (9.35), the atomic density N^j and the linear interaction coefficient μ_i^j are values for the j th material *after* mixing.

The total mass interaction coefficient μ/ρ for photons in all the elements is shown in Fig. 9.8. A typical neutron microscopic cross section is shown in Fig. 9.9. Unlike photon cross sections, which are generally smooth and slowly varying above 100 keV, neutron cross sections usually display many resonances. Because photons interact only with atomic electrons, all isotopes of the same element have the same photon cross sections. Moreover, photon cross sections vary slowly as the atomic number of the element changes. By contrast, neutron cross sections vary dramatically from isotope to isotope and from element to element. Consequently, compilations of the empirical cross sections for neutron interactions require huge databases compared to those needed for photon cross sections.

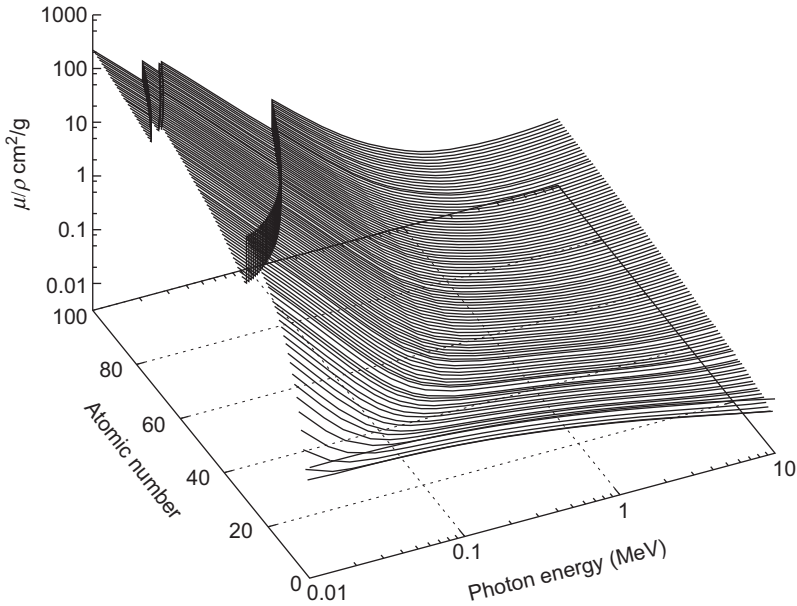


Figure 9.8 Comparison of the total mass interaction coefficients, less coherent scattering, for all the elements. The coefficients generally vary smoothly except for the discontinuities at the *electron shell edges* below which a photon has insufficient energy to remove an electron from the shell in a photoelectric interaction that dominates at low photon energies. At high energies, pair production interactions dominate, and at intermediate energies photon scattering dominates.

9.2.6 Reaction Rate Density

One of the most fundamental goals of transport theory is to compute the rate at which radiation particles interact with the host medium at some point of interest. For example, the rate of biological damage to some organ of the body is proportional to the rate at which radiation interacts in the organ. Similarly, the effectiveness of a radiation shield depends on the number of interactions that occur in the shield as the radiation migrates through it.

The rate at which radiation interacts with the medium is readily expressed in terms of the flux density and the interaction coefficient. Recall that the flux density spectrum $\phi(\mathbf{r}, E, t)$ has the interpretation that $\phi(\mathbf{r}, E, t) dE$ is the total path length traveled, in a unit time at t , by all the particles that are in a unit volume at $\mathbf{r}$ and that have energies in dE about E . Also, recall that the average distance a radiation particle must travel before undergoing an i th-type interaction with the medium is $1/\mu_i(\mathbf{r}, E, t)$, where the interaction coefficient generally varies with position, the incident particle energy, and even time if the medium's composition changes. Then the expected number of i th-type interactions occurring in a unit volume at $\mathbf{r}$, in a unit time at time t , caused by those particles with energies in dE about E , is

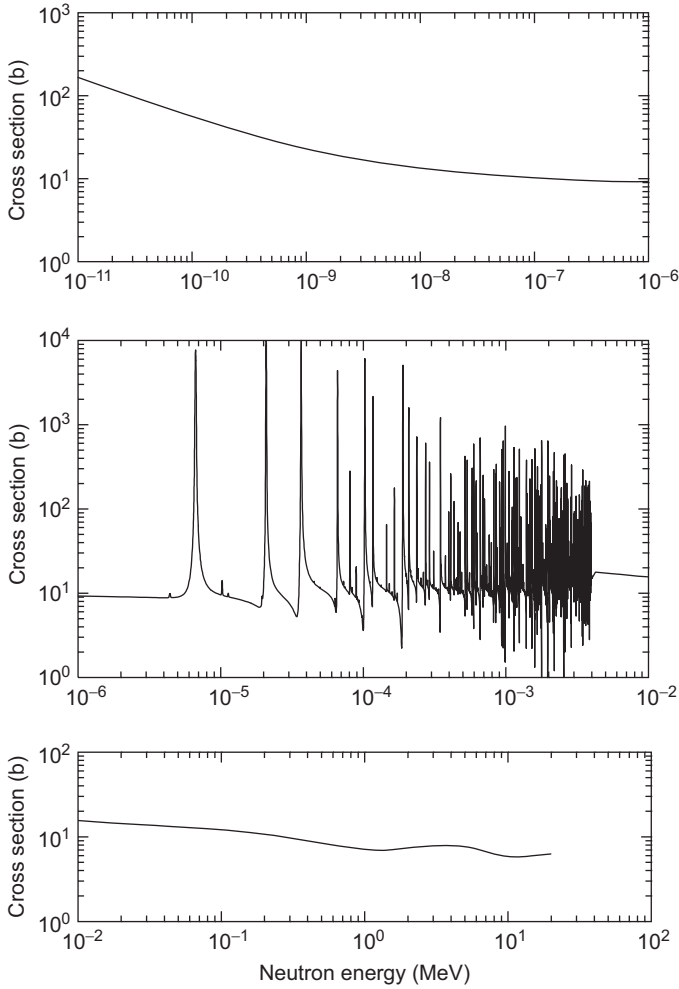


Figure 9.9 Total neutron cross section for uranium computed using NJOY-processed ENDF/B (version V) data. Above 4 keV, the resonances are no longer resolved and only the average cross-section behavior is shown.

$$\begin{aligned}
 R_i(\mathbf{r}, E, t)dE &= \{\text{path length traveled by particles in } 1 \text{ cm}^3 \text{ in one second}\} / \\
 &\quad \{\text{average travel distance for an } i\text{th-type interaction}\} \\
 &= \{\phi(\mathbf{r}, E, t)dE\} / \{1/\mu_i(\mathbf{r}, E, t)\}
 \end{aligned}$$

or

$$R_i(\mathbf{r}, E, t) = \mu_i(\mathbf{r}, E, t) \phi(\mathbf{r}, E, t). \quad (9.37)$$

This simple expression for reaction-rate densities is a key result for many radiation transport calculations.

From the interaction-rate density, all sorts of useful information can be calculated about the effects of the radiation. For example, the total number of fissions that occur in some volume V inside a reactor core between times t_1 and t_2 by neutrons of all energies is

$$\text{Number of fissions in } (t_1, t_2) = \int_{t_1}^{t_2} dt \iiint_V dV \int_0^{E_{\max}} dE \mu_f(\mathbf{r}, E, t) \phi(\mathbf{r}, E, t),$$

where μ_f is the macroscopic fission cross section and $E_{\max}$ is the maximum energy a neutron can have.

9.3 Transport Equation

One would ideally like to know the distribution of photons or neutrons everywhere in the medium of interest. For most applications the spatial and energy distribution of the particle density $n(\mathbf{r}, E, t)$ is all that is needed. Unfortunately, there is no equation for this quantity that holds rigorously in all situations. The simplest equation that accurately describes the particle distribution in a medium is for the differential energy and directional flux density $\phi(\mathbf{r}, E, \boldsymbol{\Omega}, t)$ (see [Section 9.1.3](#)).

To obtain an equation for $\phi(\mathbf{r}, E, \boldsymbol{\Omega}, t)$, consider the particle balance in an arbitrary volume V with a closed surface $\mathcal{S}$ for those particles with energies in dE about E moving in $d\boldsymbol{\Omega}$ about the direction $\boldsymbol{\Omega}$ at time t . In a radiation field, the rate of change of the number of particles in V that have energies in dE about E and that travel in directions $d\boldsymbol{\Omega}$ about $\boldsymbol{\Omega}$ at time t is the difference between (1) the creation of particles with these energies and directions (from sources and particles that scatter into these energy and direction intervals) and (2) the loss of these particles (either by leaving or entering the volume V or by changing their energy or direction of travel). Thus, the particles of the radiation field with these energies and directions must satisfy the following balance relation:

$$\left\{ \begin{array}{l} \text{(a) the rate of} \\ \text{increase of} \\ \text{particles in } V \end{array} \right\} = - \left\{ \begin{array}{l} \text{(b) the net flow} \\ \text{rate out of } V \\ \text{across } \mathcal{S} \end{array} \right\} - \left\{ \begin{array}{l} \text{(c) rate at which} \\ \text{particles interact} \\ \text{in } V \end{array} \right\} \\ + \left\{ \begin{array}{l} \text{(d) rate at which} \\ \text{secondary particles} \\ \text{of energy } E \\ \text{and direction } \boldsymbol{\Omega} \\ \text{are produced} \end{array} \right\} + \left\{ \begin{array}{l} \text{(e) rate of} \\ \text{production} \\ \text{by sources} \\ \text{in } V \end{array} \right\}. \quad (9.38)$$

The five terms have the following interpretations: (a) the rate of increase of the number of particles in V with the specified energies and directions, (b) the net number of such particles flowing out of V across $\mathcal{S}$ in a unit time, (c) the total number of these particles

suffering collisions in V in a unit time (i.e., those changing their energy and/or direction of travel or simply being absorbed), (d) the number of secondary particles in dE about E moving in $d\Omega$ about Ω that are produced by all particle-medium interactions in V in a unit time, and (e) the number of particles in dE about E moving in $d\Omega$ about Ω that are introduced in V in a unit time by sources that are independent of the radiation field—for example, gamma photons from radionuclides. The desired transport equation for $\phi(\mathbf{r}, E, \Omega, t)$ is then obtained by expressing mathematically each component of this balance relation in terms of the angular flux density $\phi(\mathbf{r}, E, \Omega, t)$. To obtain the transport equation for $\phi(\mathbf{r}, E, \Omega, t)$ each of the five terms in the balance relation must be expressed mathematically.

Term (a): The total number of particles in V with energies in dE about E and with directions in $d\Omega$ about Ω is

$$\left[\int_V dV n(\mathbf{r}, E, \Omega, t) \right] dE d\Omega = \frac{1}{v} \left[\int_V dV \phi(\mathbf{r}, E, \Omega, t) \right] dE d\Omega. \quad (9.39)$$

Thus, the rate of change of this quantity is

$$(a) = \frac{1}{v} \frac{\partial}{\partial t} \left[\int_V dV \phi(\mathbf{r}, E, \Omega, t) \right] dE d\Omega = \left[\int_V dV \frac{1}{v} \frac{\partial \phi(\mathbf{r}, E, \Omega, t)}{\partial t} \right] dE d\Omega. \quad (9.40)$$

Term (b): As discussed in Section 9.1.5, the quantity $\mathbf{n} \cdot \Omega \phi(\mathbf{r}, E, \Omega) dE d\Omega$ is the number of particles with energy in dE about E traveling in $d\Omega$ about Ω that cross a unit area of the surface (with unit outward normal $\mathbf{n}$) in a unit time at point $\mathbf{r}$. Thus, term (b) of the balance relation may be expressed mathematically as

$$(b) = \left[\int_S dS \mathbf{n} \cdot \Omega \phi(\mathbf{r}, E, \Omega, t) \right] dE d\Omega, \quad (9.41)$$

which can be expressed as a volume integral upon application of the Gauss integral theorem, namely,

$$(b) = \left[\int_V dV \nabla \cdot \Omega \phi(\mathbf{r}, E, \Omega, t) \right] dE d\Omega. \quad (9.42)$$

Because the ∇ operator affects only the spatial variable and not Ω since $\mathbf{r}$ and Ω are independent variables, this last result may be written as

$$(b) = \left[\int_V dV \Omega \cdot \nabla \phi(\mathbf{r}, E, \Omega, t) \right] dE d\Omega. \quad (9.43)$$

Term (c): The total interaction rate, term (c) of Eq. (9.38), is

$$(c) = \left[\int_V dV \mu(\mathbf{r}, E, t) \phi(\mathbf{r}, E, \Omega, t) \right] dE d\Omega, \quad (9.44)$$

where μ is the total interaction coefficient or macroscopic cross section. Any interaction in V either changes the energy and direction of the particle or removes it.

Term (d): If it is assumed that the secondary particles appear at the position where the causal interaction occurs, term (d) can be written as

$$(d) = \left[\int_V dV \int_0^\infty dE' \int_{4\pi} d\Omega' \mu(\mathbf{r}, E' \rightarrow E, \Omega' \rightarrow \Omega, t) \phi(\mathbf{r}, E', \Omega', t) \right] dE d\Omega, \quad (9.45)$$

where $\mu(\mathbf{r}, E' \rightarrow E, \Omega' \rightarrow \Omega, t) dE d\Omega$ is the probable number of secondary particles at point $\mathbf{r}$ with energies in dE about E in direction $d\Omega$ about Ω produced by an incident particle of energy E' traveling in direction Ω' , per unit differential path length of the incident particle. For example, Eq. (9.29) gives this secondary particle distribution for elastic scattering of neutrons.

Term (e): Finally, let $S(\mathbf{r}, E, \Omega, t) dE d\Omega$ denote the production rate by nonradiation-induced sources of particles, per unit volume about point $\mathbf{r}$, with energies in dE about E and in directions $d\Omega$ about Ω . Thus,

$$(e) = \left[\int_V dV S(\mathbf{r}, E, \Omega, t) \right] dE d\Omega. \quad (9.46)$$

With these results, the balance relation Eq. (9.38) can be written as

$$\int_V \left[\frac{1}{v} \frac{\partial \phi}{\partial t} - \Omega \cdot \nabla \phi - \mu \phi + \int_0^\infty dE' \int_{4\pi} d\Omega' \mu(\mathbf{r}, E' \rightarrow E, \Omega' \rightarrow \Omega, t) \phi(\mathbf{r}, E', \Omega', t) + S(\mathbf{r}, E, \Omega, t) \right] dV = 0. \quad (9.47)$$

However, the volume V is arbitrary, and hence the integrand of Eq. (9.47) must be identically zero; therefore,

$$\begin{aligned} \frac{1}{v} \frac{\partial \phi(\mathbf{r}, E, \Omega, t)}{\partial t} &= -\Omega \cdot \nabla \phi(\mathbf{r}, E, \Omega, t) - \mu(\mathbf{r}, E, t) \phi(\mathbf{r}, E, \Omega, t) \\ &+ \int_0^\infty dE' \int_{4\pi} d\Omega' \mu(\mathbf{r}, E' \rightarrow E, \Omega' \rightarrow \Omega, t) \phi(\mathbf{r}, E', \Omega', t) + S(\mathbf{r}, E, \Omega, t). \end{aligned} \quad (9.48)$$

This integro-differential equation for $\phi(\mathbf{r}, E, \Omega)$ is known as the time-dependent *transport equation* or the *linearized Boltzmann equation*, named after Ludwig Boltzmann, who first derived it over a century ago. This equation serves as a precise description of the radiation field in all neutron or photon transport problems. However, in the derivation above, several implicit assumptions have been made. The radiation particles are assumed to be point particles that stream in straight lines between collisions, which occur at distinct points in the system. In other words, the associated wavelength of the particle and the interaction distance between the particle and the atoms of the medium are assumed to be very small compared to the mean-freepath length of the particle and the size of the system. Also, all internal structure, such as spin or polarization of the

diffusing particles, is ignored. More important, the transport equation is linear because all particle–particle interactions are neglected.

The $1/v$ factor in the first term of Eq. (9.48) makes this term generally negligible, unless $\partial\phi/\partial t$ becomes exceptionally large. Such rapid transients occur rarely, such as during the detonation of a nuclear bomb or a supernova. However, in most cases, transients are much slower and the time dependence of the radiation field is that of the primary radiation sources, i.e., any time delays caused by the finite speed of the radiation particles can be ignored in most transport problems. Also, in most transport problems, the host medium does not change in time, nor does the primary source of radiation vary with time. Under these conditions in which the first term in Eq. (9.48) can be ignored and the interaction coefficients do not change in time, this equation reduces to the steady-state transport equation

$$\begin{aligned} & \boldsymbol{\Omega} \cdot \nabla \phi(\mathbf{r}, E, \boldsymbol{\Omega}) + \mu(\mathbf{r}, E) \phi(\mathbf{r}, E, \boldsymbol{\Omega}) \\ &= \int_0^\infty dE' \int_{4\pi} d\boldsymbol{\Omega}' \mu(\mathbf{r}, E' \rightarrow E, \boldsymbol{\Omega}' \rightarrow \boldsymbol{\Omega}) \phi(\mathbf{r}, E', \boldsymbol{\Omega}') + S(\mathbf{r}, E, \boldsymbol{\Omega}). \end{aligned}$$

(9.49)

This equation, or a variant thereof, is the basis of many of the large transport codes that are routinely used today (see Appendix D).

The form of the transport equation given by Eq. (9.49) is quite general and applies to any geometry, to either photons or neutrons, and allows for all types of particle-medium interactions, all of the probabilities of which are encompassed by the interaction coefficients $\mu(\mathbf{r}, E)$ and $\mu(\mathbf{r}, E' \rightarrow E, \boldsymbol{\Omega}' \rightarrow \boldsymbol{\Omega})$. To solve this equation for the differential energy and angular flux density $\phi(\mathbf{r}, E, \boldsymbol{\Omega})$, one must specify (1) the coordinate system used, so that the streaming term $\boldsymbol{\Omega} \cdot \nabla \phi$ can be expressed explicitly, (2) the coefficients $\mu(\mathbf{r}, E)$ and $\mu(\mathbf{r}, E' \rightarrow E, \boldsymbol{\Omega}' \rightarrow \boldsymbol{\Omega})$, which depend on the material composition of the host medium and the interactions of interest, (3) the distribution $S(\mathbf{r}, E, \boldsymbol{\Omega})$ of the primary radiation sources, which usually arises from the presence of radionuclides in the medium, and (4) the boundary conditions, which determine the incident flux density at the edges of the host medium.

Unfortunately, exact analytical solutions of the transport equation are known for only the simplest cases, none of which are directly applicable to realistic transport problems. Consequently, much effort has been directed toward the development of approximations to the transport equation, which can then be solved by numerical techniques or, for the simplest approximations, even by analytical methods. These numerical methods for solving the transport equation were important in the first few decades following World War II because they were generally much more computationally efficient than Monte Carlo simulations of the transport process. Today, because of the widespread availability of inexpensive fast computers, Monte Carlo approaches have largely supplanted the numerical approximation methods. Readers interested in numerical approximations of the transport equation are referred to books and review articles devoted to transport theory [e.g., Duderstadt and Martin 1979; Sanchez and McCormick 1982; Lewis and Miller 1984; Adams and Larsen 2002].

9.3.1 Integral Forms of the Transport Equation

The integro-differential form of the transport equation, Eq. (9.49), can be converted into a pure integral equation. Because this alternate integral form of the transport equation is equivalent to the integro-differential form, exact solutions to it are also not available, except for the simplest of cases. The utility of the integral form is that analytical and numerical techniques can be developed from it for obtaining approximate solutions for realistic problems. In particular, approximations to the integral transport equation are capable of more accurately treating highly anisotropic angular flux densities and scattering cross sections than are approximations based on the integro-differential transport equation. Also, the integral form of the transport equation can be used directly by Monte Carlo methods using the techniques discussed in the next chapter.

Integral equation for the angular flux density

To derive an integral equation for the angular flux density spectrum $\phi(\mathbf{r}, E, \boldsymbol{\Omega})$, begin by writing Eq. (9.49) in the form

$$\boldsymbol{\Omega} \cdot \nabla \phi(\mathbf{r}, E, \boldsymbol{\Omega}) + \mu(\mathbf{r}, E)\phi(\mathbf{r}, E, \boldsymbol{\Omega}) = \chi(\mathbf{r}, E, \boldsymbol{\Omega}), \quad \mathbf{r} \in V, \quad (9.50)$$

where

$$\chi(\mathbf{r}, E, \boldsymbol{\Omega}) \equiv \int_0^\infty dE' \int_{4\pi} d\Omega' \mu(\mathbf{r}, E' \rightarrow E, \boldsymbol{\Omega}' \rightarrow \boldsymbol{\Omega}) \phi(\mathbf{r}, E', \boldsymbol{\Omega}') + S(\mathbf{r}, E, \boldsymbol{\Omega}). \quad (9.51)$$

The quantity $\chi(\mathbf{r}, E, \boldsymbol{\Omega})$ is the *emissions-rate density*, i.e., the rate density of particles leaving a source or a collision with coordinates $(\mathbf{r}, E, \boldsymbol{\Omega})$.

It is assumed that V is a convex surface and that no particles are incident on it. Thus, the boundary condition is⁴

$$\phi(\mathbf{r}_s, E, \boldsymbol{\Omega}) = 0, \quad \text{for } \mathbf{n} \cdot \boldsymbol{\Omega} < 0, \quad (9.52)$$

where $\mathbf{n}$ is the outward normal on the surface at $\mathbf{r}_s$.

Suppose that distances s and t are measured along the directions of $\boldsymbol{\Omega}$ and $\nabla \phi$, as in Fig. 9.10, and that Θ is the angle between these two vectors. Then, because $\nabla \phi = d\phi/dt$, $dt/ds = \cos \Theta$, and $\boldsymbol{\Omega}$ is a unit vector, it follows that

$$\boldsymbol{\Omega} \cdot \nabla \phi = \cos \Theta \frac{d\phi}{dt} = \cos \Theta \frac{ds}{dt} \frac{d\phi}{ds} = \frac{d\phi}{ds}. \quad (9.53)$$

Thus, the streaming term $\boldsymbol{\Omega} \cdot \nabla \phi$ is seen to be the rate of change of ϕ along the direction of particle travel $\boldsymbol{\Omega}$, namely, $d\phi/ds$, where s is measured along the direction of $\boldsymbol{\Omega}$.

⁴ Concavities can always be filled with void regions to produce a volume V that is convex. Moreover, if radiation is incident from outside of V , the source S can be modified to include the first collision distribution of such incident radiation and the homogeneous boundary condition then applies.

Now define $\mathbf{r} \equiv \mathbf{r}' - R\boldsymbol{\Omega}$, where $R = -s$ is measured along the direction opposite to $\boldsymbol{\Omega}$ (see Fig. 9.11), then $d/ds = -d/dR$ and Eq. (9.50) can be rewritten as

$$-\frac{d}{dR}\phi(\mathbf{r}' - R\boldsymbol{\Omega}, E, \boldsymbol{\Omega}) + \mu(\mathbf{r}' - R\boldsymbol{\Omega}, E)\phi(\mathbf{r}' - R\boldsymbol{\Omega}, E, \boldsymbol{\Omega}) = \chi(\mathbf{r}' - R\boldsymbol{\Omega}, E, \boldsymbol{\Omega}). \quad (9.54)$$

Now multiply this result by the integrating factor $\exp[-\int_0^R \mu(\mathbf{r}' - R'\boldsymbol{\Omega}, E) dR']$, which has the property

$$\begin{aligned} \frac{d}{dR} \exp\left[-\int_0^R \mu(\mathbf{r}' - R'\boldsymbol{\Omega}, E) dR'\right] \\ = -\mu(\mathbf{r}' - R\boldsymbol{\Omega}, E) \exp\left[-\int_0^R \mu(\mathbf{r}' - R'\boldsymbol{\Omega}, E) dR'\right], \end{aligned} \quad (9.55)$$

to obtain

$$\begin{aligned} -\frac{d}{dR} \left\{ \phi(\mathbf{r}' - R\boldsymbol{\Omega}, E, \boldsymbol{\Omega}) \exp\left[-\int_0^R \mu(\mathbf{r}' - R'\boldsymbol{\Omega}, E) dR'\right] \right\} \\ = \chi(\mathbf{r}' - R\boldsymbol{\Omega}, E, \boldsymbol{\Omega}) \exp\left[-\int_0^R \mu(\mathbf{r}' - R'\boldsymbol{\Omega}, E) dR'\right]. \end{aligned} \quad (9.56)$$

Next, replace the dummy variable $\mathbf{r}'$ by $\mathbf{r}$ and integrate over R from zero to the distance to the boundary R_s (or to infinity if the medium is infinite) to obtain the following

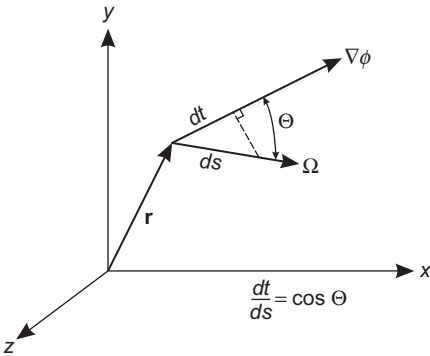


Figure 9.10 Orientation of the unit vector $\boldsymbol{\Omega}$ and the gradient of $\phi(\mathbf{r}, E, \boldsymbol{\Omega})$.

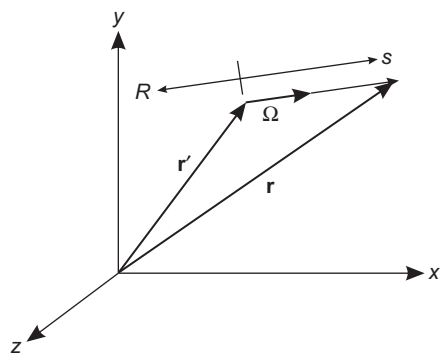


Figure 9.11 Coordinate system for integrating the transport equation.

integral equation for the angular flux density:

$$\phi(\mathbf{r}, E, \boldsymbol{\Omega}) = \int_0^{R_s} dR' \chi(\mathbf{r} - R'\boldsymbol{\Omega}, E, \boldsymbol{\Omega}) \exp\left[-\int_0^{R'} \mu(\mathbf{r} - R''\boldsymbol{\Omega}, E) dR''\right]. \quad (9.57)$$

To obtain this result, the boundary condition of Eq. (9.52) has been used, i.e., $\phi(\mathbf{r} - R_s\boldsymbol{\Omega}, E, \boldsymbol{\Omega}) \equiv \phi(\mathbf{r}_s, E, \boldsymbol{\Omega}) = 0$, so that the term in the curly brackets in Eq. (9.56) vanishes when $R = R_s$.

To simplify the notation, define the *optical thickness* as

$$\tau(E, \mathbf{r}, \mathbf{r}') \equiv \int_0^{|\mathbf{r}-\mathbf{r}'|} \mu(\mathbf{r} - R''\boldsymbol{\Omega}, E) dR'', \quad (9.58)$$

which is the number of mean-free-path lengths of material through which a particle must pass in traveling from $\mathbf{r}'$ to $\mathbf{r}$ along the direction $\boldsymbol{\Omega}$. For a homogeneous medium $\tau(E, \mathbf{r}, \mathbf{r}') = \mu(E)|\mathbf{r} - \mathbf{r}'|$. Then Eq. (9.57) can be written as

$$\phi(\mathbf{r}, E, \boldsymbol{\Omega}) = \int_0^{R_s} dR' \chi(\mathbf{r} - R'\boldsymbol{\Omega}, E, \boldsymbol{\Omega}) e^{-\tau(E, \mathbf{r}, \mathbf{r}')}. \quad (9.59)$$

This result says the flux density at some point $\mathbf{r}$ with a given direction and energy equals the production of particles from all positions backwards along the direction $-\boldsymbol{\Omega}$ to the surface times the probability these particles reach $\mathbf{r}$ without collision (the exponential term).

Eq. (9.59) can be expressed in a slightly different form by defining $\mathbf{r}' \equiv \mathbf{r} - R'\boldsymbol{\Omega}$, so that $R' = |\mathbf{r} - \mathbf{r}'|$ and $\boldsymbol{\Omega} = (\mathbf{r} - \mathbf{r}')/|\mathbf{r} - \mathbf{r}'|$. The integrals over R' can be converted into volume integrals over the volume V of the system by using the two-dimensional delta function.⁵ Thus, from Eq. (9.59),

$$\begin{aligned} \phi(\mathbf{r}, E, \boldsymbol{\Omega}) &= \int_0^{R_s} dR' \chi(\mathbf{r}', E, \boldsymbol{\Omega}) e^{-\tau(E, \mathbf{r}, \mathbf{r}')} \\ &= \int_{4\pi} d\Omega' \int_0^{R_s} dR' \chi(\mathbf{r}', E, \boldsymbol{\Omega}') e^{-\tau(E, \mathbf{r}, \mathbf{r}')} \delta(\boldsymbol{\Omega}' - \boldsymbol{\Omega}) \\ &= \int_{4\pi} d\Omega' \int_0^{R_s} dR' \chi(\mathbf{r}', E, \boldsymbol{\Omega}') e^{-\tau(E, \mathbf{r}, \mathbf{r}')} \delta\left(\boldsymbol{\Omega}' - \frac{(\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}\right). \end{aligned} \quad (9.60)$$

⁵ The two-dimensional Dirac delta function $\delta(\boldsymbol{\Omega} - \boldsymbol{\Omega}_o)$ is defined such that

$$\int_{4\pi} d\Omega' f(\boldsymbol{\Omega}') \delta(\boldsymbol{\Omega}' - \boldsymbol{\Omega}_o) = f(\boldsymbol{\Omega}_o).$$

Then multiply the integrand in the above result by $R^2/|\mathbf{r} - \mathbf{r}'|^2 = 1$ and recognize that $\int_{4\pi} \int_0^{R_s} R'^2 dR' d\Omega' = \int_V dV'$ so that the above result becomes

$$\phi(\mathbf{r}, E, \boldsymbol{\Omega}) = \int_V dV' \frac{e^{-\tau(E, \mathbf{r}, \mathbf{r}')}}{|\mathbf{r} - \mathbf{r}'|^2} \delta\left(\boldsymbol{\Omega}' - \frac{(\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}\right) \chi(\mathbf{r}', E, \boldsymbol{\Omega}'). \quad (9.61)$$

Substitution for $\chi(\mathbf{r}', E, \boldsymbol{\Omega})$ from Eq. (9.51) then gives an explicit integral form of the transport equation, namely,

$$\begin{aligned} \phi(\mathbf{r}, E, \boldsymbol{\Omega}) &= \int_V dV' \int_0^\infty dE' \int_{4\pi} d\Omega' \frac{\mu(\mathbf{r}', E' \rightarrow E, \boldsymbol{\Omega}' \rightarrow \boldsymbol{\Omega})}{|\mathbf{r} - \mathbf{r}'|^2} \exp[-\tau(E, \mathbf{r}, \mathbf{r}')] \\ &\quad \times \delta\left(\boldsymbol{\Omega}' - \frac{(\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}\right) \phi(\mathbf{r}', E', \boldsymbol{\Omega}') \\ &\quad + \int_V dV' \exp[-\tau(E, \mathbf{r}, \mathbf{r}')] \delta\left(\boldsymbol{\Omega}' - \frac{(\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}\right) \frac{S(\mathbf{r}', E, \boldsymbol{\Omega}')}{|\mathbf{r} - \mathbf{r}'|^2}, \end{aligned} \quad (9.62)$$

where $\boldsymbol{\Omega} = (\mathbf{r} - \mathbf{r}')/|\mathbf{r} - \mathbf{r}'|$ in the first term.

This integral transport equation can be written much more compactly by using $\mathbf{P}$ to denote all the phase space variables $(\mathbf{r}, E, \boldsymbol{\Omega})$. Eq. (9.61) then can be written as

$$\boxed{\phi(\mathbf{P}) = \int d\mathbf{P}' K(\mathbf{P}' \rightarrow \mathbf{P}) \phi(\mathbf{P}') + \tilde{S}(\mathbf{P})}, \quad (9.63)$$

where the kernel is given by

$$K(\mathbf{P}' \rightarrow \mathbf{P}) \equiv \frac{\mu(\mathbf{r}', E' \rightarrow E, \boldsymbol{\Omega}' \rightarrow \boldsymbol{\Omega})}{|\mathbf{r} - \mathbf{r}'|^2} \exp[-\tau(E, \mathbf{r}, \mathbf{r}')] \delta\left(\boldsymbol{\Omega} - \frac{(\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}\right), \quad (9.64)$$

and the contribution to ϕ from the source is

$$\tilde{S}(\mathbf{P}) \equiv \int_V dV' \exp[-\tau(E, \mathbf{r}, \mathbf{r}')] \delta\left(\boldsymbol{\Omega} - \frac{(\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}\right) \frac{S(\mathbf{r}', E, \boldsymbol{\Omega})}{|\mathbf{r} - \mathbf{r}'|^2}. \quad (9.65)$$

Integral equations for integrals of $\phi(\mathbf{r}, E, \boldsymbol{\Omega})$

With the integral equation of Eq. (9.63), various averages or integrated values of $\phi(\mathbf{r}, E, \boldsymbol{\Omega})$ can be estimated formally. Consider the integral transport equation in the general form of Eq. (9.63), where $\mathbf{P}$ and $\mathbf{P}'$ represent six-dimensional phase space, without time. If the integration of Eq. (9.63) were made over some subset of $\mathbf{P}$, such

as a solid angle, an equation for an expected flux density, such as the flux density spectrum, is obtained, namely,

$$\phi(\mathbf{r}, E) = \int_{4\pi} \left[\int_V dV' \int_0^\infty dE' \int_{4\pi} d\Omega' K(\mathbf{r} \rightarrow \mathbf{r}', E' \rightarrow E, \Omega' \rightarrow \Omega) \phi(\mathbf{r}', E', \Omega') + \tilde{S}(\mathbf{r}, E, \Omega) \right] d\Omega. \quad (9.66)$$

This can be decomposed into

$$\phi(\mathbf{r}, E) = \Phi(\mathbf{r}, E) + Q(\mathbf{r}, E), \quad (9.67)$$

where

$$\Phi(\mathbf{r}, E) = \int_{4\pi} d\Omega \int_V dV' \int_0^\infty dE' \int_{4\pi} d\Omega' K(\mathbf{r} \rightarrow \mathbf{r}', E' \rightarrow E, \Omega' \rightarrow \Omega) \phi(\mathbf{r}', E', \Omega') \quad (9.68)$$

and

$$Q(\mathbf{r}, E) = \int_{4\pi} \tilde{S}(\mathbf{r}, E, \Omega) d\Omega. \quad (9.69)$$

Note that Eq. (9.68) is a slightly generalized form of an eight-fold expectation, with Φ in some sense taking the place of $\langle z \rangle$ and K taking the place of a PDF, within a normalization factor. Further, Eq. (9.69) is just a definite integral, which Monte Carlo can readily evaluate. Hence, the integral form of the transport equation makes it clear that Monte Carlo is a natural method to solve radiation transport problems. One could alternatively integrate over an arbitrary volume or over an energy interval and similarly obtain equations that are in the natural form of expectations.

It, of course, is not necessary to explicitly write equations such as Eq. (9.66) in order to solve radiation transport problems, but this approach does provide a tie to the mathematical formulation provided in Chapter 2. In fact, most radiation transport simulation proceeds on physical grounds, in which a radiation particle is tracked through its history from birth to death and the results of many histories are averaged to estimate some quantity of interest. In the next chapter, it is shown how a formal solution of the integral equation can be interpreted as a simulation of particle transport. Also, the quantity of interest often is not the flux density but rather a quantity that depends on flux density such as response of a detector, energy deposited in a region, or a certain type of reaction-rate density.

Explicit form for the scalar flux density

As an example of the above integration of $\phi(\mathbf{r}, E, \Omega)$ over some subset of $\mathbf{P}$ consider the integral equation for the flux density spectrum $\phi(\mathbf{r}, E)$. Upon integration of

Eq. (9.62) over all directions, the delta functions immediately integrate to unity to give

$$\begin{aligned}\phi(\mathbf{r}, E) = & \int_V dV' \int_0^\infty dE' \int_{4\pi} d\Omega' \frac{\mu(\mathbf{r}', E' \rightarrow E, \boldsymbol{\Omega}' \rightarrow \boldsymbol{\Omega})}{|\mathbf{r} - \mathbf{r}'|^2} e^{-\tau(E, \mathbf{r}, \mathbf{r}')} \phi(\mathbf{r}', E', \boldsymbol{\Omega}') \\ & + \int_V dV' \frac{e^{-\tau(E, \mathbf{r}, \mathbf{r}')}}{|\mathbf{r} - \mathbf{r}'|^2} S(\mathbf{r}', E, \boldsymbol{\Omega}),\end{aligned}\quad (9.70)$$

where $\boldsymbol{\Omega} = (\mathbf{r} - \mathbf{r}')/|\mathbf{r} - \mathbf{r}'|$.

Note that this integral equation for the scalar flux density still depends on the angular flux density because $\chi(\mathbf{r}, E, \boldsymbol{\Omega})$ generally depends on $\phi(\mathbf{r}, E, \boldsymbol{\Omega})$ as shown in Eq. (9.51). For the special case of isotropic sources [i.e., $S(\mathbf{r}, E, \boldsymbol{\Omega}) = S_o(\mathbf{r}, E)/4\pi$] and for isotropic scattering [i.e., $\mu_s(\mathbf{r}, E' \rightarrow E, \boldsymbol{\Omega}' \rightarrow \boldsymbol{\Omega}) = \mu_{so}(\mathbf{r}, E' \rightarrow E)/4\pi$], this integral equation reduces to one involving only the scalar flux density. Under the isotropy assumptions, Eq. (9.70) reduces to

$$\phi(\mathbf{r}, E) = \int_V dV' \frac{S_o(\mathbf{r}', E) + \int_0^\infty dE' \mu_{so}(\mathbf{r}', E' \rightarrow E) \phi(\mathbf{r}', E)}{4\pi |\mathbf{r} - \mathbf{r}'|^2} e^{-\tau(E, \mathbf{r}, \mathbf{r}')}.\quad (9.71)$$

Alternate forms of the integral transport equation

Historically, different forms of the integral transport equation have been used by introducing different dependent variables. Here, two such variations are introduced. First, the *collision-rate density*, defined as the number of interactions, per unit time, experienced by particles, in a unit volume about $\mathbf{r}$, that have energies in a unit energy about E and directions in a unit solid angle about $\boldsymbol{\Omega}$, is given by

$$F(\mathbf{r}, E, \boldsymbol{\Omega}) = \mu(\mathbf{r}, E) \phi(\mathbf{r}, E, \boldsymbol{\Omega}).\quad (9.72)$$

Multiplication of Eq. (9.63) by $\mu(\mathbf{r}, E)$ gives

$$F(\mathbf{P}) = \int d\mathbf{P}' \bar{K}(\mathbf{P}' \rightarrow \mathbf{P}) F(\mathbf{P}') + \bar{S}(\mathbf{P}),\quad (9.73)$$

where

$$\bar{K}(\mathbf{P}' \rightarrow \mathbf{P}) = \frac{\mu(\mathbf{r}, E)}{\mu(\mathbf{r}', E')} K(\mathbf{P}', \mathbf{P}) \quad \text{and} \quad \bar{S}(\mathbf{P}) = \mu(\mathbf{r}, E) \tilde{S}(\mathbf{P}).\quad (9.74)$$

The integral transport equation is sometimes written in terms of the *emissions-rate density* $\chi(\mathbf{r}, E, \boldsymbol{\Omega})$ defined by Eq. (9.51). Substitution of Eq. (9.61) into Eq. (9.51) immediately gives an integral equation for $\chi(\mathbf{r}, E, \boldsymbol{\Omega})$, namely,

$$\chi(\mathbf{P}) = \int d\mathbf{P}' K(\mathbf{P}' \rightarrow \mathbf{P}) \chi(\mathbf{P}') + S(\mathbf{P}),\quad (9.75)$$

where the kernel $K(\mathbf{P}' \rightarrow \mathbf{P})$ is given by Eq. (9.64).

9.4 Adjoint Transport Equation

In the next chapter, which discusses the use of Monte Carlo techniques for particle transport problems, the concept of *adjoint* flux density is very important when importance sampling is used for variance reduction. In this section, the adjoint transport equation is derived and an important property of the adjoint flux density is introduced.

Operators

First, some mathematical preliminaries. An *operator* $\mathcal{O}$ maps a function $f(\mathbf{P})$ into another function $g(\mathbf{P})$, i.e.,

$$\mathcal{O}f(\mathbf{P}) = g(\mathbf{P}), \quad \mathbf{P} \in R,$$

where $\mathbf{P}$ are all the independent variables of phase space and R is some region of phase space. For example, in this section $\mathbf{P} = (\mathbf{r}, E, \boldsymbol{\Omega})$, the independent variables of the transport equation. An operator is *linear* if $\mathcal{O}[f_1(\mathbf{P}) + f_2(\mathbf{P})] = \mathcal{O}f_1(\mathbf{P}) + \mathcal{O}f_2(\mathbf{P})$.

Inner product

An *inner product* assigns a number to a pair of functions,⁶ namely,

$$(f, g) \equiv \int_R f(\mathbf{P})g(\mathbf{P}) d\mathbf{P} = \int_V dV \int_0^\infty dE \int_{4\pi} d\Omega f(\mathbf{r}, E, \boldsymbol{\Omega})g(\mathbf{r}, E, \boldsymbol{\Omega}),$$

where it is assumed that such integrals exist.

Adjoint operator

Let $\mathcal{O}$ be a real continuous linear operator that operates on square integrable functions $f(\mathbf{P})$. $\mathcal{O}^\dagger$ is the *adjoint* of $\mathcal{O}$ if

$$(g^\dagger, \mathcal{O}f) = (\mathcal{O}^\dagger g^\dagger, f) \quad (9.76)$$

or, equivalently,

$$\int_R g^\dagger(\mathbf{P})[\mathcal{O}f(\mathbf{P})] d\mathbf{P} = \int_R [\mathcal{O}^\dagger g^\dagger(\mathbf{P})]f(\mathbf{P}) d\mathbf{P}. \quad (9.77)$$

Here, $g^\dagger$ are square integrable functions of $\mathbf{P}$ often with different properties (although not necessarily) than those of the functions $f(\mathbf{P})$. For example, if $\mathcal{O}$ is a differential operator, $f(\mathbf{P})$ and $g^\dagger(\mathbf{P})$ may have different behavior at the boundary of V .

⁶ Throughout this section, it is assumed that all operators and functions are real.

9.4.1 Derivation of the Adjoint Transport Equation

With the above definitions, the adjoint of the various operators in the transport equation Eq. (9.49) can be found. First, consider the streaming operator $\mathbf{\Omega} \cdot \nabla$. Because $\mathbf{\Omega} \cdot \nabla = \nabla \cdot \mathbf{\Omega}$, the vector identity

$$\nabla \cdot (\mathbf{\Omega} \phi^\dagger) = \phi^\dagger \mathbf{\Omega} \cdot \nabla \phi + \phi \mathbf{\Omega} \cdot \nabla \phi^\dagger$$

can be integrated to give

$$\int_S dS \mathbf{n} \cdot \mathbf{\Omega} \phi^\dagger = \int_V dV \phi^\dagger \mathbf{\Omega} \cdot \nabla \phi + \int_V dV \phi \mathbf{\Omega} \cdot \nabla \phi^\dagger, \quad (9.78)$$

where the divergence theorem has been used to convert the volume integral on the left to a surface integral and $\mathbf{n}$ a unit outward normal on the surface of the medium. Notice that the boundary condition of Eq. (9.52) makes the integrand of this surface integral vanish for $\mathbf{n} \cdot \mathbf{\Omega} < 0$. This integrand can be forced to vanish entirely if the adjoint flux density is required to satisfy the boundary condition

$$\phi^\dagger(\mathbf{r}_s, E, \mathbf{\Omega}) = 0, \quad \mathbf{r}_s \in S, \quad \mathbf{n} \cdot \mathbf{\Omega} \geq 0. \quad (9.79)$$

Finally, integration of Eq. (9.78) over all E and $\mathbf{\Omega}$ gives

$$\int_R \phi^\dagger(\mathbf{P}) [\mathbf{\Omega} \cdot \nabla \phi(\mathbf{P})] d\mathbf{P} = \int_R [-\mathbf{\Omega} \cdot \nabla \phi^\dagger(\mathbf{P})] \phi(\mathbf{P}) d\mathbf{P}. \quad (9.80)$$

Comparison of this result to the definition of the adjoint given by Eq. (9.77) immediately reveals that $(\mathbf{\Omega} \cdot \nabla)^\dagger = -(\mathbf{\Omega} \cdot \nabla)$.

The operator for the collision term $\mu\phi$ in Eq. (9.51) is simply $\mathcal{O} = \mu(\mathbf{r}, E)$, a multiplicative operator. This is a self-adjoint operator because $\phi^\dagger \mu \phi = [\mu \phi^\dagger] \phi$.

Finally, consider a general integral operator $\mathcal{O}[\cdot] \equiv \int H(\mathbf{P}' \rightarrow \mathbf{P})[\cdot] d\mathbf{P}'$. By definition,

$$(\phi^\dagger, \mathcal{O}\phi) = \int_R \phi^\dagger(\mathbf{P}) \left[\int_R H(\mathbf{P}' \rightarrow \mathbf{P}) \phi(\mathbf{P}') d\mathbf{P}' \right] d\mathbf{P}.$$

Interchange the order of integration and interchange $\mathbf{P}$ and $\mathbf{P}'$ to obtain

$$(\phi^\dagger, \mathcal{O}\phi) = \int_R \left[\int_R H(\mathbf{P} \rightarrow \mathbf{P}') \phi^\dagger(\mathbf{P}') d\mathbf{P}' \right] \phi(\mathbf{P}) d\mathbf{P}.$$

This must equal $(\mathcal{O}^\dagger \phi^\dagger, \phi)$ so it is seen that

$$\mathcal{O}^\dagger[\cdot] = \int H(\mathbf{P} \rightarrow \mathbf{P}')[\cdot] d\mathbf{P}'. \quad (9.81)$$

Thus, the effect of taking the adjoint of an integral operator is to interchange the primed and unprimed variables in the kernel H .

From these results, the integro-differential adjoint equation is found from Eq. (9.49) to be

$$\begin{aligned} & -\mathbf{\Omega} \cdot \nabla \phi^\dagger(\mathbf{r}, E, \mathbf{\Omega}) + \mu(\mathbf{r}, E) \phi^\dagger(\mathbf{r}, E, \mathbf{\Omega}) \\ & = \int_0^\infty dE' \int_{4\pi} d\Omega' \mu(\mathbf{r}, E \rightarrow E', \mathbf{\Omega} \rightarrow \mathbf{\Omega}') \phi^\dagger(\mathbf{r}, E', \mathbf{\Omega}') + S^\dagger(\mathbf{r}, E, \mathbf{\Omega}), \end{aligned} \quad (9.82)$$

where $S^\dagger(\mathbf{r}, E, \mathbf{\Omega})$ is some arbitrary distributed adjoint source and the boundary condition is

$$\phi^\dagger(\mathbf{r}_s, E, \mathbf{\Omega}) = 0, \quad \mathbf{n} \cdot \mathbf{\Omega} \geq 0. \quad (9.83)$$

Similarly, the adjoint integral transport equation for the adjoint flux density is obtained from Eq. (9.63) as

$$\phi^\dagger(\mathbf{P}) = \int d\mathbf{P}' K(\mathbf{P} \rightarrow \mathbf{P}') \phi^\dagger(\mathbf{P}') + \tilde{S}^\dagger(\mathbf{P}). \quad (9.84)$$

Later, the adjoint equation for the collision density is needed, which from Eq. (9.73) is

$$F^\dagger(\mathbf{P}) = \int d\mathbf{P}' \bar{K}(\mathbf{P} \rightarrow \mathbf{P}') F^\dagger(\mathbf{P}') + \bar{S}^\dagger(\mathbf{P}). \quad (9.85)$$

9.4.2 Utility of the Adjoint Solution

Suppose it is desired to evaluate some functional of the flux density

$$I = \int_R f(\mathbf{P}) \phi(\mathbf{P}) d\mathbf{P}, \quad (9.86)$$

where $f(\mathbf{P})$ is a specified *weight function*. For example, if $f(\mathbf{P}) = 1/V$, then I would give the average scalar flux density in the spatial volume V . Likewise, if $f(\mathbf{P}) = \delta(\mathbf{P} - \mathbf{P}_o)$, then $I = \phi(\mathbf{P}_o)$, the flux density at a particular position, energy, and direction. Almost any property of the particle field can be obtained with the proper choice of the weight function.

To evaluate I one could first find ϕ by solving the transport equation

$$\mathcal{O}\phi(\mathbf{P}) = S(\mathbf{P}), \quad (9.87)$$

where the transport operator $\mathcal{O}$ is found by writing Eq. (9.49) in the above form. Then once ϕ is known, Eq. (9.86) could be evaluated numerically. It so happens that the same value of I can be found from the corresponding adjoint transport problem

$$\mathcal{O}^\dagger \phi^\dagger = f(\mathbf{P}), \quad (9.88)$$

where the adjoint source is the weight function in I . To see this alternate method for evaluating I , multiply Eq. (9.87) by $\phi^\dagger$ and Eq. (9.88) by ϕ and integrate the results over all $\mathbf{P}$. The results are

$$(\phi^\dagger, \mathcal{O}\phi) = \int_R S(\mathbf{P})\phi^\dagger(\mathbf{P}) d\mathbf{P} \quad \text{and} \quad (\mathcal{O}^\dagger\phi^\dagger, \phi) = \int_R f(\mathbf{P})\phi(\mathbf{P}) d\mathbf{P}.$$

From the definition of the adjoint, the two inner products on the left sides are equal. Thus, it is seen that

$$I = \int_R f(\mathbf{P})\phi(\mathbf{P}) d\mathbf{P} = \int_R S(\mathbf{P})\phi^\dagger(\mathbf{P}) d\mathbf{P}. \quad (9.89)$$

This important result has practical applications and provides insight into the significance of the adjoint flux.

To understand the significance of the adjoint flux density, suppose that $f(\mathbf{P})$ is some detector response function $\mu_d(\mathbf{r}, E)$ defined for some subregion of V and is zero outside that subregion. The units of μ_d are the probability, per unit path length, that a particle interacts with the detector material and produces a “count,” i.e., cm^{-1} . Hence, the units of I , from the first integral in Eq. (9.89), are “counts per unit time.” Further, suppose that one particle is born per unit time at $\mathbf{r}_o$ with energy E_o and in direction $\boldsymbol{\Omega}_o$, i.e., $S(\mathbf{P}) = \delta(\mathbf{r} - \mathbf{r}_o)\delta(E - E_o)\delta(\boldsymbol{\Omega} - \boldsymbol{\Omega}_o) \{s^{-1}\}$. Then, from the second integral in Eq. (9.89),

$$I = \phi^\dagger(\mathbf{r}_o, E_o, \boldsymbol{\Omega}_o) = \text{detector count rate}. \quad (9.90)$$

From Eq. (9.82) with $S^\dagger = \mu_d$, it is seen that $\phi^\dagger$ is dimensionless and, from the above result, it is seen that the adjoint flux density gives the *importance* of a particle born at $(\mathbf{r}_o, E_o, \boldsymbol{\Omega}_o)$ to the detector response. This interpretation is consistent with the adjoint boundary condition that $\phi^\dagger = 0$ for particles reaching the surface traveling in the outward direction because such particles can never return and contribute to the detector response. That the adjoint flux density describes how important different regions of phase space are to the tally or score in a Monte Carlo simulation gives a powerful tool for how to pick the importance functions [Coveyou et al. 1967]. Indeed, it will be shown that if $\phi^\dagger$ is used as the importance function, the tally has zero variance!

Another very practical use of the adjoint transport equation is to transform a hard problem into an easier one. For example, if the average flux in some small region of phase space is sought and the source is widely distributed, then in a Monte Carlo simulation few histories will even reach the small region and contribute to the tally. In other words, the first integral in Eq. (9.89) is difficult to evaluate. By contrast, the adjoint problem has the source in the small volume and the source region becomes the tally region, making the second integral in Eq. (9.89) much easier to evaluate.

9.5 Summary

In this chapter, the fundamentals were presented: (1) how the strength of a radiation field of neutral particles, such as neutrons and photons, is characterized by the flux density $\phi(\mathbf{r}, E, \Omega)$, (2) how the flow across a surface is characterized by the current vector $\mathbf{J}(\mathbf{r}, E, \Omega)$, (3) how the interaction coefficient determines the probability a radiation particle interacts with the host medium, and (4) how the reaction rate per unit volume of the radiation with the medium is given by $R_i(\mathbf{r}, E) = \mu_i(\mathbf{r}, E)\phi(\mathbf{r}, E)$. It was then shown that uncollided neutral-particle radiation is attenuated exponentially as it traverses the host medium, a key result that is used in the next chapter on Monte Carlo transport simulation. The interaction coefficient and associated microscopic cross sections for both photons and neutrons were illustrated.

Two forms of the linearized Boltzmann equation were derived that describe exactly how radiation migrates through a host medium. The integro-differential form, which is widely used in deterministic transport calculations, was then converted into an integral form, the form first used in Monte Carlo analysis of radiation transport and that is also used in the following chapter. Both the direct and adjoint integral forms were introduced. As shown in the next chapter, the adjoint solution plays an important role in maximizing variance reduction of Monte Carlo evaluation of the direct solution.

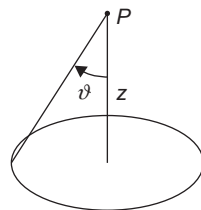
Bibliography

- ADAMS, M.L. AND E.L. LARSEN 2002, "Fast Iterative Methods for Discrete-Ordinates Particle Transport Calculations," *Prog. Nucl. Energ.*, **40**(1), 3–159.
- BELL, G.I. AND S. GLASSTONE 1970, *Nuclear Reactor Theory*, Van Nostrand Reinhold, New York.
- COVEYOU R.R., V.R. CAIN, AND K.J. YOST 1967, "Adjoint and Importance in Monte Carlo Applications," *Nucl. Sci. Eng.*, **27**, 219–234.
- DUDERSTADT, J.J. AND W.R. MARTIN 1979, *Transport Theory*, Wiley, New York.
- LEWIS, E.E. AND W.F. MILLER 1984, *Computational Methods of Neutron Transport Theory*, Wiley, New York.
- SANCHEZ, R. AND N.J. MCCORMICK 1982, "A Review of Neutron Transport Approximations," *Nucl. Sci. Eng.*, **80**, 481–535.
- SHULTIS, J.K. AND R.E. FAW 2000, *Radiation Shielding*, American Nuclear Society, LaGrange Park, IL.
- SHULTIS, J.K. AND R.E. FAW 2008, *Fundamentals of Nuclear Science and Engineering*, 2nd ed., CRC Press, Boca Raton, FL.

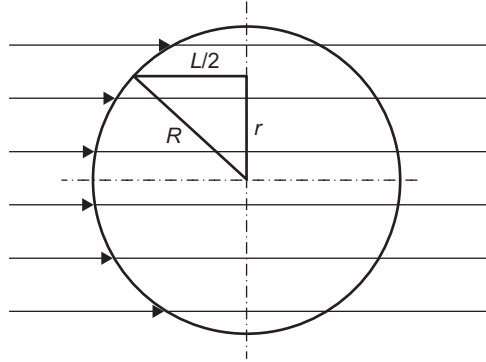
Problems

1. Show that the solid angle subtended by the disk shown to the right at the point P , which is a distance z from the disk on a line perpendicular to the center of the disk, is

$$\Omega = 2\pi(1 - \cos \vartheta).$$



2. Equations (9.13) and (9.14) are two definitions of the fluence. Consider the sphere illustrated to the right, with a uniform parallel beam of particles incident on its surface, so that ΔN_p particles enter the sphere. Verify that the mean chord length is $\langle L \rangle = 4R/3$ and use this result to demonstrate the equivalence of the two definitions.



3. Show that if the angular flux density $\phi(\mathbf{r}, \mathbf{\Omega})$ is isotropic, the current density $\mathbf{j}(\mathbf{r})$ is equal to zero.
4. A broad beam of neutrons is normally incident on a homogeneous slab 6-cm thick. The intensity of neutrons transmitted through the slab without interactions is found to be 30% of the incident intensity. (a) What is the total interaction coefficient μ_t for the slab material? (b) What is the average distance a neutron travels in this material before undergoing an interaction?
5. A small homogeneous sample of mass m (g) with atomic mass A is irradiated uniformly by a constant flux density ϕ ($\text{cm}^{-2} \text{s}^{-1}$). If the total atomic cross section for the sample material with the irradiating particles is denoted by σ_t (cm^2), derive an expression for the fraction of the atoms in the sample that interact during a one-hour irradiation. State any assumptions made.
6. The transport equation derived in Section 9.3 is based on the angular flux density $\phi(\mathbf{r}, E, \mathbf{\Omega})$ in terms of the particle energy E and direction of travel $\mathbf{\Omega}$. However, an angular flux density $\phi(\mathbf{r}, \mathbf{v})$ based on the particle velocity $\mathbf{v}$ could be used instead. Derive the relation between these two flux densities.
7. Show that $\nabla \cdot \mathbf{\Omega} f(\mathbf{r}, \mathbf{\Omega}) = \mathbf{\Omega} \cdot \nabla f(\mathbf{r}, \mathbf{\Omega})$, where f is a continuous function in $\mathbf{r}$.
8. In many radiation shielding analyses, the uncollided flux density is of interest. The uncollided angular flux density can be obtained from the transport equation by treating the medium as a purely absorbing one (i.e., any interaction removes the particle from its uncollided status and no secondary uncollided particles are created). Show that for a homogeneous medium, the uncollided angular flux density is given by

$$\phi(\mathbf{r}, E, \mathbf{\Omega}) = \int_0^\infty dR S(\mathbf{r} - R\mathbf{\Omega}, E, \mathbf{\Omega}) e^{-\mu(E)R}.$$

9. Consider a half-space of purely absorbing material in which there is a uniformly distributed, isotropic, volumetric source that emits S_v photons of energy E_o per unit time. Determine the angular flux density and current of photons at the surface of the half-space.
10. Consider a homogeneous slab of thickness $2T$ of which the cross sections and sources vary only in the x direction, which is perpendicular to the slab faces and of which the origin is at the slab center. Show that the integral transport equation for isotropic scattering and

sources, given by Eq. (9.71), reduces to

$$\phi(x, E) = \frac{1}{2} \int_{-T}^T dx' \int_{|x-x'|}^{\infty} \frac{e^{-\mu(E)R'}}{2R'} q_o(x', E) = \frac{1}{2} \int_{-T}^T dx' E_1(\mu|x-x'|) q_o(x', E),$$

where

$$q_o(x', E) = S_o(x', E) + \int_0^{\infty} dE' \mu_{so}(x', E' \rightarrow E) \phi(x', E').$$

Hint: Express $\int dV'$ as $\int dx' \int 2\pi\rho d\rho$, where $\rho^2 + |x - x'|^2 = R'^2$.

10 Monte Carlo Simulation of Neutral Particle Transport

*Monte Carlo transport calculations
may sometimes not meet expectations.
The trouble, of course,
is the use of brute force
rather than clever implementation.*

Numerical solutions of the transport equation, called *deterministic* calculations, usually require a discretization of the time, energy, angular, and spatial variables. Each such approximation introduces a *systematic* error. Moreover, the geometry for deterministic calculations must be rather simple, for example, planar, cylindrical, or spherical surfaces, and a full three-dimensional calculation requires tremendous computer memory to hold all the discretized values of the energy and angular flux density.

Monte Carlo calculations, by contrast, are capable of treating very complex three-dimensional geometries. The continuous treatment of time, energy, direction, and space avoids all systematic discretization errors. However, the stochastic simulation of the transport process introduces *statistical* errors in the results, errors that are not present in deterministic calculations. Historically, Monte Carlo analyses could be performed only on large mainframe computers because of the intensive computational requirements. But today's inexpensive personal computers are now capable of running very complex Monte Carlo simulations.

Monte Carlo calculations are not always better than deterministic approaches. Monte Carlo excels at estimating some quantity averaged over, at most, a few regions of phase space $\mathbf{P} = (\mathbf{r}, E, \boldsymbol{\Omega}, t)$. But, if, for example, detailed three-dimensional spatial profiles of the flux in a reactor core are needed, deterministic methods often are superior. The reason for this is that, to obtain with Monte Carlo the value of the flux density at a large number of points, the spatial domain would have to be broken into a vast number of small contiguous volumes ΔV and the flux density estimated in each of these volumes. The number of histories that contribute to the flux estimates in the volumes decreases rapidly as ΔV is made smaller to increase the spatial resolution, thereby causing the statistical uncertainty in the estimated flux densities to increase, sometimes to unacceptable values. However, Monte Carlo techniques are, or can be made, efficient when the number of results or tallies are small.

10.1 Basic Approach for Monte Carlo Transport Simulations

A Monte Carlo radiation transport calculation simulates a finite number N of particle histories by sampling from the appropriate PDFs governing the various events that may happen to a particle from its birth by a source to its eventual demise by being absorbed or escaping through the problem boundary. A history begins by randomly selecting a particle's initial position, energy, and direction from the PDFs that describe the sources for the problem. As the particle travels on various legs of its trajectory, sampling is performed to determine random values for distance to the next interaction, type of interaction, scattering angle, energy change, and so on. Between interactions, the neutrons or photons travel in straight lines. On some interactions, secondary particles are created, such as neutrons from fission reactions or X-rays from photon photoelectric interactions. These secondary particles are "banked" and their initial energies, interaction positions, and travel directions recorded. After the original particle finishes its history, these secondary banked particles are processed sequentially to create a history for each.

After each history is terminated (or after its interactions or surface crossings are determined), its contribution z_i is added to the tally being used to estimate the quantity of interest, for example, average flux in a cell, probability of escape, energy deposited in a cell, current through a surface, or some other quantity related to the radiation field. After N histories, the estimate of $\langle z \rangle$ is

$$\langle z \rangle \simeq \bar{z} = \frac{1}{N} \sum_{i=1}^N z_i.$$

Along with such an estimate, various measures of the quantity's statistical reliability must be made (see Section 4.2).

Such an analog simulation mimics the stochastic events that befall an actual particle if the problem were converted to an equivalent experiment. In such calculations, extremely few of the histories contribute anything to the tally of interest. For example, if the probability a neutron incident on a thick shield eventually reaches the other side is 1 in 10^9 , then in an analog simulation only 1 out of 10^9 histories, on the average, terminate on the back side of the shield. For such problems, it is vital to use variance reduction techniques (see Chapter 5) and perform *nonanalog* Monte Carlo calculations.

10.2 Geometry

Almost all Monte Carlo codes use a Cartesian coordinate system, even if the problem possesses cylindrical or spherical symmetry. The reason for this is that, as a particle travels along the direction $\mathbf{\Omega} = \mathbf{i}u + \mathbf{j}v + \mathbf{k}w$, the direction cosines $u = \mathbf{i} \cdot \mathbf{\Omega}$, $v = \mathbf{j} \cdot \mathbf{\Omega}$, and $w = \mathbf{k} \cdot \mathbf{\Omega}$ with respect to the coordinate axes do not change. If a particle moves a distance s from $\mathbf{r}_0$ in direction $\mathbf{\Omega}$, its new position $\mathbf{r}$ is

$$\mathbf{r} = \mathbf{r}_0 + s\mathbf{\Omega} = (x_0 + su, y_0 + sv, z_0 + sw). \quad (10.1)$$

The three-dimensional space in which the simulated histories are constructed is usually assumed to be composed of contiguous homogeneous volumes or *cells*, each cell being bounded by one or more surfaces (or portions of surfaces). A cell can be a void or composed of any homogeneous material. All of the three-dimensional space must belong to some cell—there can be no “holes” where a particle would become “lost.” Although not necessary, the geometry of the problem is often surrounded by a *problem boundary*, usually far from tally regions. The cell containing all space beyond the problem boundary is the “graveyard” of which the purpose is to “kill” any particle that enters it, thereby saving time by not tracking particles beyond the boundary and that have negligible chance of contributing to the tally. For example, in calculating the doses to a patient from an X-ray therapy unit, the walls of the treatment room should be modeled, but nearby buildings and airplanes passing overhead have little chance of affecting the dose to the patient. Similarly, objects far from tally regions often can be modeled crudely compared to objects near the important regions. For example, door knobs in the treatment room need not be modeled unless the problem is to determine how much radiation leaks through the key hole. Constructing an effective but efficient geometry model for a particular problem is a matter of skill and experience on the modeler’s part.

10.2.1 Combinatorial Geometry

Surfaces are usually defined by first-degree functions (planes) or second-degree functions (cylinders, cones, ellipsoids, etc.) as $f(x, y, z) = 0$. Sometimes fourth-degree functions (toroids) are used, although such surfaces present greater numerical difficulties for evaluating the intersection point with a particle trajectory. Commonly used surfaces are given in Table 10.1. Surface A , defined by $f_A(x, y, z) = 0$, divides all space into two subdomains, namely, those points $\mathbf{r} = (x, y, z)$ for which $f_A(x, y, z) > 0$ (denoted by A^+) and those points for which $f_A(x, y, z) < 0$ (denoted by A^-). For example, consider the spherical surface S : $f_S(x, y, z) = x^2 + y^2 + z^2 - R^2 = 0$. All points inside the sphere (S^-) are to the negative side of the surface, while all points outside (S^+) are to the positive side of the surface.

By using Boolean unions ($\cup$, “and”) and intersections ($\cap$, “or”) of the two subdomains for each surface, volumes of quite complex shapes can be approximated. As a simple example, consider the construction of a finite cylinder of radius R and height $2H$, centered at the origin, and with its axis on the x -axis. Such a cylindrical volume is bounded by two parallel planes perpendicular to the x -axis, $f_{P_1}(x, y, z) \equiv x - H = 0$ and $f_{P_2}(x, y, z) \equiv x + H = 0$, and the infinite cylinder $f_C(x, y, z) \equiv y^2 + z^2 - R^2 = 0$. The space inside this cylinder is specified as

$$P_1^- \cap P_2^+ \cap C^-,$$

while all of the space outside this finite cylinder is described as

$$P_1^+ \cup P_2^- \cup C^+.$$

To permit rapid particle tracking through the geometry, each region (or cell) and each surface is uniquely numbered. Then, for each cell, the numbers of the surfaces bounding it are tabulated as are the specification of the regions on the other side of the bounding surfaces. Although somewhat redundant, this bookkeeping makes tracking particles very efficient, because when a particle leaves a region, the adjacent

Table 10.1 Some first-, second-, and fourth-degree surfaces. The location and orientation of the surfaces are determined by the values of the surface parameters denoted by capital and Greek letters. Extracted from MCNP [2003].

Type	Description	Surface $f(x, y, z) = 0$
plane	general	$Ax + By + Cz - D = 0$
	normal to x -axis	$x - D = 0$
	normal to y -axis	$y - D = 0$
	normal to z -axis	$z - D = 0$
sphere	centered at origin	$x^2 + y^2 + z^2 - R^2 = 0$
	general	$(x-A)^2 + (y-B)^2 + (z-C)^2 - R^2 = 0$
	centered on x -axis	$(x-A)^2 + y^2 + z^2 - R^2 = 0$
	centered on y -axis	$x^2 + (y-B)^2 + z^2 - R^2 = 0$
	centered on z -axis	$x^2 + y^2 + (z-C)^2 - R^2 = 0$
cylinder	parallel to x -axis	$(y-B)^2 + (z-C)^2 - R^2 = 0$
	parallel to y -axis	$(x-A)^2 + (z-C)^2 - R^2 = 0$
	parallel to z -axis	$(x-A)^2 + (y-B)^2 - R^2 = 0$
	on x -axis	$y^2 + z^2 - R^2 = 0$
	on y -axis	$x^2 + z^2 - R^2 = 0$
	on z -axis	$x^2 + y^2 - R^2 = 0$
cone	parallel to x -axis	$(y-B)^2 + (z-C)^2 - D(x-A)^2 = 0$
	parallel to y -axis	$(x-A)^2 + (z-C)^2 - D(y-B)^2 = 0$
	parallel to z -axis	$(x-A)^2 + (y-B)^2 - D(z-C)^2 = 0$
	on x -axis	$y^2 + z^2 - D(x-A)^2 = 0$
	on y -axis	$x^2 + z^2 - D(y-B)^2 = 0$
	on z -axis	$x^2 + y^2 - D(z-C)^2 = 0$
ellipsoid hyperboloid paraboloid	axis parallel to x -, y -, or z -axis	$A(x-\alpha)^2 + B(y-\beta)^2 + C(z-\gamma)^2$ $+ 2D(x-\alpha) + 2E(y-\beta)$ $+ 2F(z-\gamma) + G = 0$
cylinder, cone ellipsoid paraboloid hyperboloid	axis not parallel to x -, y -, or z -axis	$Ax^2 + By^2 + Cz^2 + Dxy + Eyz$ $+ Fzx + Gz + Hy + Jz + K = 0$
elliptical or circular torus axis is parallel to x -, y -, or z -axis		$(x-\alpha)^2/B^2 + (\sqrt{(y-\beta)^2 + (z-\gamma)^2} - A)^2/C^2 - 1 = 0$
		$(y-\beta)^2/B^2 + (\sqrt{(x-\alpha)^2 + (z-\gamma)^2} - A)^2/C^2 - 1 = 0$
		$(z-\gamma)^2/B^2 + (\sqrt{(x-\alpha)^2 + (y-\beta)^2} - A)^2/C^2 - 1 = 0$

regions are immediately known as are the new surfaces toward which the particle is traveling.

10.3 Sources

Each particle history is begun by sampling from the spatial, energy, and angular distribution of the source to determine the starting position, energy, and direction of each particle history. The source may be specified explicitly (a *fixed-source* problem) or calculated by the simulation itself (an *eigenvalue* problem). In the latter case, initial simulations are used to estimate the spatial distribution of sources such as a fission source and the resulting equilibrium distribution in a critical system. However, in this chapter only the fixed-source problem is discussed.

Sources in a transport calculation can be distributed over many regions, be localized to a few cells, or be singular sources such as point, line, and plane sources. The energy of particles emitted by the sources can consist of a discrete set of energies (typical of gamma rays produced by radioactive decay) or have a continuous distribution of energies (such as neutrons produced by spontaneous fission). While many sources emit radiation isotropically, some sources can be anisotropic (such as a beam of radiation from a reactor beam port). To start a particle history, one must sample from the appropriate spatial, energy, and angular distributions. To complicate matters, these three distributions are often interrelated. For example, different source regions may emit particles with different energy and angular distributions. Modeling the source distributions can often be quite a complex task.

10.3.1 Isotropic Sources

Many physical sources emit radiation isotropically. The probability $p(\mathbf{\Omega})d\mathbf{\Omega}$ a source particle is emitted in $d\mathbf{\Omega}$ about a direction $\mathbf{\Omega}$ is thus

$$p(\mathbf{\Omega})d\mathbf{\Omega} = \frac{d\mathbf{\Omega}}{4\pi} = \frac{d\psi}{2\pi} \frac{d\omega}{2}, \quad (10.2)$$

where $\omega = \cos \theta$. Thus, the PDF of the azimuthal angle is seen to be $p(\psi) = 1/(2\pi)$, $0 \leq \psi < 2\pi$, and the PDF for the cosine of the polar angle θ is $p(\omega) = 1/2$, $-1 \leq \omega \leq 1$, both uniform distributions. Sampling from these distributions gives

$$\psi_i = 2\pi\rho_j \quad \text{and} \quad \omega_i = -1 + 2\rho_{j+1}, \quad (10.3)$$

where ρ_j and ρ_{j+1} are two random numbers uniformly distributed between (0, 1). The directed cosines of the initial direction $\mathbf{\Omega}_i$ of a source particle are then calculated as

$$u_i = \mathbf{i} \cdot \mathbf{\Omega} = \cos \psi_i \sqrt{1 - \omega_i^2}, \quad v_i = \mathbf{j} \cdot \mathbf{\Omega} = \sin \psi_i \sqrt{1 - \omega_i^2}, \quad w_i = \mathbf{k} \cdot \mathbf{\Omega} = \omega_i. \quad (10.4)$$

where f_m^i , $m = 1, 2, \dots$ are all the surfaces bounding region i , and by picking s_i , equal to the smallest positive real solution. If $f_m^i(x, y, z)$ are quadratic surfaces,¹ then Eq. (10.8) is a quadratic equation for s_i , namely,

$$as_i^2 + 2bs_i + c = 0, \quad (10.9)$$

where a , b , and c depend on the parameters of f_m^i and on $(x_{i-1}, y_{i-1}, z_{i-1})$. Of course, in performing this calculation for all the surfaces bounding region i , most of the double roots will be imaginary (meaning no intersection) or negative (an intersection in the backward direction). The smallest positive real root gives the path distance s_i in region i .

10.4.2 Convex versus Concave Cells

So far, it has been tacitly assumed that the various regions are all convex, i.e., a straight line between any two points lies wholly within the region. Actually, concavities in a cell due to single surfaces (see Fig. 10.2) present no difficulties with the previous procedure for calculating path lengths in a cell. However, if the concavities are produced by multiple surfaces (the dashed lines in Fig. 10.3) additional tests are required to eliminate irrelevant surface intersections by determining if the same cell is on the other side of these spurious intersections.

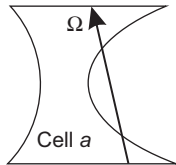


Figure 10.2 A concave cell whose concavities are single surfaces.

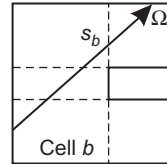


Figure 10.3 A concave cell whose concavities are from multiple surfaces.

10.4.3 Effect of Computer Precision

Although the previous description of how a particle is traced along its trajectory of straight-line segments, punctuated by point interactions, is conceptually simple, its implementation in computer code is not so easy because of the finite precision with which numbers can be represented. The position $\mathbf{r}_i$ of the intersection of a trajectory with surface f_i , bounding the cells i and $i + 1$, may be rounded by the finite precision of the computer to (1) still remain in region i , (2) be placed in cell $i + 1$, or (3) actually

¹ The most general form is $f(x, y, z) = Ax^2 + By^2 + Cz^2 + Dxy + Exz + Fyz + Gx + Hy + Iz + J$. The ten parameters $A - J$ define ten possible surfaces: ellipsoids, cones, cylinders, hyperboloids of one sheet, hyperboloids of two sheets, elliptic paraboloids, hyperbolic cylinders, hyperbolic paraboloids, parabolic cylinders, and planes [Olmsted 1947].

be on the surface (see Fig. 10.4). In all cases, the computer “thinks” the particle is entering cell $i + 1$.

The calculation of the distance s_{i+1} of the path length in the next cell $i + 1$ is done by finding the smallest positive distance to the surfaces bounding cell $i + 1$ (here, surfaces f_i and f_{i+1}). All three cases give essentially the correct distance to surface f_i ; however, the distance to f_i is a small positive value for case 1 (undershoot), a small negative value for case 2 (overshoot), and zero for case 3 (exact).² Only in cases 2 and 3 is the correct distance s_{i+1} to the exit point in cell i obtained.

It is tempting to add a small positive distance to s_i to ensure there are never any undershoots, but such a fix prevents the code from being “scale invariant,” i.e., the size of the correction depends on the size of the geometry. Also, use of higher-precision arithmetic does not fix this problem, but only makes the under/overshoots smaller. Fortunately, there is a simple remedy to avoid these problems.

Rule 1. Do not calculate distances to boundaries from which the particle is receding based on its known direction Ω and the cell the particle thinks it is in. This resolves all problems associated with undershoot, overshoot, or being exactly on the boundary.

A further complicating factor occurs if the particle changes direction on reaching the boundary. This change in direction occurs if the boundary is a “reflecting” boundary (used, for example, to take advantage of problem symmetries) or if (very rarely) the next interaction is on a boundary and is a scattering event. The ambiguities on how near the boundary the particle is and a possible change in its direction can cause serious errors in the calculation of the next leg of the trajectory. In Fig. 10.5, the particle has been transported to surface f_i , its direction reversed, but the computer still thinks it is entering region $i + 1$. In the next step, only the distance to surface f_i is made because the particle is receding from surface $i + 1$. Cases 4, 5, and 6 return a small negative, small positive, and a zero value, respectively. Such problems are resolved by using another tracking rule.

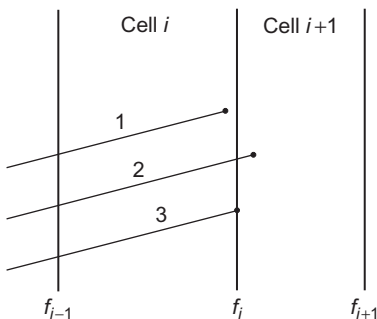


Figure 10.4 Three possibilities when a particle is transported to surface f_i .

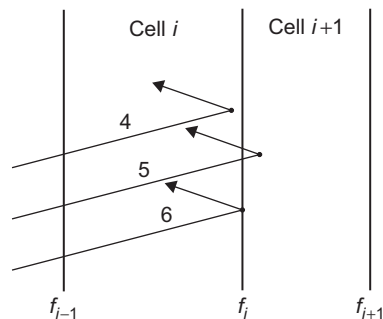


Figure 10.5 Three possibilities when a particle is transported to and reflected from surface f_i .

² Roundoff also affects these small or zero distances to f_i , and any one of them could be rounded into a small positive distance and, thus, give the incorrect distance to the exit point in cell $i + 1$.

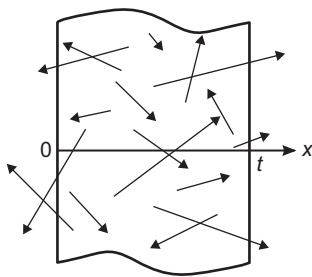
Rule 2. For a flight distance that is negative or zero, use a zero step size and make a region change. This resolves all undershoot or exact intersections when the particle is backscattered at the surface.

Finally, it should be mentioned that, with finite-precision arithmetic, it is possible in complex geometries for a particle to become “lost.” For example, if several cells have a common vertex, a particle reaching the vertex (within machine precision) may not be able to resolve which region it is entering on the other side of the vertex. Such problems, however, are extremely rare (the probability any trajectory will exactly pass through any specified point is zero). For problems using millions of histories, one or two “lost” particles has a negligible effect on the results.

10.5 Purely Absorbing Media

A degenerate form of transport theory is to calculate the probability that a neutron born in a region escapes without an interaction. In essence, each collision is treated as an absorption, and the problem is to estimate the probability of escape from the region without undergoing any collision. In other words, what is the fraction of the particles born in the region that have a path length to the first collision site that lies outside the region? For this special case, exact analytic solutions for simple geometries are known [Case et al. 1953]. More importantly, such escape-probability calculations are readily evaluated with the Monte Carlo method using only track-length simulation. Such a calculation is given in Example 10.1 and compared to the analytical result.

Example 10.1 Escape Probability from an Infinite Slab



A simple application of sampling path lengths of neutral particles is the determination of the escape probability P_{esc} of particles emitted uniformly in some purely absorbing body, for which $\mu = \mu_a$. For simple bodies, such as infinite slabs, spheres, and infinite cylinders, analytical solutions of the transport equation of Eq. (9.49) are known [Case et al. 1953]. As an example, consider the case in which particles are emitted uniformly and isotropically throughout an infinite absorbing slab of thickness t as shown in the figure to the left.

Because the slab is infinite in the y and z directions, only the x position where the particle is born affects its chance of escape. Further, for isotropic emission, the cosine of the emission angle θ with respect to the positive polar x -axis is uniformly distributed in $(-1, 1)$. The azimuthal angle of emission does not affect the escape probability, so only the polar angle θ needs to be simulated.

The following Monte Carlo game is played N times to determine the number of simulated particles that escape, N_{esc} .

1. Selected an emission point uniformly in $(0, t)$ as $x_i = t\rho_i$.
2. Selected emission direction $\omega \equiv \cos \theta$ as $\omega_i = 2\rho_i - 1$.
3. Calculate the distance d to the surface: if $\omega_i > 0$, then $d_i = (t - x_i)/\omega_i$; otherwise, $d_i = x_i/\omega_i$.

4. Generate a random path length s for the particle's first collision (absorption) in the material from the distribution $f(s) = \mu e^{-\mu s}$ as (see Example 4.2) $s_i = [-\ln \rho_i]/\mu$.
5. If $s_i < d_i$, the particle is absorbed and $z_i = 0$; otherwise, the particle escapes and $z_i = 1$.

After performing this simulation N times the escape probability is estimated from the two summations $\bar{z} = (1/N) \sum_{i=1}^N z_i$ and $\overline{z^2} = (1/N) \sum_{i=1}^N z_i^2$ as

$$P_{esc} \simeq \bar{z} \pm \frac{1}{\sqrt{N}} \left[\overline{z^2} - \bar{z}^2 \right].$$

The analytic solution for this problem is [Case et al. 1953]

$$P_{esc} = \mu t \left[\frac{1}{2} - E_3(\mu t) \right],$$

where $E_3(x) \equiv x^2 \int_x^\infty u^{-3} e^{-u} du$ is the exponential integral function of order 3. Results using $N = 1000$ particles are compared to the analytical result for several slab widths in Fig. 10.6.

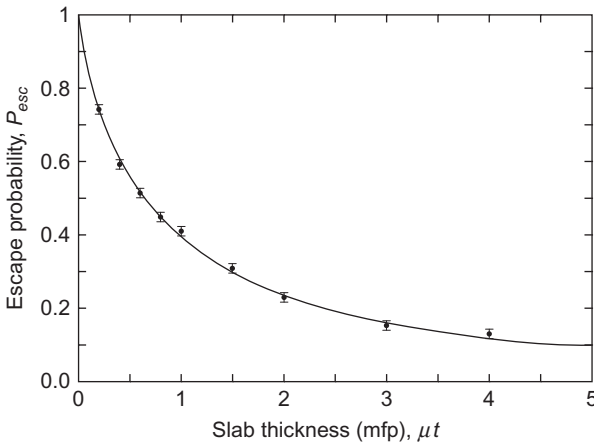


Figure 10.6 The escape probability P_{esc} for a purely absorbing slab with a uniform isotropic source of particles. The line is the analytical result, and the data are Monte Carlo simulation results using 1000 histories for each point. Error bars are the estimated standard deviations.

10.6 Type of Collision

At the end of a straight-line path segment, the particle undergoes an interaction with the material in the current cell.³ The probability that the particle with energy E interacts

³ Of course, if the collision site is outside the problem boundary, the history is terminated, tallies are updated, and a new source particle history is started.

with species j (an isotope for neutrons or an element for photons) in a reaction of type i is

$$p_i^j = \frac{\mu_i^j(E)}{\sum_i \sum_j \mu_i^j(E)}, \quad (10.10)$$

where the summations are over all species in the cell and over all possible reaction types.

If the reaction is one in which no particle of the type being tracked emerges from the interaction, the particle is absorbed and the history ends. If more than one particle of the type being considered results from the interaction (e.g., fission for neutrons or fluorescence and X-rays for photons), the energy and direction of all but one are “banked” for later processing and one particle continues the presented history.

10.6.1 Scattering Interactions

Scattering is a frequent type of interaction in which the incident particle is deflected into a new direction $\mathbf{\Omega}'$ from its old direction $\mathbf{\Omega}$ through a scattering angle $\theta_s = \cos^{-1} \mathbf{\Omega} \cdot \mathbf{\Omega}'$. In a scattering interaction with species j , the constraints of conservation of energy and momentum force the doubly differential scattering interaction coefficient (see Section 9.2.4) $\mu_s^j(E \rightarrow E', \mathbf{\Omega} \cdot \mathbf{\Omega}')$ to be proportional to $\delta(\mathbf{\Omega} \cdot \mathbf{\Omega}' - S(E, E'))$, i.e., if any two of the variables E , E' , and $\cos \theta_s = \mathbf{\Omega} \cdot \mathbf{\Omega}'$ are specified, the third is forced to have a value that makes the argument of the delta function zero. Thus, the cosine of the scattering angle is sampled (usually with the rejection technique) from the PDF formed from the empirical⁴ singly differential scattering interaction coefficient

$$f(\mathbf{\Omega} \cdot \mathbf{\Omega}') = \frac{\mu_s^j(E, \mathbf{\Omega} \cdot \mathbf{\Omega}')}{\int_{-1}^1 \mu_s^j(E, \mathbf{\Omega} \cdot \mathbf{\Omega}') d(\mathbf{\Omega} \cdot \mathbf{\Omega}')} \quad (10.11)$$

Once $\cos \theta_s$ is determined, the new energy of the particle is obtained by solving

$$\mathbf{\Omega} \cdot \mathbf{\Omega}' = \cos \theta_s = S(E, E') \quad (10.12)$$

for E' .

Direction cosines of the scattered particle

Once the scattering angle and energy of the scattered particle are determined, the direction cosines (u' , v' , w') of $\mathbf{\Omega}'$ need to be calculated. The geometry is illustrated in Fig. 10.7. The change in azimuth $\Delta\psi_s = \psi' - \psi$ is uniformly distributed in $(0, 2\pi)$ radians. Let θ and ψ be the polar and azimuthal angles before scatter, with direction cosines $u = \sin \theta \cos \psi$, $v = \sin \theta \sin \psi$, and $w = \cos \theta$.

⁴ Only for the case of Compton scattering of photons is there an analytical expression for the differential scattering interaction coefficient.

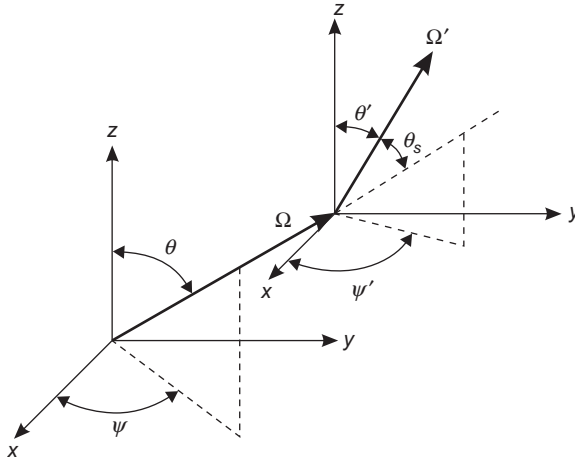


Figure 10.7 The relation between the coordinates of the particle in direction Ω before a scatter and the direction Ω' after the scatter. The angle between Ω and Ω' is the scattering angle θ_s , namely, $\cos \theta_s = \Omega \cdot \Omega'$.

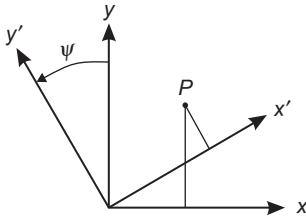


Figure 10.8 A rotation in the positive sense about the z -axis.

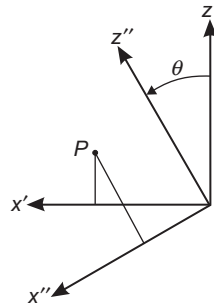


Figure 10.9 A rotation about the y' -axis in the positive sense.

To calculate the direction cosines (u' , v' , w') of Ω' it is first necessary to introduce rotation matrices $\mathbf{R}$. Consider first a rotation about the z -axis through an angle ψ , as shown in Fig. 10.8. The coordinates of the point P in the original system (x, y, z) and in the rotated systems (x', y', z') are readily shown to be related by

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \cos \psi & -\sin \psi & 0 \\ \sin \psi & \cos \psi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} \equiv \mathbf{R}_z(\psi) \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix}. \quad (10.13)$$

If this rotated coordinate system is then itself rotated about the y' -axis by an angle θ , as shown in Fig. 10.9, the relation between the coordinates of the point P in both systems

is given by

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \cos \theta & 0 & \sin \theta \\ 0 & 1 & 0 \\ -\sin \theta & 0 & \cos \theta \end{pmatrix} \begin{pmatrix} x'' \\ y'' \\ z'' \end{pmatrix} \equiv \mathbf{R}_y(\theta) \begin{pmatrix} x'' \\ y'' \\ z'' \end{pmatrix}. \quad (10.14)$$

Combining these two rotations then gives

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \cos \theta \cos \psi & -\sin \psi & \sin \theta \cos \psi \\ \cos \theta \sin \psi & \cos \psi & \sin \theta \sin \psi \\ -\sin \theta & 0 & \cos \theta \end{pmatrix} \begin{pmatrix} x'' \\ y'' \\ z'' \end{pmatrix}. \quad (10.15)$$

The resulting rotation matrix is defined as

$$\mathbf{R}_{zy}(\theta, \psi) = \mathbf{R}_z(\psi) \mathbf{R}_y(\theta) = \begin{pmatrix} \cos \theta \cos \psi & -\sin \psi & \sin \theta \cos \psi \\ \cos \theta \sin \psi & \cos \psi & \sin \theta \sin \psi \\ -\sin \theta & 0 & \cos \theta \end{pmatrix}. \quad (10.16)$$

Rotation matrices have several interesting properties such as the fact that their determinants are always ± 1 . More useful is the fact that $\mathbf{R}^{-1} = \mathbf{R}^T$, i.e., the inverse is the transpose of the matrix.⁵

Now consider a particle moving in direction $\mathbf{\Omega}(u, v, w) = \mathbf{i} \sin \theta \cos \psi + \mathbf{j} \sin \theta \sin \psi + \mathbf{k} \cos \theta \equiv (\sin \theta \cos \psi, \sin \theta \sin \psi, \cos \theta)$ that is scattered through an angle θ_s with a change in azimuth of $\Delta \psi_s$. How does one calculate $\mathbf{\Omega}'(u', v', w')$? First, rotate $\mathbf{\Omega}$ in the laboratory system to a local coordinate system in which the vector lies along the z -axis. It is readily verified that $\mathbf{\Omega}(\theta, \psi) = \mathbf{R}_{zy}(\theta, \psi)(0, 0, 1)$, so it follows that $\mathbf{R}_{zy}^{-1}(\theta, \psi)\mathbf{\Omega} = (0, 0, 1)$.

Next, the transformation $\mathbf{R}_{zy}(\theta_s, \Delta \psi_s)(0, 0, 1)$ produces a vector, in the local coordinate system, with the direction cosines given by the scattering angles, namely, $(\sin \theta_s \cos \Delta \psi_s, \sin \theta_s \sin \Delta \psi_s, \cos \theta_s)$. Finally, this scattered vector is rotated back to the laboratory system with the rotation matrix $\mathbf{R}_{zy}(\theta, \psi)$. Combining these three steps, the direction of the scattered particle is given by

$$\mathbf{\Omega}'(u', v', w') = \mathbf{R}_{zy}(\theta, \psi) \mathbf{R}_{zy}(\theta_s, \Delta \psi_s) \mathbf{R}_{zy}^{-1}(\theta, \psi) \mathbf{\Omega}. \quad (10.17)$$

To relate the new direction cosines (u', v', w') to the before-scattered direction cosines (u, v, w) , the rotation matrix of Eq. (10.16) can be written as

$$\mathbf{R}_{zy}(\theta, \psi) = \begin{pmatrix} w \cos \psi & -\sin \psi & u \\ w \sin \psi & \cos \psi & v \\ -\sin \theta & 0 & w \end{pmatrix}. \quad (10.18)$$

⁵ Such matrices are often called *orthogonal* matrices.

With this result and the relation $\mathbf{R}_{zy}^{-1}(\theta, \psi)\mathbf{\Omega} = (0, 0, 1)$, Eq. (10.17) gives the following explicit expressions:

$$u' = u \cos \theta_s + \sin \theta_s (w \cos \psi \cos \Delta \psi_s - \sin \psi \sin \Delta \psi_s), \quad (10.19)$$

$$v' = v \cos \theta_s + \sin \theta_s (w \sin \psi \cos \Delta \psi_s + \cos \psi \sin \Delta \psi_s), \quad (10.20)$$

$$w' = w \cos \theta_s - \sin \theta \sin \theta_s \cos \Delta \psi_s. \quad (10.21)$$

Because $\sin \theta = \sqrt{1 - w^2}$, $u = \sin \theta \cos \psi$, and $v = \sin \theta \sin \psi$, these results can be written in an alternate form that minimizes trigonometric function calls, namely,

$$u' = u\mu_s + \sqrt{1 - \mu_s^2} (wu \cos \Delta \psi_s - v \sin \Delta \psi_s) / \sqrt{1 - w^2}, \quad (10.22)$$

$$v' = v\mu_s + \sqrt{1 - \mu_s^2} (wv \cos \Delta \psi_s + u \sin \Delta \psi_s) / \sqrt{1 - w^2}, \quad (10.23)$$

$$w' = w\mu_s - \sqrt{1 - w^2} \sqrt{1 - \mu_s^2} \cos \Delta \psi_s, \quad (10.24)$$

where $\mu_s \equiv \cos \theta_s$.

From Eq. (10.21), explicit expressions for the polar and azimuthal angles after scatter can be obtained from

$$\cos \theta' = \cos \theta_s \cos \theta - \sin \theta_s \sin \theta \cos \Delta \psi_s \quad (10.25)$$

and

$$\psi' = \psi + \cos^{-1} \left[\frac{\cos \theta \cos \theta_s - \cos \theta'}{\sin \theta \sin \theta_s} \right]. \quad (10.26)$$

The above results for calculating (u', v', w') are sufficient for most cases. However, if $\mathbf{\Omega}(u, v, w)$ is almost parallel to the z -axis, then $w \simeq 1$ and the $1/\sqrt{1 - w^2}$ term becomes very large and leads to numerical errors. For the case that, say, $w < 0.9$, an alternative approach is to use a rotation matrix that first rotates about the x -axis followed by a rotation about the y -axis to create a local coordinate system in which $\mathbf{\Omega}$ is pointed along the y -axis. In this case, the rotation matrix is

$$\mathbf{R}_{yx}(\theta, \psi) = \begin{pmatrix} \cos \psi & \sin \theta \sin \psi & \cos \theta \sin \psi \\ 0 & \cos \theta & -\sin \theta \\ -\sin \psi & \sin \theta \cos \psi & \cos \theta \cos \psi \end{pmatrix}. \quad (10.27)$$

Now $\mathbf{R}_{yx}^{-1}(\theta, \psi)\mathbf{\Omega} = (0, 1, 0)$, and the direction vector, as above, is given by

$$\mathbf{\Omega}' = \mathbf{R}_{yx}(\theta, \psi) \mathbf{R}_{yx}(\theta_s, \Delta \psi_s) \mathbf{R}_{yx}^{-1}(\theta, \psi) \mathbf{\Omega}. \quad (10.28)$$

Explicitly, it is found that, with $\mu_s = \cos \theta_s$,

$$u' = u\mu_s + \sqrt{1 - \mu_s^2}(vu \cos \Delta\psi - w \sin \Delta\psi_s)/\sqrt{1 - v^2}, \quad (10.29)$$

$$v' = v\mu_s - \sqrt{1 - v^2}\sqrt{1 - \mu_s^2} \cos \Delta\psi_s, \quad (10.30)$$

$$w' = w\mu_s + \sqrt{1 - \mu_s^2}(wv \cos \Delta\psi - u \sin \Delta\psi_s)/\sqrt{1 - v^2}. \quad (10.31)$$

In the sections that follow specific approaches for simulating neutron and photon scattering are detailed. For sampling from other reactions that produce secondary particles (e.g., fission, $(n, 2n)$, fluorescence, etc.), the reader is referred to the MCNP manual [MCNP 2003].

10.6.2 Photon Scattering from a Free Electron

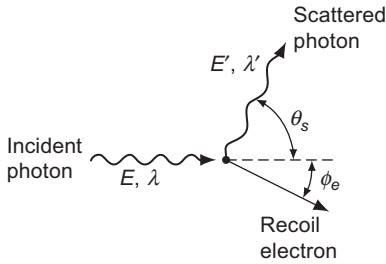


Figure 10.10 A photon with energy E and wavelength λ is scattered by a free electron. After scattering, the photon has a reduced energy E' and larger wavelength λ' .

In Compton scattering, a photon of energy E and dimensionless wavelength $\lambda = m_e c^2 / E$ scatters from a free electron as shown in Fig. 10.10. Here, $m_e c^2$ is the rest mass energy equivalent of the electron, namely, 0.511 MeV. The differential scattering cross section is given by the Klein-Nishina formula [Shultis and Faw 2007]. The PDF to sample from this cross-section formula can be written in terms of the ratio $r \equiv \lambda' / \lambda = E / E'$ as

$$f(r|\lambda) = \frac{K}{r^2} \left[\frac{1}{r} + r - 1 + (1 + \lambda - r\lambda)^2 \right], \quad (10.32)$$

where K is the normalization constant to make $f(r|\lambda)$ a PDF. The cosine of the scattering angle is given by

$$\cos \theta_s = 1 + \lambda - r\lambda. \quad (10.33)$$

The range of the scattering angle $0 \leq \theta_s \leq \pi$ thus gives the constraint $1 \leq r \leq 1 + (2/\lambda)$. This PDF can be separated into a sum of two PDFs amenable to sampling by the composition-rejection method (see Section 4.1.7) as follows [Kahn 1956]:

$$\begin{aligned} f(r|\lambda) &= \frac{1}{K} \left\{ \left(\frac{1}{r} - \frac{1}{r^2} \right) + \frac{1}{r^2} \left(\frac{1}{r} + [1 + \lambda - r\lambda]^2 \right) \right\} \\ &= A_1 g_1(r) f_1(r) + A_2 g_2(r) f_2(r). \end{aligned} \quad (10.34)$$

The constants $A_1 \equiv K/(2\lambda)$ and $A_2 \equiv 4K/[\lambda + 2]$ have the properties

$$\frac{A_1}{A_1 + A_2} = \frac{\lambda + 2}{2 + 9\lambda} \quad \text{and} \quad \frac{A_2}{A_1 + A_2} = \frac{8\lambda}{2 + 9\lambda}. \quad (10.35)$$

The functions f_1 and f_2 are defined as

$$f_1(r) \equiv 4 \left(\frac{1}{r} - \frac{1}{r^2} \right) \quad \text{and} \quad f_2(r) \equiv \frac{1}{2} \left[\frac{1}{r} + (1 + \lambda - r\lambda)^2 \right] \quad (10.36)$$

and have the property that $f_i(r) \leq 1$ for $1 \leq r \leq 1 + (2/\lambda)$. Finally, the PDFs $g_1(r) \equiv \lambda/2$ and $g_2(r) \equiv (\lambda + 2)/(2r^2)$ have the easily invertible CDFs

$$G_1(r) = \frac{\lambda}{2}(r - 1) \quad \text{and} \quad G_2(r) = \frac{\lambda + 2}{2} \left(1 - \frac{1}{r} \right). \quad (10.37)$$

The algorithm for sampling from Eq. (10.34) is as described in Section 4.1.7, namely,

1. If $\rho_i < (\lambda + 2)/(2 + 9\lambda)$, calculate $r = 1 + (2\rho_{i+1} + 1)/\lambda$ and accept this value if $\rho_{i+2} < (4/r)[1 - (1/r)]$.
2. Otherwise, if $\rho_i \geq (\lambda + 2)/(2 + 9\lambda)$, calculate $r = (\lambda + 2)/(2\rho_{i+1} + \lambda)$ and accept this value if $\rho_{i+2} < [(1/r) + (1 + \lambda - r\lambda)^2]/2$.
3. If no acceptable value for r is found, return to step 1 and try again.

Once a value of r is accepted, the wavelength of the scattered photon is simply $\lambda' = r\lambda$ corresponding to an energy of $E' = E/r$. Lastly, $\cos \theta_s$ is found from Eq. (10.33) and $\sin \theta_s = \sqrt{1 - \cos^2 \theta_s}$. Comparison of this and other sampling methods for the Klein-Nishina cross section are given by Blomquist and Gelbard [1983].

10.6.3 Neutron Scattering

Neutron scattering data are almost always tabulated in the center-of-mass coordinate system [Shultis and Faw 2007] in which there is zero net linear momentum before and after the scatter. The microscopic differential cross section for either elastic or inelastic scattering is most often approximated as

$$\sigma_s(E, \omega_c) \simeq \frac{\sigma_s(E)}{4\pi} \sum_{n=0}^N (2n + 1) f_n(E) P_n(\omega_c) \quad (10.38)$$

in which $P_n(\omega_c)$ is the Legendre polynomial of order n and ω_c is the cosine of the scattering angle in the center-of-mass system. Cross-section data consists of tabulating the Legendre coefficients $f_n(E)$ (with f_0 always equal to 1). The reason for this representation is that, in the center-of-mass system, scattering is usually more isotropic

than in the laboratory system (target at rest system) so that only a low-order expansion is needed. Indeed, for isotropic scattering $N = 0$, and most data libraries seldom exceed $N = 8$ even for highly anisotropic scattering.⁶

The first step in simulating a neutron scattering event is to construct a PDF with the same shape as Eq. (10.38) and to obtain a sample of ω_c from it. The corresponding cosine of the scattering angle in the laboratory system is [Shultis and Faw 2007]

$$\boldsymbol{\Omega} \cdot \boldsymbol{\Omega}' \equiv \cos \theta_s = \frac{\gamma + \omega_c}{\sqrt{1 + 2\gamma\omega_c + \gamma^2}}, \quad (10.39)$$

where

$$\gamma = \left[A^2 + \frac{A(A+1)Q}{E} \right], \quad (10.40)$$

with Q being the Q-value [Shultis and Faw 2007] of the scatter ($= 0$ for elastic scatter and < 0 for inelastic scatter) and A being the ratio of the atom's mass to that of the neutron ($\simeq$ atomic mass number). From conservation of energy and momentum, the energy of the scattered neutrons is given by [Shultis and Faw 2007]

$$E' = \frac{1}{2}E(1 + \alpha) + \frac{1}{2}(1 - \alpha)E\omega_c\sqrt{1 + \Delta} + \frac{QA}{A + 1} \quad (10.41)$$

in which $\alpha \equiv (A - 1)^2/(A + 1)^2$ and $\Delta \equiv Q(A + 1)/(AE)$.

10.7 Time Dependence

Although most transport problems are independent of time, Monte Carlo simulations are easily extended to time-dependent cases when needed, for example, to estimate when a particle reaches some surface. For neutrons, the energy determines the neutron's speed from which the time for various segments of its history are easily calculated. Tallies are then binned by the time the contribution to the tally was made. Photons are even easier because they all travel at the speed of light. Even time-varying sources can be treated by sampling from an appropriate time distribution to obtain the starting time for each history and tracking the times of the various events that occur

⁶ For the case of isotropic scattering (often a good approximation for heavier atoms since the center-of-mass system becomes the target-at-rest system for infinitely heavy scattering centers), the direction cosines of the scattered particle can be calculated in the same manner as used for the direction cosines of an isotropically emitted source particle (see Section 10.3.1).

during the history. In the following discussion, however, it is assumed that the tallies and flux densities are independent of time.

10.8 Particle Weights

The efficiency of a Monte Carlo transport calculation depends both on the speed with which the calculation can be made and the variance of the result. The speed is, to a large extent, controlled by the hardware and, to a lesser extent, by the computer programmer and the operating system. The variance is controlled primarily by the type of tally used and the number of particles that contribute to the tally. In a strict analog simulation, the only way to reduce the variance of a tally is to run more histories. But this brute-force approach only reduces the variance as $1/N$. As explained in Chapter 5, a better way to increase the precision of the tally is to use nonanalog techniques to force more particles to the regions of phase space where particles are more likely to score without increasing (and perhaps decreasing) the sampling in less important regions of phase space.

In nonanalog simulation, as explained in Section 4.2.3, samplings for the various physical events used to create a particle history are biased so that events more likely to contribute to the tally are produced more frequently than in an analog simulation. This bias introduced into particle histories must, of course, be corrected when scoring the histories so as to produce an unbiased estimator. For example, if a particular bias makes a history α times more likely to score, then its contribution to the tally must be multiplied by a factor of $1/\alpha$. This removal of biases is readily accomplished by assigning a *weight* W to each particle, and, at every event in its history that introduces a bias, its weight is adjusted so that its score *times* its weight produces an unbiased contribution to the score. For example, if, in a biased sampling for some event, the particle's chance of scoring is changed by a factor of α , its current weight W is modified to W/α . If there is no bias in starting a particle history, the history begins with a weight of unity. Note that weights are not constrained to integer values and may be greater or less than unity.

In short, in a nonanalog Monte Carlo simulation, the scoring of the particles reaching the tally region or detector must be modified from just the sum of the particles' scores to the sum of the particles' weighted scores. In this way, unbiased results can be achieved with less computing time, if proper variance reduction methods are used, compared to an analog calculation producing the same results and precision.

10.9 Scoring and Tallies

After a particle history ends or, more usually, after a particle leaves a cell, its contribution to the score (or tally) is added to the running score of interest. Almost anything of interest that depends on the radiation field can be estimated from the particle histories used in a Monte Carlo simulation. Here, some of the more common tallies are reviewed. For more thorough treatments and other estimators, the reader is referred to Kalos et al. [1968], Carter and Cashwell [1975], and MCNP [2003].

10.9.1 Fluence Averaged Over a Surface

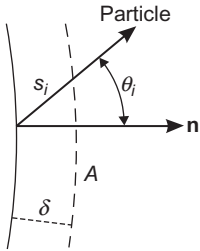


Figure 10.11 Particle crossing a surface at an angle θ_i from the outward normal $\mathbf{n}$.

Often the fluence, averaged over some surface (or portion of a surface) is sought. Imagine a parallel surface a very small distance δ from the surface of interest as shown in Fig. 10.11. From Eq. (9.13), the fluence (flux density integrated over all time or all simulated histories) is just the weighted sum of the path lengths of all the particles passing through this incremental volume $\Delta V = A\delta$ divided by the volume of the extended region, i.e., $\Phi = \sum_i W_i s_i / \Delta V$, where W_i is the weight of particle i . The path length for the particle shown is $s_i = \delta / |\cos \theta_i|$, where θ_i is the angle between the particle's exit direction and the outward normal $\mathbf{n}$. Thus, the fluence contribution of the i th particle crossing the surface is $\Phi_i = \lim_{\delta \rightarrow 0} (W_i \delta / |\cos \theta_i|) / A\delta = W_i / (A |\cos \theta_i|) = W_i / (A |\mathbf{\Omega}_i \cdot \mathbf{n}|)$. Every time a particle crosses the surface, the value of its $W_i / |\cos \theta_i|$ is added to the tally. Of

course, many histories may not cross the surface in question and so no score is contributed to the tally. Finally, after N histories the average fluence, *per source particle*, is estimated as

$$\bar{\Phi}_S \equiv \frac{1}{A} \int_A dA \int_0^\infty dE \int_{4\pi} d\mathbf{\Omega} \Phi(\mathbf{r}, E, \mathbf{\Omega}) \simeq \frac{1}{NA} \sum_{i=1}^N \sum_{j=1}^{n_i} \frac{W_i^j}{|\cos \theta_i^j|}, \quad (10.42)$$

where n_i is the number of surface crossings made by the i th particle,⁷ W_i^j the i th particle's weight on its j th crossing, and θ_i^j is the angle with respect to the outward normal at the j th crossing. Here, $\Phi(\mathbf{r}, E, \mathbf{\Omega})$ is the energy and angular fluence normalized to one source particle.

It should be noted that the variance of this fluence estimator is infinite and, hence, the central limit theorem cannot be used to estimate confidence intervals. If many particles cross the surface in nearly tangential directions, i.e., $\cos \theta$ is very small, the tally contributions become very large. To avoid such large scoring events, many Monte Carlo transport codes use an ad hoc fix-up procedure to avoid large variances. If $|\cos \theta| < \epsilon$, where ϵ is some small value, say 0.01, then replace the $|1/\cos \theta_i^j|$ in Eq. (10.42) by $2/\epsilon$, the expected value of $1/|\cos \theta|$ when $|\cos \theta| < \epsilon$ and the fluence is nearly isotropic. Although this procedure is not rigorous, in most cases, reliable estimates of the surface averaged fluence are obtained.

If the angular distribution and energy spectra of the particles crossing the surface are of interest, then the tally procedure requires the use of a multidimensional array for scoring. Polar angle cosines, $\cos \theta$, can be divided into some number of intervals from -1 to 1 . Likewise, deviations in azimuthal angle at the plane surface, as measured from some azimuthal angle of incidence, can be divided into some number of intervals from 0 to 2π radians. The full range of possible particle energies may be divided

⁷ Typically, n_i is 0 or 1, but a particle may make several crossings, especially if the surface is reentrant.

likewise into some number of intervals. When a particle is found to cross the surface, a score is made in the *bin* associated with the energy and direction of the particle. Binning by energy also allows each bin to be multiplied by some energy-dependent response function $\mathcal{R}(E)$, so that, summing over all energy bins, the average dose at the surface can be estimated. Likewise, multiplication by the j th reaction interaction coefficient $\mu_j(E)$ and summation of the results yields the average reaction density $\int \mu_j(E)\Phi(E) dE$ at the surface.

10.9.2 Average Fluence in a Volume: Path-Length Estimator

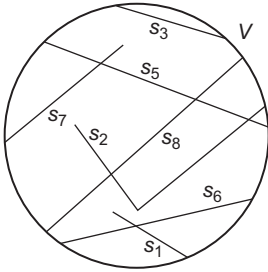


Figure 10.12 Particle paths in cell of volume V .

The idea behind Eq. (9.13) can also be extended to determine the average fluence in the volume of a cell. In this type of tally, the sum of path lengths s_i each particle makes in the volume V of interest is accumulated (see Fig. 10.12). Again, if a particle history never reaches the cell in question, then $s_i = 0$. The average fluence, per source particle, is then estimated as

$$\bar{\Phi}_V \simeq \frac{1}{NV} \sum_{i=1}^N \sum_{j=1}^{n_i} W_i^j s_i^j, \quad (10.43)$$

where n_i is the number of times the i th particle enters V , s_i^j is the j th track length in V , and W_i^j is the particle's weight when entering V for the j th time. By breaking the energy range into a contiguous set of bins and adding each weighted path length to the tally in the appropriate energy bin, the energy spectrum of the volume-average fluence can be estimated as

$$\bar{\Phi}_V(E) \equiv \frac{1}{V} \int_V dV \int_{4\pi} d\Omega \Phi(\mathbf{r}, E, \Omega) \simeq \frac{S}{NV\Delta E}, \quad (10.44)$$

where ΔE is the width of the energy bin containing E and S is the weighted sum of path lengths tallied in that energy bin. As before the tally for each particle can be multiplied by some function $\mathcal{R}(E)$ so that when the results in each energy bin are summed, the average value of $\int dE \mathcal{R}(E) \bar{\Phi}_V(E)$ is estimated.

If the volume V is large so that many histories enter it, the path-length estimator is generally a good tally. If the region is a thin curved shell, most track lengths have similar lengths and the estimator has a small variance. By contrast, for a region bounded by two closely spaced parallel planes, there may be a wide variation in the track lengths through the region, and the resulting path-length estimator may have a large variance. The path-length estimator is also computationally quite efficient, because particle tracking already computed the track lengths in the regions, so little extra effort is needed for this tally.

10.9.3 Average Fluence in a Volume: Reaction-Density Estimator

An alternative estimator for the average fluence in a region is obtained by tallying the number of interactions that occur in the region. From Eq. (9.37), the average density of reactions in a homogeneous volume V is found to be

$$\bar{R}(E) \equiv \frac{1}{V} \int_V \mu_t(\mathbf{r}, E) \Phi(\mathbf{r}, E) dV = \frac{\mu_t(E)}{V} \int_V \Phi(\mathbf{r}, E) dV \equiv \mu_t(E) \bar{\Phi}_V(E). \quad (10.45)$$

Thus, the contribution to the average fluence in V made by a particle of which the energy E lies in ΔE and that collides in V is made by adding $1/\mu_t(E)$ to the tally energy bin. Thus, after N histories, the estimate of the average fluence is

$$\bar{\Phi}_V(E) \equiv \frac{1}{V} \int_V dV \int_{4\pi} d\Omega \Phi(\mathbf{r}, E, \Omega) \simeq \frac{1}{NV\Delta E} \sum_{i=1}^N \sum_{j=1}^{n_i} \frac{W_i^j}{\mu_t(E)}, \quad (10.46)$$

where n_i is the number of collisions in V made by particle i when it had an energy in ΔE , and the superscript j refers the j th such interaction in V made by the i th particle.

The probability a particle interacts while crossing a cell is

$$p = 1 - \exp[-\mu_t(E)s_{\max}],$$

where $s_{\max}$ is the maximum distance the particle could travel in the cell along its flight path. For thin or small regions, this probability is small and the above estimator will produce poor results. A more efficient way to estimate the fluence is to tally the interaction contribution times the probability the particle interacts, i.e.,

$$\bar{\Phi}_V(E) \simeq \frac{1}{NV\Delta E} \sum_{i=1}^N \sum_{j=1}^{n_i} W_i^j \frac{1 - e^{-\mu_t(E)s_{\max,j}}}{\mu_t(E)}, \quad (10.47)$$

where n_i now is the number of cell crossings made by particle i when it had energy E in ΔE .

For the reaction-density estimator of Eq. (10.46) to work properly, particles must cross the cell of interest. If the cell is far from the source, even if a particle heads toward the cell, it is unlikely to reach it and contribute to the tally. But all is not lost. The contribution to the fluence in the cell, however, can be calculated for particles heading toward the cell but that interact at some point before reaching it. Denote the distance from the particle interaction point to the nearest cell surface by d along the original trajectory. One simply includes in the estimator of Eq. (10.46) a term that gives the probability p_{hit} that the particle *does* reach the cell,

$$p_{hit} = \exp \left[- \int_0^d \mu_t(\mathbf{r}, E) ds \right],$$

where the integration is along the trajectory. Thus, the argument of the exponential is just the number of mean-free-path lengths the particle would have to travel from its interaction point to reach the cell. The average fluence in the cell is then estimated as

$$\bar{\Phi}_V(E) \simeq \frac{1}{NV\Delta E} \sum_{i=1}^N \sum_{j=1}^{n_i} W_i^j \frac{1 - e^{-\mu_t(E)s_{\max,j}}}{\mu_t(E)} \exp \left[-\int_0^{d_j} \mu_t(\mathbf{r}, E) ds \right]. \quad (10.48)$$

Here, n_i is the number of collisions, both inside and outside the cell, made by particle i . If a collision is inside the cell, then, of course, $d_j = 0$.

For a large thin region, subtending a large solid angle, Eq. (10.48) may be a good estimator, but for small distant regions, few particles will even interact while heading toward it. Even this can be overcome by including the probability that, if the collision were a scatter, the particle would have an energy in ΔE and head toward the cell. This idea of forcing a scatter toward a cell of interest is developed further in Section 10.9.5.

10.9.4 Average Current Through a Surface

Sometimes the average total flow or current through a surface is sought (see Section 9.1.5). For particles with energy E in ΔE and direction $\boldsymbol{\Omega}$ in $\Delta\Omega$, the estimate of the average current, per source particle, is⁸

$$\bar{J}(E, \boldsymbol{\Omega}) \equiv \frac{1}{A} \int_A dA \int_{\Delta\Omega} d\boldsymbol{\Omega} |\boldsymbol{\Omega} \cdot \mathbf{n}| \Phi(\mathbf{r}, E, \boldsymbol{\Omega}) \simeq \frac{1}{NA \Delta E \Delta\Omega} \sum_{i=1}^N n_i W_i, \quad (10.49)$$

where n_i is the number of crossings of the surface made by particle i when it had energy in ΔE and direction in $\Delta\Omega$. Typically, n_i is 0 or 1, but a particle may cross a reentrant surface several times.

If the net flows across a surface are wanted (see Eq. (9.20)), then use this procedure with only the two cosine ($\mu \equiv \mathbf{n} \cdot \boldsymbol{\Omega}$) angular bins $-1 \leq \mu < 0$ and $0 < \mu \leq 1$ to estimate $J^-(E_i)$ and $J^+(E_i)$, respectively. The net flow is $J_{\text{net}} = J^+ - J^-$.

10.9.5 Fluence at a Point: Next-Event Estimator

A very powerful technique for scoring is to combine deterministic tally contributions and the stochastic collisions that occur during a history. One such estimator is the *next-event estimator*, also known as the no-further-collision estimator. Consider the small spherical detector, with cross-sectional area dA , shown in Fig. 10.13. A particle traveling in direction $\boldsymbol{\Omega}'$ has a collision at $\mathbf{r}'$. The collision may not be a scatter, and even if it were a scatter, it is very unlikely to scatter in the direction of dV or even reach it to contribute to the fluence tally for dV . However, one can analytically calculate the probability the particle will scatter at $\mathbf{r}'$ and reach the detector without further

⁸ Here, J is the time-integrated current vector discussed in Section 9.1.5 just as the fluence Φ is the time-integrated flux ϕ (see Eq. (9.15)).

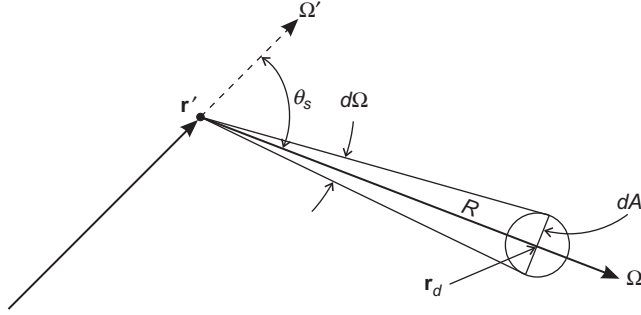


Figure 10.13 A particle moving in direction Ω' makes a collision at $\mathbf{r}'$ and scatters toward a small spherical detector at $\mathbf{r}_d$ of cross-sectional area dA and a distance $R = |\mathbf{r}' - \mathbf{r}|$ from the scattering point.

interaction, and thus provide a contribution to the fluence tally for the detector, thereby shortcutting the Monte Carlo process.

The probability the particle with energy E scatters at $\mathbf{r}'$ through a scattering angle θ_s and has a new direction of travel in $d\Omega$ about Ω is

$$p_1 = \frac{\mu_s(\mathbf{r}', E, \omega_s) d\Omega}{\int_0^{2\pi} d\psi \int_{-1}^1 d\omega_s \mu_s(\mathbf{r}', E, \omega_s)} = \frac{1}{2\pi} \left[\frac{\mu_s(\mathbf{r}', E, \omega_s)}{\int_{-1}^1 d\omega_s \mu_s(\mathbf{r}', E, \omega_s)} \right] d\Omega = \frac{p(\omega_s)}{2\pi} d\Omega,$$

where $p(\omega_s)$ is the PDF (defined by the terms in the square brackets) for scattering through an angle $\theta_s = \cos^{-1} \Omega \cdot \Omega'$. The distance between $\mathbf{r}'$ and $\mathbf{r}_d$ is $R = |\mathbf{r}' - \mathbf{r}_d|$ so that $d\Omega = dA/R^2$. The probability that such a scattered particle actually reaches the detector without further collision is

$$p_2 = \exp \left[- \int_0^R \mu_t(\mathbf{r}, E) ds \right],$$

where the argument of the exponential is the total number of mean-free-path lengths between $\mathbf{r}'$ and $\mathbf{r}_d$.

As stated by Eq. (9.14), the flux density can be interpreted as the number of particles entering a spherical detector of cross-sectional area dA divided by dA . Thus, the contribution to the fluence made by the particle with weight W that interacts at $\mathbf{r}'$ is

$$\delta\Phi = W \frac{p_1 p_2}{dA} = W \frac{p(\omega_s)}{2\pi R^2} \exp \left[- \int_0^R \mu_t(\mathbf{r}, E) ds \right]. \quad (10.50)$$

This result is independent of dA and, thus, is used as a fluence tally for a point detector at $\mathbf{r}_d$. After each interaction in a particle's history, an estimate is made deterministically for the expected contribution of that interaction to the fluence at point $\mathbf{r}_d$.

A difficulty with this point-detector tally occurs when the tally point lies within a scattering medium. Because of the $1/R^2$ term in Eq. (10.50) an enormous contribution

is made to the tally if an interaction occurs very near the tally point. In fact, in a scattering region the variance of this estimator is infinite! An infinite variance, however, does not mean the tally cannot be used; indeed, the tally will converge, but the asymptotic behavior is reduced to $1/\sqrt[3]{N}$, slower than the $1/\sqrt{N}$ convergence for a tally with a finite variance [Kalos et al. 1968]. Of course, if the tally point is in a vacuum or nonscattering medium no such problem occurs.

The way to avoid an infinite variance in a scattering medium is to surround the tally point by a small sphere of radius R_o , and do no estimation if an interaction occurs at $R < R_o$. Alternatively, if an interaction occurs with $R < R_o$, tally the average fluence uniformly distributed in the volume, i.e.,

$$\begin{aligned}\delta\Phi(R < R_o) &= \frac{\int \Phi dV}{\int dV} = W \frac{p(\omega_s)}{2\pi} \frac{\int_0^{R_o} \frac{1}{r^2} e^{-\mu_t(E)r} (4\pi r^2) dr}{\frac{4}{3}\pi R_o^3} \\ &= W p(\omega_s) \frac{3(1 - e^{-\mu_t R_o})}{2\pi R_o^3}.\end{aligned}\quad (10.51)$$

Although this next-event estimator can produce a good prediction of the fluence at a point, its convergence may be rather slow. This estimator needs many statistical tests to ensure that sound results are achieved. Lux and Koblinger [1991] and MCNP [2003] discuss variance issues associated with the use of next-event estimators in the estimation of point values of fluence and related quantities.

10.9.6 Flow Through a Surface: Leakage Estimator

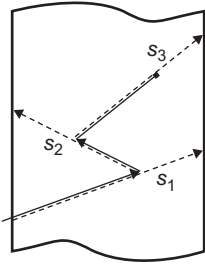


Figure 10.14 A three-segment particle path ending in an absorption.

Another partially deterministic method is the *leakage estimator* used to find the probability P_{esc} that a particle entering a region leaves through some surface of the region. The idea is very similar to the next-event estimator. At the point of entry of a particle into the region (either entering from outside or born in the region), and after each scattering interaction in the region, a determination is made of the probability that, without further interaction, the photon can escape from the slab. That probability is then scored as a contribution to the flow across the surfaces. Thus, for the three-segment path ending in an absorption shown in Fig. 10.14, the leakage tally for this particle history is

$$\delta P_{esc} = e^{-\mu s_1} + e^{-\mu s_2} + e^{-\mu s_3}.$$

Also note that if the particle passes through the surface on its last segment in the region no score is tallied. This method can also be applied to scoring the energy spectrum and angular distribution of the flow rates through each slab face. For deep penetration problems, score contributions may vary considerably among histories, so that large variances result. An example of an analog-leakage calculation and one using the leakage estimator is given in Example 10.2.

Example 10.2 Escape Probability from an Infinite Slab

Consider an infinite homogeneous slab surrounded by a vacuum. One-speed particles are born uniformly throughout the slab. The particles scatter isotropically with $c = \mu_s/\mu_t$, the scattering probability per interaction. The problem is to calculate the probability P_{esc} that a particle born in the slab eventually escapes or leaks from it.

In a straight analog simulation of N histories in which n_{esc} particles leak from the slab,

$$P_{esc} = \frac{n_{esc}}{N}.$$

However, for thick slabs with small values of c only particles born near the surface have much chance of escape and of contributing to the tally. With a leakage estimator every particle contributes something to the tally, although many contributions are quite small. In Fig. 10.15, results are shown for the case $c = 0.1$ using analog scoring (left figure) and the leakage estimator (right figure). It is seen that, although both methods give comparable results for thin slabs, the leakage estimator is better for the thick slabs.

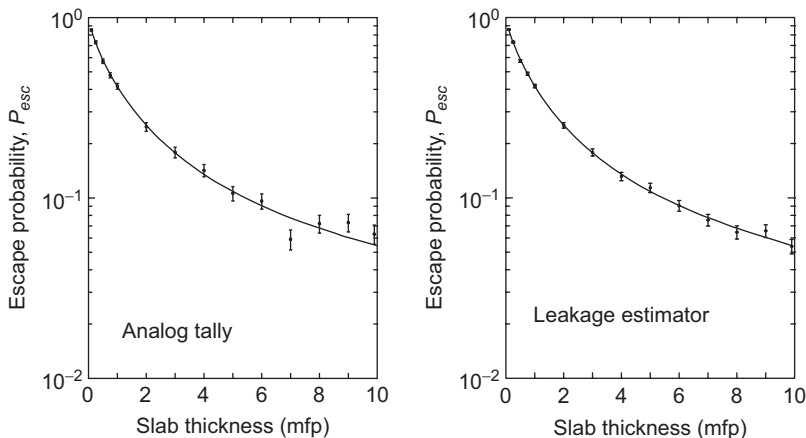
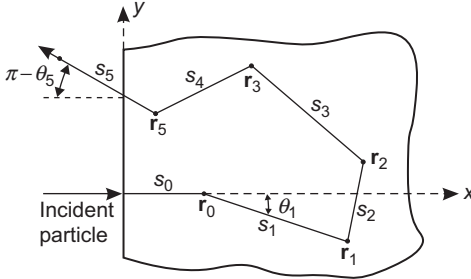


Figure 10.15 Escape probability for particles born uniformly in an infinite homogeneous slab with $c = 0.1$. Monte Carlo results with $\pm\sigma$ error bars are for $N = 1,000$ particle histories. Results on the left are for an analog tally, i.e., $P_{esc} = (\text{no. escaping})/(\text{no. born})$. Results on the right were obtained using the leakage estimator. The solid lines on these plots are cubic spline fits to results obtained with 10^7 histories and that contain negligible error.

10.10 An Example of One-Speed Particle Transport

Before venturing into more detailed aspects of Monte Carlo transport theory, it might be useful to consider a simple example that uses many of the concepts introduced so far. In Example 10.3, a homogeneous half-space is normally and uniformly illuminated by incident particles. The problem is to find the distribution of reflected particles.

Example 10.3 Particle Reflection from a Half-Space



The challenge for solving the transport equation occurs when particles can scatter as well as be absorbed. Analytic solutions are known only for highly simplified forms of the transport equation of Eq. (9.49), and then only with substantial mathematical effort. To allow a comparison between a Monte Carlo calculation and an exact analytical solution, the following simplifications are applied to Eq. (9.49): (1) particles

are assumed to have only a single energy, (2) scattering is isotropic, (3) the medium is a source-free homogeneous half-space $x \geq 0$, and (4) particles are incident uniformly and normally on the surface of the half-space. Under these conditions, Eq. (9.49) can be written as

$$\omega \frac{\partial \phi(x, \omega)}{\partial x} + \phi(x, \omega) = \frac{c}{2} \int_{-1}^1 \phi(x, \omega') d\omega', \quad 0 \leq x < \infty, \quad -1 \leq \omega \leq 0, \quad (10.52)$$

where x is measured in mean-free-path lengths, $\omega = \mathbf{i} \cdot \mathbf{\Omega}$ and $c = \mu_s/\mu \leq 1$. The boundary conditions of this *albedo problem* are (1) $\phi(0, \omega) = \delta(\omega - 1)$ and (2) $\lim_{x \rightarrow \infty} \phi(x, \omega) < \infty$.⁹ In this formulation, the units of $\phi(x, \omega)$ are no longer $\text{cm}^{-2} \text{s}^{-1} \text{sr}^{-1}$ but $\text{cm}^{-2} \text{s}^{-1}$ per unit ω , because, to obtain Eq. (10.52), integration over all azimuthal angles was performed.

Several different (and complex) analytical methods can be used to obtain equivalent closed-form analytical expressions for the angular flux density $\phi(x, \omega)$. In particular, the reflected flux density is [Chandrasekhar 1960]

$$\phi(0, \omega) = \frac{c}{2} \frac{1}{1 + |\omega|} H(c, |\omega|) H(c, 1), \quad -1 \leq \omega \leq 0, \quad (10.53)$$

where $H(c, \cos \theta)$ is Chandrasekhar's H -function, which can be evaluated by an iterative solution of the nonlinear integral equation [Shultis and Faw 2007]

$$\frac{1}{H(c, \cos \theta)} = \sqrt{1 - c} + \frac{c}{2} \int_0^1 \frac{u H(c, u)}{u + \cos \theta} du. \quad (10.54)$$

Not only is the analysis to obtain Eq. (10.53) difficult, so is its numerical evaluation, which requires a numerical solution of Eq. (10.54). By contrast, the reflected flux density is easily calculated by a Monte Carlo simulation if the problem is first transformed to an easier equivalent one, as discussed in Section 5.1. Rather than having particles incident all over the infinite surface of the half-space and scoring only those that flow back through a given unit area on the surface after scattering in the half-space, it is far better to have all incident particles impinge on the same point on the surface (say, $y = z = 0$) and to score particles that emerge anywhere on the surface. Because the half-space is infinite in the y and z directions, these coordinates have no effect on whether a particle escapes, and, hence, only the x -coordinate of a particle's interaction site needs to be calculated.

⁹ This albedo problem is used, for example, to describe light reflected from a planetary atmosphere or thermal neutrons reflected from the surface of a shield.

The Monte Carlo simulation proceeds as follows. An incident particle penetrates a distance s_0 , randomly sampled from $f(s) = \mu e^{-\mu s}$, where it is scattered, with probability c , or absorbed, with probability $1 - c$. If a scatter occurs, a new angular direction θ_0 is selected, where $\cos \theta_0$ is sampled from $\mathcal{U}(-1, 1)$, and another path length is randomly chosen to the next interaction point $\mathbf{r}_1$. This procedure is continued, as shown in the figure on the previous page, until the particle is either absorbed or escapes through the surface. The escaping particle is then binned according to its angle of escape.

One possible algorithm for this albedo simulation is given below. Here, each ρ refers to a new random number from $\mathcal{U}(0, 1)$:

1. Generate the first collision site $x_0 = s_0 = -\ln \rho_0$.
2. If $\rho > c$ particle is absorbed, go to step 1; otherwise:
 - (a) Pick a new cosine direction $\omega_{i+1} = 2\rho - 1$.
 - (b) Pick the distance to the next interaction site as $s_{i+1} = -\ln \rho$.
 - (c) Calculate the x -coordinate of the interaction site as $x_{i+1} = x_i + \omega_{i+1}s_{i+1}$.
 - (d) If $x_{i+1} < 0$, the particle escapes. Bin the flux density estimator $1/\omega_{i+1}$ in the tally vector $g(\Delta\omega_i) = g(\Delta\omega_i) + 1/\omega_{i+1}$, where the cosine range for $\Delta\omega_i$ includes ω_{i+1} . It is time for another history, so go to step 1.
 - (e) If $x_{i+1} > 0$, the particle is still in the half-space. If $\rho < c$, the particle scatters, so go to step 2a; otherwise, the particle is absorbed, so go to step 1.

After a large number of histories N the estimate of the reflected angular flux density $\phi(0, \omega)$, $\omega \leq 0$ is simply $g(\Delta\omega_i)/N$. In Fig. 10.16, a comparison is shown between the analytical result of Eq. (10.53) and the above Monte Carlo procedure. To obtain such good agreement, 10^7 histories were used.

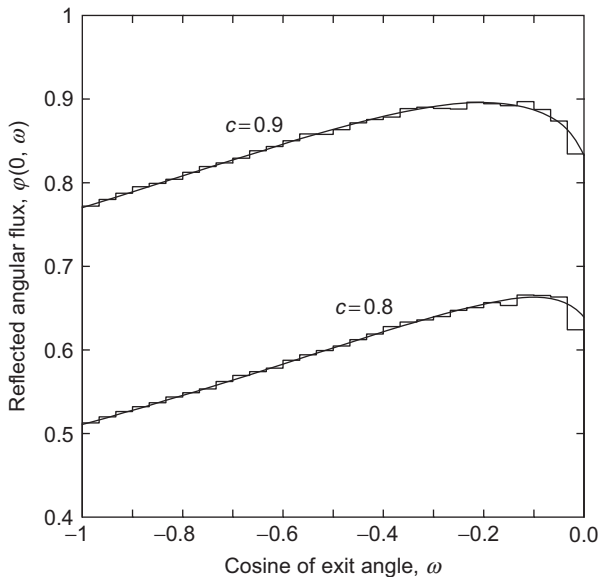


Figure 10.16 Analytical (smooth lines) and Monte Carlo (histograms) results for the albedo problem. The Monte Carlo results used 10^7 histories and 30 cosine bins.

10.11 Monte Carlo Based on the Integral Transport Equation

As discussed in earlier chapters, Monte Carlo is ideally suited for evaluating multidimensional integrals (see Section 2.6) and for solving integral equations (see Section 8.2.2). Thus, it is not surprising that early transport calculations used Monte Carlo techniques on the integral form of the transport equation (see Section 9.3.1).

10.11.1 Integral Transport Equation

One widely used way to develop Monte Carlo calculational procedures for particle transport is to begin with some form of the integral transport equation (see Section 9.3.1). It doesn't really matter which form of the integral equation is used, but here the procedure of [Carter and Cashwell \[1975\]](#) is used, which is based on the integral equation for the collision rate density $F(\mathbf{r}, E, \boldsymbol{\Omega})$ that is given by Eq. (9.73), namely,

$$F(\mathbf{P}) = \int_R d\mathbf{P}' \bar{K}(\mathbf{P}' \rightarrow \mathbf{P}) F(\mathbf{P}') + \bar{S}(\mathbf{P}). \quad (10.55)$$

Here, $\mathbf{P}$ are the particle phase space coordinates $(\mathbf{r}, E, \boldsymbol{\Omega})$, and region R of phase space contains the volume V of the ambient medium, the energy range of E , and the 4π steradians of directions. $\bar{S}(\mathbf{P})$ is the first-flight collision rate density of particles streaming from the source and is given by

$$\bar{S}(\mathbf{P}) \equiv \mu(\mathbf{r}, E) \int_V dV' \exp[-\tau(E, \mathbf{r}, \mathbf{r}')] \delta\left(\boldsymbol{\Omega} - \frac{\mathbf{r} - \mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|}\right) \frac{S(\mathbf{r}', E, \boldsymbol{\Omega})}{|\mathbf{r} - \mathbf{r}'|^2},$$

and $\bar{K}(\mathbf{P}' \rightarrow \mathbf{P})$ is the next-flight collision rate density at $\mathbf{P}$ due to a collision at $\mathbf{P}'$ and is given by

$$\bar{K}(\mathbf{P}' \rightarrow \mathbf{P}) \equiv \frac{\mu(\mathbf{r}, E)}{\mu(\mathbf{r}', E')} \frac{\mu(\mathbf{r}', E' \rightarrow E, \boldsymbol{\Omega}' \rightarrow \boldsymbol{\Omega})}{|\mathbf{r} - \mathbf{r}'|^2} \exp[-\tau(E, \mathbf{r}, \mathbf{r}')] \delta\left(\boldsymbol{\Omega} - \frac{\mathbf{r} - \mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|}\right).$$

Further, suppose the purpose of the Monte Carlo calculation is to estimate the value of some weighted average of $F(\mathbf{P})$, such as

$$G = \int d\mathbf{P} g(\mathbf{P}) F(\mathbf{P}), \quad (10.56)$$

where $g(\mathbf{P})$ is a prescribed weight function, such as a detector response function, that gives the contribution to G from a collision at $\mathbf{P}$.

Formal solution of the integral transport equation

The integral equation of Eq. (10.55) can be solved using the Neumann iteration scheme of Eq. (8.49), namely,

$$F_n(\mathbf{P}) = \int d\mathbf{P}' \bar{K}(\mathbf{P}' \rightarrow \mathbf{P}) F_{n-1}(\mathbf{P}') + \bar{S}(\mathbf{P}), \quad n = 1, 2, 3, \dots, \quad (10.57)$$

where $F_0(\mathbf{P}) = 0$. Thus,

$$\begin{aligned} F_1(\mathbf{P}) &= \bar{S}(\mathbf{P}), \\ F_2(\mathbf{P}) &= S(\mathbf{P}) + \int d\mathbf{P}_1 \bar{K}(\mathbf{P}_1 \rightarrow \mathbf{P}) \bar{S}(\mathbf{P}_1), \\ F_3(\mathbf{P}) &= S(\mathbf{P}) + \int d\mathbf{P}_2 \bar{K}(\mathbf{P}_2 \rightarrow \mathbf{P}) F_2(\mathbf{P}_2) \\ &= S(\mathbf{P}) + \int d\mathbf{P}_1 \bar{K}(\mathbf{P}_1 \rightarrow \mathbf{P}) \bar{S}(\mathbf{P}_1) \\ &\quad + \int d\mathbf{P}_1 \int d\mathbf{P}_2 \bar{K}(\mathbf{P}_1 \rightarrow \mathbf{P}_2) \bar{K}(\mathbf{P}_2 \rightarrow \mathbf{P}) \bar{S}(\mathbf{P}_1), \end{aligned}$$

or, in general,

$$F_n(\mathbf{P}) = \bar{S}(\mathbf{P}) + \sum_{m=1}^{n-1} \int d\mathbf{P}_1 \dots \int d\mathbf{P}_m \bar{K}(\mathbf{P}_1 \rightarrow \mathbf{P}_2) \dots \bar{K}(\mathbf{P}_m \rightarrow \mathbf{P}) \bar{S}(\mathbf{P}_1).$$

If the series converges, the solution of Eq. (10.55) is $F(\mathbf{P}) = \lim_{n \rightarrow \infty} F_n(\mathbf{P})$, i.e.,

$$F(\mathbf{P}) = \bar{S}(\mathbf{P}) + \sum_{m=1}^{\infty} \int d\mathbf{P}_1 \dots \int d\mathbf{P}_m \bar{K}(\mathbf{P}_1 \rightarrow \mathbf{P}_2) \dots \bar{K}(\mathbf{P}_m \rightarrow \mathbf{P}) \bar{S}(\mathbf{P}_1). \quad (10.58)$$

Formal Evaluation of G

The value of G can now be formally obtained by substitution of the collision rate density given by Eq. (10.58) into Eq. (10.56). The result is

$$\begin{aligned} G &= \int d\mathbf{P}_1 \bar{S}(\mathbf{P}) + \sum_{m=1}^{\infty} \int d\mathbf{P}_1 \dots \int d\mathbf{P}_{m+1} \\ &\quad \times \bar{S}(\mathbf{P}_1) \bar{K}(\mathbf{P}_1 \rightarrow \mathbf{P}_2) \dots \bar{K}(\mathbf{P}_m \rightarrow \mathbf{P}_{m+1}) \bar{S}(\mathbf{P}_1) g(\mathbf{P}_{m+1}), \end{aligned}$$

which can be written more succinctly as

$$G = \sum_{m=0}^{\infty} \int d\mathbf{P}_1 \dots \int d\mathbf{P}_{m+1} \bar{S}(\mathbf{P}_1) \bar{K}(\mathbf{P}_1 \rightarrow \mathbf{P}_2) \dots \bar{K}(\mathbf{P}_m \rightarrow \mathbf{P}_{m+1}) \bar{S}(\mathbf{P}_1) g(\mathbf{P}_{m+1}). \quad (10.59)$$

Evaluation of G by Monte Carlo

To evaluate Eq. (10.59) by Monte Carlo methods, the kernel $\bar{K}(\mathbf{P}' \rightarrow \mathbf{P})$ is first decomposed into a product of a nonabsorption probability, a normalization factor to account

for multiplying interactions such as fission or $(n, 2n)$ reactions, and a normalized kernel, namely,

$$\bar{K}(\mathbf{P}' \rightarrow \mathbf{P}) = [1 - \alpha(\mathbf{P}')] \eta(\mathbf{P}') \kappa(\mathbf{P}' \rightarrow \mathbf{P}). \quad (10.60)$$

Here, $\alpha(\mathbf{P}') = \mu_a(\mathbf{P}')/\mu(\mathbf{P}')$ is the absorption (capture) probability at $\mathbf{P}'$ ¹⁰ and $\eta(\mathbf{P}')$ is the normalization factor

$$\eta(\mathbf{P}') = \frac{\int d\mathbf{P} \bar{K}(\mathbf{P}' \rightarrow \mathbf{P})}{1 - \alpha(\mathbf{P}')}. \quad (10.61)$$

The normalized collision kernel $\kappa(\mathbf{P}' \rightarrow \mathbf{P})$ is

$$\kappa(\mathbf{P}' \rightarrow \mathbf{P}) = \frac{\bar{K}(\mathbf{P}' \rightarrow \mathbf{P})}{\int d\mathbf{P}'' \bar{K}(\mathbf{P}' \rightarrow \mathbf{P}'')}, \quad (10.62)$$

which has the property that $\int d\mathbf{P} \kappa(\mathbf{P}' \rightarrow \mathbf{P}) = 1$ so that $\kappa(\mathbf{P}' \rightarrow \mathbf{P})$ is a conditional PDF from which the next collision point $\mathbf{P}$ can be sampled given that the previous nonabsorption collision was at $\mathbf{P}'$.

Now apply the factorization of Eq. (10.60) to Eq. (10.59) to obtain

$$\begin{aligned} G = & \sum_{m=0}^{\infty} \int d\mathbf{P}_1 \int d\mathbf{P}_2 \cdots \int d\mathbf{P}_{m+1} \bar{S}(\mathbf{P}_1) [1 - \alpha(\mathbf{P}_0)] \kappa(\mathbf{P}_1 \rightarrow \mathbf{P}_2) \\ & \times [1 - \alpha(\mathbf{P}_2)] \kappa(\mathbf{P}_2 \rightarrow \mathbf{P}_3) \cdots [1 - \alpha(\mathbf{P}_m)] \kappa(\mathbf{P}_m \rightarrow \mathbf{P}_{m+1}) \\ & \times \alpha(\mathbf{P}_{m+1}) Z(\mathbf{P}_0, \mathbf{P}_1, \dots, \mathbf{P}_{m+1}), \end{aligned} \quad (10.63)$$

where Z is defined as

$$W(\mathbf{P}_1, \mathbf{P}_2, \dots, \mathbf{P}_{m+1}) \equiv \frac{g(\mathbf{P}_{m+1})}{\alpha(\mathbf{P}_{m+1})} \prod_{j=1}^m \eta(\mathbf{P}_j). \quad (10.64)$$

Eq. (10.63) suggests a Monte Carlo algorithm for evaluating G . Each particle history is created in the following manner: (1) The coordinates of the first collision, $\mathbf{P}_1$, are obtained by sampling from the first-flight collision density rate $\bar{S}(\mathbf{P}_1)$ (after renormalization of $\bar{S}(\mathbf{P}_0)$ to make it a proper PDF). (2) Then for $i = 1, 2, \dots, m + 1$, sample $\alpha(\mathbf{P}_i)$ to see if the history ends in an absorption at the i th collision. (3) Sample $\kappa(\mathbf{P}_{i-1} \rightarrow \mathbf{P}_i)$ to find the next collision coordinates (assuming the particles survived

¹⁰ If the spatial coordinates of $\mathbf{P}'$ lie outside V , the particle has leaked from the system and $\alpha(\mathbf{P}')$ is set to unity, the same as if the particle had been absorbed.

the previous collision and the chain is to be continued). With this procedure, the quantity

$$\{\bar{S}(\mathbf{P}_1)[1-\alpha(\mathbf{P}_1)]\kappa(\mathbf{P}_1 \rightarrow \mathbf{P}_2) \cdots \\ \cdots [1-\alpha(\mathbf{P}_m)]\kappa(\mathbf{P}_m \rightarrow \mathbf{P}_{m+1})\alpha(\mathbf{P}_{m+1})\} d\mathbf{P}_1, d\mathbf{P}_2, \dots, d\mathbf{P}_{m+1} \quad (10.65)$$

can be identified as the probability that the initial coordinates are in $d\mathbf{P}_1$ about $\mathbf{P}_1$, the second in $d\mathbf{P}_2$ about $\mathbf{P}_2$, and so on until the chain is terminated by an absorption at collision $m + 1$.

The quantity $Z(\mathbf{P}_1, \mathbf{P}_2, \dots, \mathbf{P}_{m+1})$ is the contribution of the history to the tally for G . Moreover, from the interpretation of the expression inside the brackets of Eq. (10.65) as the joint PDF for $(\mathbf{P}_1, \dots, \mathbf{P}_{m+1})$, it is seen that the right side of Eq. (10.63) is just the expected value of Z , i.e., $\langle Z \rangle = G$. Thus, sampling from the PDFs $\bar{S}$, α , and κ , and scoring the resulting value of Z defined by Eq. (10.64), yields an unbiased estimator for G . The Monte Carlo estimate of G is then obtained by sampling N histories, calculating Z for each history, and averaging the scores, i.e.,

$$G \simeq \frac{1}{N} \sum_{i=1}^N Z(\mathbf{P}_1^i, \dots, \mathbf{P}_{n_i}^i),$$

where n_i is collision number when the i th particle was absorbed or escaped.

10.11.2 Integral Equation Method as Simulation

It may not be readily apparent how the previous sampling scheme, based on the integral transport equation, is related to the Monte Carlo simulation method outlined in Section 10.1. The simulation approach had no need for a transport equation. However, the previous scheme is really an equivalent approach for transport simulations. Here, it is shown that the scheme based on the integral transport equation for estimating the functional G can be cast into a Monte Carlo simulation of which the histories are constructed much more physically.

For simplicity, it is assumed here that the medium is nonmultiplying, i.e., the medium can only absorb or scatter particles. For such a medium, $\eta(\mathbf{P}) = 1$. A history begins by sampling the source function $S(\mathbf{r}, E, \boldsymbol{\Omega})$ (properly normalized) to select starting values $(\mathbf{r}_0, E_0, \boldsymbol{\Omega}_0)$. Then, the flight distance R to the first collision point $\mathbf{r}_1$ is found by sampling R from

$$\mu(\mathbf{r}_0, E_0) \exp \left\{ - \int_0^R \mu(\mathbf{r}_0 - R' \boldsymbol{\Omega}_0, E_0) dR' \right\},$$

(again properly normalized) to obtain the first collision point $\mathbf{r}_1 = \mathbf{r}_0 + R\boldsymbol{\Omega}_0$.

The next step is to determine if the first collision is an absorption, in which case the history is terminated. This is done by sampling from the absorption probability $\alpha(\mathbf{r}_1, E_0) = \mu_a(\mathbf{r}_1, E_0)/\mu(\mathbf{r}_1, E_0)$. However, as discussed later in Section 10.12.4, it

is a bad idea to terminate a history when an absorption event occurs. Rather, it is more efficient to continue the particle history after reducing the weight of the particle by multiplying its weight by the nonabsorption probability $[1 - \alpha(\mathbf{r}_0, E_0)] = \mu_s(\mathbf{r}_0, E_0)/\mu(\mathbf{r}_0, E_0)$. In this manner, every interaction is deemed to be a scatter.

After this (perhaps forced) scatter, a new particle energy and direction are determined from the appropriate scattering distributions. The new energy is obtained by sampling from the PDF

$$f(E) = \frac{\mu_s(\mathbf{r}_1, E_0 \rightarrow E)}{\int dE \mu_s(\mathbf{r}_1, E_0 \rightarrow E)},$$

and the new direction is found by sampling from the PDF

$$f(\Omega) = \frac{\mu_s(\mathbf{r}_0, E_0 \rightarrow E_1, \Omega_0 \cdot \Omega)}{\int_{4\pi} \mu_s(\mathbf{r}_0, E_0 \rightarrow E_1, \Omega_0 \cdot \Omega)}.$$

This sampling procedure is continued to successive collision events until the particle weight falls below some predetermined minimum value (at which point, Russian roulette—see [Section 10.12.3](#)—is used to, perhaps, end the history) or until the particle leaves the phase space of interest in the problem. When the particle history is ended, its contribution is added to the tally for G , and a new particle history is begun.

10.12 Variance Reduction and Nonanalog Methods

To define the efficiency of a Monte Carlo calculation, one must take into account both speed and variance. The speed is, to a large extent, determined by the hardware. Of course, the skill of the computer programmer and the operating system of the computer also determine speed. The emphasis in this section, however, is the variance of the result of a calculation and ways to minimize it.

As discussed in Chapter 5, the power of Monte Carlo depends on using many nonanalog techniques to reduce the variance of the estimator. In this section, some of the more useful variance reduction techniques for particle transport are briefly reviewed.

10.12.1 Importance Sampling

The use of importance sampling can greatly increase the flexibility of scoring and sampling procedures. For any single step in the Monte Carlo process involving the probability of survival of a particle or of its reaching a certain point, the result of the sampling process, insofar as its effect on the final answer is concerned, may be taken as the product of the particle weight and the probability of occurrence of the event. Thus, if a PDF $f(x)$ describes some random process, one is allowed to use some alternative, perhaps simpler, probability density function $\tilde{f}(x)$; however, the ultimate expectation of the process must be left unchanged, and therefore the particle's weight W before entering the process must be adjusted by the multiplicative factor $w(x)$, where

$w(x) \equiv f(x)/\tilde{f}(x)$. The function $\tilde{f}(x)$ is somewhat arbitrarily chosen, although it should not deviate too markedly from $f(x)$ and should have the nature of a proper PDF. In an analog simulation, the use of an alternate PDF instead of $f(x)$ is compensated by the subsequent testing for rejection or acceptance. In the importance sampling technique, the value of x selected is accepted without further ado, but the particle weight after the process is obtained by multiplying its previous weight by $w(x)$, where the argument is the value of x selected.

The use of importance sampling, which is a very effective variance reduction technique for any Monte Carlo problem, can also be used with great effect for particle transport. Consider the problem of [Section 10.11.1](#) in which

$$G = \int d\mathbf{P} g(\mathbf{P}) F(\mathbf{P}) \quad (10.66)$$

is to be estimated. The collision density $F(\mathbf{P})$ is given by the integral transport equation of Eq. (9.73)

$$F(\mathbf{P}) = \int d\mathbf{P}' \bar{K}(\mathbf{P}' \rightarrow \mathbf{P}) F(\mathbf{P}') + \bar{S}(\mathbf{P}). \quad (10.67)$$

The function $g(\mathbf{P})$ can be viewed as the contribution to G from a collision at $\mathbf{P}$. In many problems, $g(\mathbf{P})$ is very small over large portions of phase space, and, to reduce the number of collisions in such regions, it is desirable to bias the samplings that generate the simulated particle histories so that more collisions occur in regions where $g(\mathbf{P})$ is large. Ideally, if one could find an importance function $I(\mathbf{P})$ that is proportional to the expected contribution to G from a particle at $\mathbf{P}$, this function could be used to bias the density functions used to generate the sequence of points in phase space that define a particle history.

The importance function can be used to define the following two functions:

$$\tilde{S}(\mathbf{P}) \equiv \frac{\bar{S}(\mathbf{P}) I(\mathbf{P})}{\int \bar{S}(\mathbf{P}') I(\mathbf{P}') d\mathbf{P}'}, \quad (10.68)$$

$$\tilde{K}(\mathbf{P}' \rightarrow \mathbf{P}) \equiv \frac{\bar{K}(\mathbf{P}' \rightarrow \mathbf{P}) I(\mathbf{P})}{I(\mathbf{P}')}. \quad (10.69)$$

As in [Section 10.11.1](#), these functions can be used to construct biased histories with the weight of each particle adjusted at each step to compensate for the bias. Thus, a source particle has its weight modified by a multiplicative factor of $\tilde{S}(\mathbf{P})/\bar{S}(\mathbf{P})$, and each particle that enters a collision from a previous collision has its weight multiplied by $\bar{K}(\mathbf{P}' \rightarrow \mathbf{P})/\tilde{K}(\mathbf{P}' \rightarrow \mathbf{P})$.

Now what function to choose for $I(\mathbf{P})$? It can be shown [[Carter and Cashwell 1975](#)] that the optimum importance function I_{opt} is the solution of the equation

$$I_{opt}(\mathbf{P}) = \int d\mathbf{P}' \bar{K}(\mathbf{P} \rightarrow \mathbf{P}') I_{opt}(\mathbf{P}') + g(\mathbf{P}). \quad (10.70)$$

This is just the integral adjoint transport equation of Eq. (9.85) with the adjoint source specified as $\bar{S}^\dagger(\mathbf{P}) = g(\mathbf{P})$. With the optimum importance function, the Monte Carlo calculation gives a score of G for *every* history and, thus, is an estimator with zero variance! However, solving Eq. (10.70) is as much work as solving the original transport equation of Eq. (10.67). Indeed, if $I_{opt}(\mathbf{P})$ were known, G could be estimated with simple numerical quadrature by using Eq. (9.89), namely,

$$G = \int g(\mathbf{P})F(\mathbf{P}) d\mathbf{P} = \int \bar{S}(\mathbf{P})I_{opt}(\mathbf{P}) d\mathbf{P}. \quad (10.71)$$

Although this optimum importance function is not known beforehand, its existence tempts one to try to develop various approximations for it without unduly increasing the computation time. These approximations are often based on deterministic methods such as discrete ordinates calculations using a simplified geometry [Tang et al. 1976]. Alternatively, a Monte Carlo calculation of the adjoint flux can be performed using simplified physics models such as energy multigroup cross sections [MCNP 2003].

10.12.2 Truncation Methods

During the course of its random walk, a particle may reach a location, direction, energy, or weight for which the likelihood of scoring is negligible. The energy may be so low or the position so remote that terminating the tracking of the particle is justified. Truncation is a particularly valuable tool in the tracking of charged particles, which have reasonably well-defined ranges, so that a truncation decision is straightforward. A special type of truncation usually associated with small particle weights is called Russian roulette. The Russian roulette method of truncation is addressed, as follows, along with *splitting*, its companion procedure used for variance reduction.

10.12.3 Splitting and Russian Roulette

Suppose that by virtue of its position, direction, or some other feature, a particle is relatively likely to make an appreciable contribution to a score. The variance may be reduced by *splitting* the particle into two or more, say, m particles, each having the same characteristics as the original particle, except each has a weight given by the fraction $1/m$ of the weight of the original particle. Each particle is then tracked independently. On the other hand, a particle may be relatively unlikely to make an appreciable contribution to a score. Then, a random number ρ may be selected, and if less than some fraction, say $1/m$, the particle may be *killed* by the Russian roulette scheme. Otherwise, tracking the particle would continue, but the particle's weight would be multiplied by the factor m .

Splitting and Russian roulette are normally used together. Different spatial or energy regions, for example, may be assigned *importances* I . As a particle moves from region j to region k , if $n \equiv I_k/I_j$ exceeds unity, the particle is split into n particles¹¹ with

¹¹ If n is not an integer, the particle may be split into one of the bounding-integer number of particles and the weights adjusted as discussed in Section 5.8.

weights adjusted by the factor $1/n$. If n is less than unity, then Russian roulette may be played, with the particle surviving with probability $1 - n$ and its weight multiplied by the factor $1/n$. In a shielding problem with particles penetrating a thick-slab shield, [Carter and Cashwell \[1975\]](#) suggest that, for geometric splitting and Russian roulette, the slab be divided into layers of one mean-free-path thickness. As particles pass from region to region away from the entry face they should be split 2 for 1. Russian roulette with $1/2$ survival probability should be applied as particles pass from region to region toward the entrance face.

10.12.4 Implicit Absorption

An example of the usefulness of weighting is in the avoidance of the absorption process for a particle. Because particles absorbed in a medium give zero contribution to the final answer in many transport calculations, one intuitively has the feeling that a history terminating in this way is, in a sense, a waste of computer time. Thus, if a collision is deemed an absorption, the particle weight is reduced by multiplying its incident weight by the probability of survival, i.e., $[1 - (\mu_a/\mu)]$, and the type of collision redetermined by sampling from all the nonabsorption interactions probabilities. This numerical avoidance of absorption is sometimes called *implicit absorption* or *survival biasing*, distinct from *analog absorption*, in which the particle history is actually terminated.

However, in such a scheme the only way for a history to terminate is for the particle to leak from the system, and this can result in the buildup of a large number of particles with very low weights and that contribute little to the tally. This implicit capture is almost always combined with Russian roulette (see [Section 10.12.3](#)) to artificially reduce the population of low-weight particles when a particle's weight falls below some specified value.

10.12.5 Interaction Forcing

Consider, for example, the passage of photons or neutrons through important regions, but regions with dimensions of only one or so mean-free paths, so that a particle is likely to pass through the region without interaction. The effort wasted by tracking a particle through such a region can be eliminated by *interaction forcing*. Suppose that the distance through the region along the particle's path is S . Collisions may be forced if path lengths are selected by sampling the following alternate PDF:

$$\tilde{f}(s) = \frac{\mu e^{-\mu s}}{1 - e^{-\mu S}}, \quad 0 \leq s \leq S. \quad (10.72)$$

The particle weight must then be adjusted by the factor $(1 - e^{-\mu S})$.

10.12.6 Exponential Transformation

For Monte Carlo calculations involving penetration or reflection from a slab shield, there is motivation to artificially increase or decrease, respectively, the distance

between collisions. To do this, the normal PDF for sampling path lengths, $f(s) = \mu \exp(-\mu s)$, is modified by replacing μ by $(1 - \beta\omega)\mu$, in which ω is the cosine of the angle between the preferred direction (toward the scoring region) and the direction of flight of the particle. The factor β is a biasing parameter, normally restricted to values less than unity. For deep-penetration problems, this *exponential transformation* is often applied only to particles traveling in the generally favored direction (i.e., with $\omega > 0$). In other words, path lengths may be selected by sampling the following alternate PDF:

$$\tilde{f}(s) = (1 - \beta\omega)\mu e^{-(1-\beta\omega)\mu s}. \quad (10.73)$$

For $\omega > 0$, the effective interaction coefficient $(1 - \beta\omega)\mu$ is less than μ ; thus, the particle path is longer than would be the case in a strictly analog calculation. The corresponding weight factor is

$$w(s) = \frac{e^{-\beta\omega\mu s}}{1 - \beta\omega}. \quad (10.74)$$

By contrast, for $\omega < 0$, as may be the case for particles moving away from the scoring region, the effective interaction coefficient is greater than μ and the particle path is foreshortened. The parameter β must not be so near unity that excessively large weight factors result. In a particular application, it may be necessary to adjust β by trial and error to optimize calculations.

10.13 Summary

In a Monte Carlo neutral-particle transport simulation, the geometry of the system is first specified, typically by combinatorial geometry, although other approaches can be used. Then by sampling from many distributions a complete simulation of a particle's track as it migrates through phase space can be obtained. Source sampling is required to pick the initial starting location, direction, and energy of a particle. Then sampling is used to pick a flight distance before a collision, followed by sampling to determine the type of collision. If the particle is not absorbed, more sampling is done to determine the type, direction, and energy of secondary particles. Each subsequent leg of the simulated track continues as for the first leg. The particle is tracked until it is absorbed or leaves the problem boundary. As a particle moves along its trajectory, various tallies are updated so that, after many histories, some desired property of the radiation field can be estimated.

To reduce computational effort, transport simulations often change the physical sampling distributions to bias a particle's track to increase the chance of it scoring. To remove the bias from the score, a weight is assigned to each particle, and as the particle track evolves, the weight is adjusted to remove the bias each time a biased sampling is used. Such biasing can greatly reduce the variance of the tally and, consequently, most Monte Carlo simulations employ one or more variance reduction techniques.

With the advent of inexpensive and powerful personal computers and computer clusters and with the availability of powerful general-purpose Monte Carlo codes, Monte Carlo has become the method of choice for the vast majority of neutron- and photon-transport analyses.

Bibliography

- BLOMQUIST, R.N. AND E.M. GELBARD 1983, "An Assessment of Existing Klein-Nishina Monte Carlo Sampling Methods," *Nucl. Sci. Eng.*, **83**, 380–384.
- CARTER, L.L. AND E.D. CASHWELL 1975, *Particle-Transport Simulation with the Monte Carlo Method*, TID-26607, National Technical Information Service, U.S. Department of Commerce, Springfield, VA.
- CASE, K.M., F. DE HOFFMAN, AND G. PLACZEK 1953, "Introduction to the Theory of Neutron Diffusion," Vol. I, Los Alamos Scientific Laboratory, NM.
- CHANDRASEKHAR, S. 1960, "Radiative Transfer," Dover, New York.
- KAHN, H. 1956, "Applications of Monte Carlo," Report AECU-3259, The Rand Corp., Santa Monica, CA.
- KALOS, M.H., F.R. NAKACHE, AND J. CELNIK 1968, "Monte Carlo Methods in Reactor Computations," in *Computing Methods in Reactor Physics*, H. Greenspan, C.N. Kelber, and D. Okrent, eds., Gordon and Breach, New York.
- LUX, I. AND L. KOBLINGER 1991, *Monte Carlo Particle Transport Methods: Neutron and Photon Calculations*, CRC Press, Boca Raton, FL.
- MCNP 2003, *A General Monte Carlo N-Particle Transport Code, Version 5*, Vol. I, *Overview and Theory*, LA-UR-03-1987, Los Alamos National Lab, NM.
- OLMSTEDED, J.M.H. 1947, *Solid Analytic Geometry*, Appelton-Century-Crafts Inc., New York.
- SHULTIS, J.K. AND R.E. FAW 2000, *Radiation Shielding*, American Nuclear Society, LaGrange Park, IL.
- SHULTIS, J.K. AND R.E. FAW 2007, *Fundamentals of Nuclear Science and Engineering*, 2nd ed., CRC Press, Boca Raton, FL.
- TANG, J.S., P.N. STEVENS, AND T.J. HOFFMAN 1976, *Methods of Monte Carlo Biasing Using Two-Dimensional Discrete Ordinates Adjoint Flux*, ORNL-TM-5454, Oak Ridge National Laboratory, TN.

Problems

1. Verify that the functions $f_1(r)$ and $f_2(r)$ defined by Eq. (10.36) are ≤ 1 for $1 \leq r \leq 1 + 2/\lambda$.
2. Evaluate the constant K required to make Eq. (10.32) a properly normalized PDF. Then derive an expression for the efficiency of the sampling method described in Section 10.6.2. Plot the efficiency as a function of the incident photon energy.
3. Perform a Monte Carlo analysis, similar to that of Example 10.1, to determine the escape probability of a particle born uniformly in a purely absorbing sphere of radius R . Compare your results to the analytical result

$$P_{esc} = \frac{3}{8(\mu R)^3} \left[2(\mu R)^2 - 1 + (1 + 2\mu R)e^{-2\mu R} \right].$$

4. Verify Eqs. (10.19) to (10.21) by evaluating Eq. (10.17).

5. Derive the rotation matrix $R_{xy}(\theta, \psi)$ for first a rotation about the x -axis followed by a rotation about the y -axis. Verify that the inverse of this matrix transforms the vector $\mathbf{\Omega}(\theta, \psi)$ in the vector $(0, 1, 0)$.
6. Consider a purely absorbing slab of thickness z mean-free-path lengths that is uniformly and normally illuminated by particles. The average flux density in the slab can be estimated analytically by (a) the path-length estimator and (b) the reaction-density estimator. Show that the relative error (ratio of the standard deviation to the mean) of these two estimators is as follows:

$$\text{collision estimator} = \sqrt{\frac{e^{-z}}{(1 - e^{-z})}}$$

$$\text{path-length estimator} = \frac{\sqrt{1 - 2ze^{-z} - e^{-2z}}}{(1 - e^{-z})}.$$

- (a) Plot these results as a function of z . (b) Also determine the limiting values as $z \rightarrow 0$ and as $z \rightarrow \infty$. (c) Discuss which estimator is best for optically thin and thick media and explain why these results are expected.
7. This problem requires development of a computer code for calculation of transmission and reflection of photons normally incident on a homogeneous shielding slab surrounded by a vacuum. For simplicity, assume one-speed particles and isotropic scattering. Estimate (a) the reflected and transmitted currents and (b) the reflected and transmitted angular flux densities. Plot the leakage currents as a function of slab thickness for various values of $c = \mu_s/\mu$. Also plot the angular distribution of the escaping angular flux densities for $c = 0.9$ for different slab thicknesses.

A Some Common Probability Distributions

In this appendix, several frequently encountered probability distributions are presented. For each distribution explicit expressions are given for both the PDF and the CDF, and the mean value μ and variance σ^2 are given or derived. More details on these and other distributions can be obtained from references such as [Kendall and Stuart \[1977\]](#), [Snedecor and Cochran \[1971\]](#), [Johnson and Kotz \[1987\]](#), and [Johnson et al. \[1992, 1994, 1995\]](#).

Also presented here are some methods for generating random values from these distributions; however, the emphasis is on simplicity and suggested sampling schemes may not be the fastest or most efficient. Indeed, the literature is filled with literally hundreds of research papers on “better” random variate generators, many duplicative because of the widespread use of Monte Carlo methods in very different disciplines. Also, many are of incremental improvement, a result of a minor tweak in some line of computer code. The reader interested in random-number generators faster (and more complex) than those given here are referred to several excellent works on the subject, notably [Fishman \[2003\]](#), [Gentle \[1998\]](#), Devroye [1986], [McGrath and Irving \[1975\]](#), and [Rubinstein \[1981\]](#).

A.1 Continuous Distributions

In this section, 11 continuous univariate probability distributions are presented, namely, the uniform, the exponential, the gamma, the beta, the Weibull, the normal, the lognormal, the Cauchy or Lorentzian, the chi-squared, the Student’s t , and the Pareto. There are many other continuous distributions; the interested reader is referred to [Johnson et al. \[1994, 1995\]](#), [Kendall and Stuart \[1977\]](#), [McGrath and Irving \[1975\]](#), or many other books on probability and statistics.

Sampling and linear transformations

In many of the sampling schemes for continuous distributions, the sampling is first performed from a simpler *standardized* form of the distribution with some parameters set to zero or unity. Then a linear transformation is used to obtain a sample from the more general form of the distribution. For example, the shifted gamma distribution,

$$f(x) = \frac{1}{\beta\Gamma(\alpha)} \left(\frac{x-a}{\beta} \right)^{\alpha-1} e^{-(x-a)/\beta}, \quad a \leq x < \infty,$$

has a shift parameter a , a shape parameter α , and a scale parameter β . The linear transformation $(x - a)/\beta = y$ transforms the above PDF to the much simpler standardized PDF

$$g(y) = \frac{1}{\Gamma(\alpha)} y^{\alpha-1} e^{-y}, \quad 0 \leq y \leq \infty.$$

It is much easier to obtain a sample y from this single-parameter PDF, and then transform it back to obtain $x = \beta y + a$ as a sample from the three-parameter form of the distribution. This trick is useful for many of the distributions discussed here.

Inverse CDF method

The inverse CDF method (see Section 4.1.1) is an excellent way to generate samples if the inverse CDF function is easily calculated or if an accurate approximation is known. Thus, one simply has to evaluate $F^{-1}(\rho)$ to obtain a sample from the distribution. In [Table A.1](#), the inverse functions for several continuous distributions are summarized.

A.1.1 Uniform Distribution

The uniform or rectangular distribution can be defined over any range $[a, b]$, where a and b are finite and $b > a$. In this distribution, denoted by $\mathcal{U}(a, b)$, all values within the range are equally likely. Thus, the PDF is

$$f(x) = \frac{1}{b-a}, \quad a \leq x \leq b, \quad (\text{A.1})$$

and the CDF is

$$F(x) = \int_a^b \frac{1}{b-a} dx = \frac{x-a}{b-a}, \quad a \leq x \leq b. \quad (\text{A.2})$$

The mean value is

$$\mu = \frac{b+a}{2} = a + \frac{b-a}{2}, \quad (\text{A.3})$$

and the variance is given by

$$\sigma^2 = \overline{x^2} - \mu^2 = \frac{b^3 - a^3}{3(b-a)} - \left(\frac{b+a}{2} \right)^2 = \frac{(b-a)^2}{12}. \quad (\text{A.4})$$

An example uniform distribution over the interval $(-6, 4)$ is shown in [Fig. A.1](#).

Table A.1 Some continuous distributions for which the inverse CDF method can be used to obtain random variates $x = F^{-1}(\rho)$, where ρ is from $\mathcal{U}(0, 1)$ (after Fishman [2003]).

Distribution	PDF $f(x)$	Inverse $F^{-1}(\rho)$	Efficient Form
uniform $\mathcal{U}(a, b)$ $a \leq x \leq b$	$\frac{1}{b-a}$	$a + (b-a)\rho$	—
exponential $\mathcal{E}(\lambda)$ $\lambda > 0; x \geq 0$	$\lambda e^{-\lambda x}$	$-\frac{1}{\lambda} \ln(1-\rho)$	$-\frac{1}{\lambda} \ln(\rho)$
beta $\mathcal{B}_e(\alpha, 1)$ $\alpha > 0, 0 \leq x \leq 1$	$\alpha x^{\alpha-1}$	$\rho^{1/\alpha}$	—
beta $\mathcal{B}_e(1, \beta)$ $\beta > 0, 0 \leq x \leq 1$	$\beta(1-x)^{\beta-1}$	$1 - (1-\rho)^{1/\beta}$	$1 - \rho^{1/\beta}$
logistic $\mathcal{L}(\alpha, \beta)$ $\beta > 0; -\infty < x, \alpha < \infty$	$\frac{e^{-(x-\alpha)/\beta}}{\beta[1 + e^{-(x-\alpha)/\beta}]^2}$	$\alpha + \beta \ln[\rho/(1-\rho)]$	—
Weibull $\mathcal{W}(\alpha, \beta)$ $\alpha, \beta > 0; x \geq 0$	$\frac{\alpha}{\beta} \left(\frac{x}{\beta}\right)^{\alpha-1} e^{-(x/\beta)^\alpha}$	$\beta[-\ln(1-\rho)]^{1/\alpha}$	$\beta[-\ln \rho]^{1/\alpha}$
normal ^a $\mathcal{N}(\mu, \sigma)$ $\sigma^2 > 0; -\infty < x, \mu < \infty$	$\frac{e^{(\mu-x)^2/(2\sigma^2)}}{\sqrt{2\pi\sigma^2}}$	$\mu + \text{sign}(\rho - 1/2)\sigma \times$ $\left(t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3}\right)$ $t \equiv \sqrt{-\ln[\min(\rho, 1-\rho)]^2}$	—
Cauchy $\mathcal{C}(\alpha, \beta)$ $\beta > 0; -\infty < x, \alpha < \infty$	$\frac{1}{\pi} \frac{\beta}{(x-\alpha)^2 + \beta^2}$	$\alpha + \beta \tan \pi[\rho - (1/2)]$	$\alpha + \beta \tan \pi \rho$
Pareto $\mathcal{P}_a(\alpha, \beta)$ $\alpha > 0; x \geq \beta > 0$	$\frac{\alpha \beta^\alpha}{x^{\alpha+1}}$	$\frac{\beta}{(1-\rho)^{1/\alpha}}$	$\frac{\beta}{\rho^{1/\alpha}}$

^a $c_0 = 2.515517, c_1 = 0.802853, c_2 = 0.010328, d_1 = 1.432788, d_2 = 0.189269, d_3 = 0.001308$; absolute error $< 0.45 \times 10^{-3}$; this approximation is from Hastings [1955].

Sampling from the rectangular distribution

Of obvious importance is the unit rectangular distribution $\mathcal{U}(0, 1)$, which forms the basis of almost all sampling schemes for other probability distributions. Random variates from the unit rectangular distribution are denoted in this book by ρ_i to distinguish them from other random variates. The generation of sequences of ρ_i is the subject of Chapter 3. To obtain a random sample x_i from $\mathcal{U}(a, b)$, one first takes a sample ρ_i from

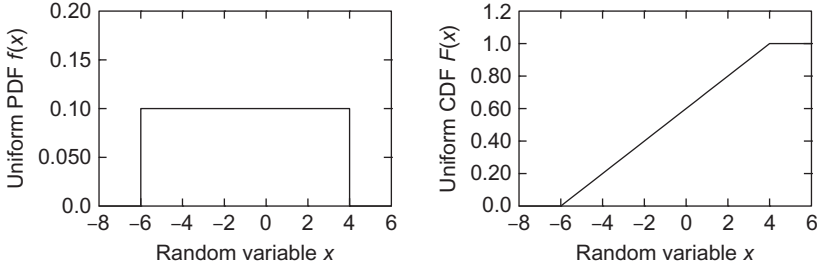


Figure A.1 The PDF and the CDF for the uniform distribution with $a = -6$ and $b = 4$.

$\mathcal{U}(0, 1)$ and then applies a linear transformation to obtain

$$x_i = a + \rho_i(b - a). \quad (\text{A.5})$$

This result is equivalent to the inverse CDF method (see Example 4.1).

A.1.2 Exponential Distribution

The PDF of the exponential distribution is given by

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0, \quad (\text{A.6})$$

and is denoted by $\mathcal{E}(\lambda)$. The scale parameter $\lambda > 0$ determines the rate of descent of the exponential. The CDF is

$$F(x) = \int_0^x \lambda e^{-\lambda u} du = 1 - e^{-\lambda x}. \quad (\text{A.7})$$

The moments of the PDF are

$$\langle x^n \rangle = \int_0^\infty u^n \lambda e^{-\lambda u} du = \frac{n!}{\lambda^n}. \quad (\text{A.8})$$

It then follows that the mean is $\mu \equiv \langle x \rangle = 1/\lambda$ and the variance is $\sigma^2 = \langle x^2 \rangle - \mu^2 = 1/\lambda^2$. Fig. A.2 shows the exponential distribution for three values of λ . Clearly, λ serves as a scaling parameter.

The exponential distribution is very important because many processes behave exponentially. In Chapter 2, radioactive decay was considered. In the decay process, the number of radioactive atoms at time t is given by

$$N(t) = N(0)e^{-\lambda t}, \quad (\text{A.9})$$

where $N(0)$ is the number of radioactive atoms at time 0 and λ is the decay constant of the particular radioactive species. Also, in Chapter 9, it was found that neutral particles from a source are attenuated exponentially according to

$$\phi(r) = \phi^o(0)e^{-\mu r}, \quad (\text{A.10})$$

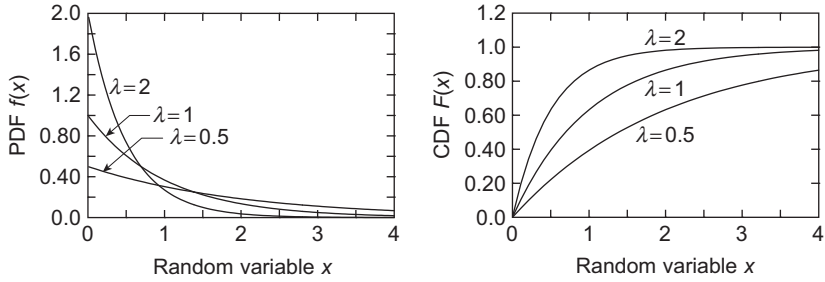


Figure A.2 The PDF and the CDF for the exponential distribution for three values of the parameter λ .

where $\phi^o(0)$ is the flux density at $r = 0$, μ is the total linear attenuation coefficient of the medium, and r is the distance from the source. More generally, the exponential distribution applies to any process for which the rate of change of a quantity f with respect to a variable x is proportional to the quantity, i.e.,

$$\frac{df}{dx} = \lambda f, \quad (\text{A.11})$$

where λ is the constant of proportionality.

Sampling from an exponential distribution

Generation of variates from the exponential distribution is readily done using the inverse CDF method (see Example 4.2), namely,

$$x_i = -\ln(\rho_i)/\lambda, \quad (\text{A.12})$$

where the ρ_i are sampled from $\mathcal{U}(0, 1)$.

Translated exponential distribution

A slightly more generalized version of the exponential distribution is defined on the half-open interval $[a, \infty)$. The PDF in this case is just translated along the x -axis by an amount a and takes the form

$$f(x) = \lambda e^{-\lambda(x-a)}, \quad x \geq a, \quad \lambda > 0. \quad (\text{A.13})$$

This distribution is denoted by $\mathcal{E}(\lambda, a)$. The CDF is seen to be

$$F(x) = 1 - e^{-\lambda(x-a)}. \quad (\text{A.14})$$

Note that the PDF integrates to unity regardless of the value of a and the CDF reduces to Eq. (A.7) for $a = 0$. The mean value is

$$\mu = \int_a^\infty \lambda e^{-\lambda(u-a)} du = \frac{1}{\lambda} + a \quad (\text{A.15})$$

and the variance is

$$\sigma^2 = \frac{2e^{\lambda a} - (\lambda a + 1)^2}{\lambda^2}. \quad (\text{A.16})$$

Sampling from a translated exponential distribution

To generate variates from the shifted exponential distribution, first sample from the unshifted distribution (see Eq. (A.12)) and add the shift, i.e.,

$$x_i = -\ln(\rho_i)/\lambda + a, \quad (\text{A.17})$$

where the ρ_i are uniformly distributed in $(0, 1)$.

A.1.3 Gamma Distribution

Somewhat related to the exponential distribution is the gamma distribution $\mathcal{G}(\alpha, \beta)$, the PDF of which can be written as

$$f(x) = \frac{1}{\beta\Gamma(\alpha)} \left(\frac{x}{\beta}\right)^{\alpha-1} e^{-x/\beta} \quad \alpha > 0, \quad \beta > 0, \quad x \geq 0, \quad (\text{A.18})$$

where $\Gamma(\alpha)$ is the gamma function defined by

$$\Gamma(u) = \int_0^\infty x^{u-1} e^{-x} dx. \quad (\text{A.19})$$

The gamma distribution is defined on the half-open interval $[0, \infty)$ and has two parameters, a shape parameter $\alpha > 0$ and a scale parameter $\beta > 0$. The gamma distribution for $\alpha = 1$ is in fact the exponential distribution of Eq. (A.6) with $\beta = 1/\lambda$. This close relationship is, however, somewhat muddled by the fact that no explicit closed-form expression for the CDF is available for general values of α . Thus, the best one can say is

$$F(x) = \frac{1}{\beta\Gamma(\alpha)} \int_0^x \left(\frac{x}{\beta}\right)^{\alpha-1} e^{-x/\beta} dx = P(\alpha, x/\beta), \quad (\text{A.20})$$

where $P(\alpha, y)$ is the *incomplete gamma function* defined as

$$P(\alpha, y) \equiv \frac{\gamma(y)}{\Gamma(\alpha)} = \frac{1}{\Gamma(\alpha)} \int_0^y e^{-u} u^{\alpha-1} du. \quad (\text{A.21})$$

When $\beta = 1$, Eq. (A.18) becomes the *standardized* gamma function $\mathcal{G}(\alpha, 1)$. This standardized gamma function, when α is a positive integer, gives the distribution of waiting times x to the α th Poisson event. Thus, for $\alpha = 1$, $\mathcal{G}(1, 1) \rightarrow e^{-x}$, which is the distribution of waiting times for the first event, such as a radioactive decay in a radionuclide sample.

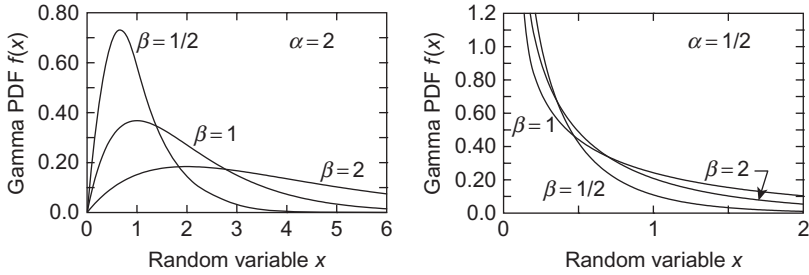


Figure A.3 The PDF for the gamma distribution for $\alpha > 1$ (left) and for $\alpha < 1$ (right).

Another special case for the gamma distribution occurs when the scale parameter $\beta = 2$. The gamma distribution then becomes the chi-squared distribution with 2α equal to the ν degrees of freedom, i.e., $\mathcal{G}(\alpha, 2) \rightarrow \chi^2(\nu = 2\alpha)$ (see Section A.1.9).

The moments of the gamma PDF are given by

$$\langle x^n \rangle \equiv \int_0^\infty x^n f(x) dx = \beta^n \prod_{i=0}^{n-1} (\alpha + i). \quad (\text{A.22})$$

From this it follows that the mean value is $\mu = \langle x \rangle = \alpha\beta$ and the variance is $\sigma^2 = \langle x^2 \rangle - \mu^2 = \alpha\beta^2$, which, of course, reduce to the appropriate results for an exponential distribution when $\alpha = 1$. Examples of the gamma distribution are plotted in Fig. A.3. Notice that when the shape parameter $\alpha > 1$, $f(x|\alpha, \beta)$ is unimodal with a maximum at $x_{\max} = (1 - \alpha^{-1})/\beta$. For $\alpha < 1$, $f(x|\alpha, \beta)$ is a monotonic decreasing function that is unbounded at $x = 0$.

Sampling from a gamma distribution

Generating samples from the gamma distribution is not straightforward because the sampling method depends on whether α is greater or less than unity. Special and very efficient techniques can be used if α is a positive integer [Press et al. 1996]. For the general case, methods for sampling from the gamma distribution have been described by McGrath and Irving [1975], Rubinstein [1981], and Fishman [2003]. Gentle [1998] reports the following algorithms for sampling from the standardized gamma distribution.

Case $\alpha > 1$. The Cheng-Feast [1979] algorithm:

1. Generate ρ_1 and ρ_2 from a uniform distribution on $(0, 1)$ and set

$$\xi = \frac{\left(\alpha - \frac{1}{6\alpha}\right)\rho_1}{(\alpha - 1)\rho_2}.$$

2. If $\frac{2(\rho_2 - 1)}{\alpha - 1} + \xi + \frac{1}{\xi} \leq 2$, then return $x = (\alpha - 1)\xi$; otherwise, if $\frac{2 \ln \rho_2}{\alpha - 1} - \ln \xi + \xi \leq 1$, then return $x = (\alpha - 1)\xi$.

3. If no value is returned, go to step 1 and try again.

Case $\alpha < 1$. The Ahrens-Dieter [1974] algorithm, modified by Best [1983]:

1. Set $t = 0.07 + 0.75\sqrt{1 - \alpha}$ and $b = \alpha[1 + (e^t/t)]$.
2. Generate ρ_1 and ρ_2 from a uniform distribution on $(0, 1)$ and set $\xi = b\rho_1$.
3. If $\xi \leq 1$, then set $x = t\xi^{1/\alpha}$; then, if $\rho_2 \leq (2-x)/(2+x)$, return x ; otherwise, if $\rho_2 \leq e^{-x}$, return x ; otherwise, set $x = \ln[t(b - \xi)/\alpha]$ and $y = x/t$ and, if $\rho_2[\alpha + y(1 - \alpha)] \leq 1$, return x .
4. If no value is returned, go to step 2 and try again.

To generate samples z_i from the nonstandardized gamma distribution ($\beta \neq 1$), first generate samples x_i from the standardized distribution using these algorithms and set $z_i = x_i/\beta$.

A.1.4 Beta Distribution

The beta distribution $\mathcal{B}(\alpha, \beta)$ has the PDF

$$f(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}, \quad \alpha > 0, \beta > 0, 0 \leq x \leq 1, \quad (\text{A.23})$$

in which α and β are shape parameters. Like the gamma distribution, numerical integration must be used to evaluate the CDF. The CDF is

$$F(x) = I_x(\alpha, \beta), \quad (\text{A.24})$$

where $I_x(\alpha, \beta)$ is the *incomplete beta function*

$$I_x(\alpha, \beta) \equiv \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^x u^{\alpha-1} (1-u)^{\beta-1} du. \quad (\text{A.25})$$

The moments of the beta distribution are

$$\langle x^n \rangle \equiv \int_0^1 x^n f(x) dx = \prod_{i=1}^{n-1} \frac{\alpha + i}{\alpha + \beta + i} \quad (\text{A.26})$$

from which the mean value is

$$\mu \equiv \langle x \rangle = \frac{\alpha}{\alpha + \beta} \quad (\text{A.27})$$

and the variance is

$$\sigma^2 \equiv \langle x^2 \rangle - \mu^2 = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}. \quad (\text{A.28})$$

Examples of the beta function are shown in Figs. A.4 to A.7. It is seen from these figures that the beta PDF can assume a wide variety of shapes over the interval $[0, 1]$.

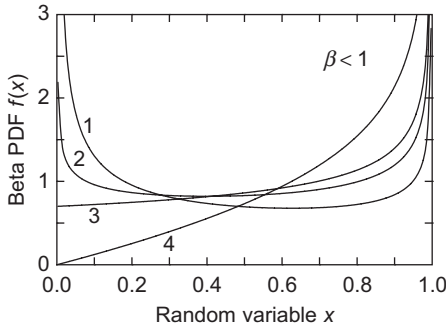


Figure A.4 The PDF for the beta distribution for $\beta < 1$. Here, $\beta_1 = \beta_2 = \beta_3 = \beta_4 < 1$ and $\alpha_1 < \beta_i < \alpha_2 < \alpha_3 = 1 < \alpha_4$.

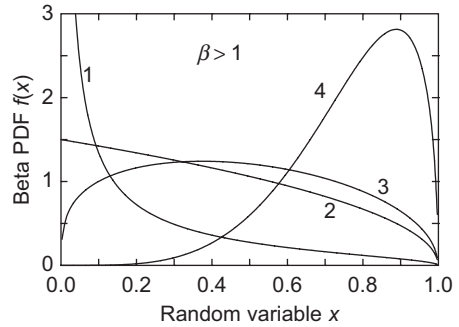


Figure A.5 The PDF for the beta distribution for $\beta > 1$. Here, $\beta_1 = \beta_2 = \beta_3 = \beta_4 > 1$ and $\alpha_1 < \alpha_2 = 1 < \alpha_3 < \beta_i < \alpha_4$.

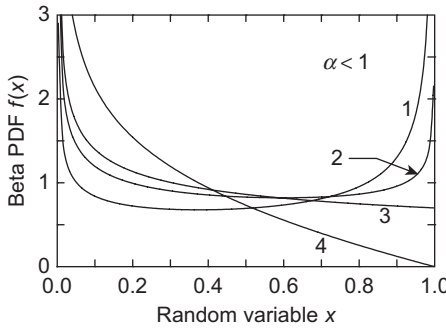


Figure A.6 The PDF for the beta distribution for $\alpha < 1$. Here, $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 < 1$ and $\beta_1 < \alpha_i < \beta_2 < \beta_3 = 1 < \beta_4$.

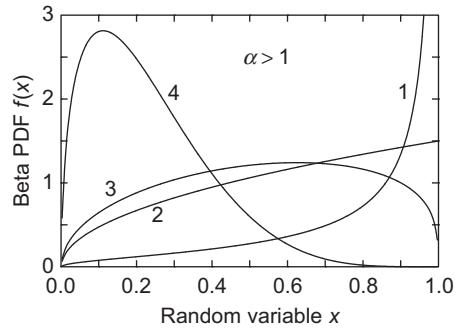


Figure A.7 The PDF for the beta distribution for $\alpha > 1$. Here, $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 > 1$ and $\beta_1 < \beta_2 = 1 < \beta_3 < \alpha_i < \beta_4$.

For this reason the beta function is often used to represent, for example, the distribution of component failure probabilities for components in a complex mechanical system. The parameter α controls the behavior of the function near $x=0$, while β determines the behavior near $x=1$. If $\alpha < 1$, $f(x)$ is unbounded at $x=0$, and if $\beta < 1$, $f(x)$ is unbounded at $x=1$. If $\alpha, \beta > 1$, there is a maximum (or mode) at $x = (\alpha - 1)/(\alpha + \beta - 2)$. If $\beta > \alpha$ or $\alpha > \beta$, the distribution is skewed toward $x=0$ or $x=1$, respectively. If $\alpha = \beta$, the distribution is symmetric about $x=0.5$.

Sampling from a beta distribution

Generating samples from the beta distribution is not straightforward, except for the simple cases of $\alpha = 1$ or $\beta = 1$ or $\alpha = \beta = 1$ (see Table A.1). For $\max(\alpha, \beta) < 1$, Jöhnk [1964] gives the following simple algorithm as reported by Gentle [1998]:

Case $\max(\alpha, \beta) < 1$.

1. Generate ρ_1 and ρ_2 from a uniform distribution on $(0, 1)$ and set $v_1 = \rho_1^{1/\alpha}$ and $v_2 = \rho_2^{1/\beta}$.

2. Set $w = v_1 + v_2$.
3. If $w < 1$, return $x = v_1/w$; otherwise, go to step 1 and try again.

For this case of $\max(\alpha, \beta) < 1$, [Atkinson and Whittaker \[1976\]](#) also give a very fast algorithm. For other cases in which either or both of the shape parameters are greater than unity, fast algorithms have been devised by [Cheng \[1978\]](#), [Schmeiser and Babu \[1980\]](#), and [Zechner and Stadlober \[1993\]](#). [Fishman \[2003\]](#) gives explicit algorithms for these cases.

A.1.5 Weibull Distribution

The PDF of the Weibull distribution $\mathcal{W}(\alpha, \beta)$ can be written as

$$f(x) = \frac{\alpha}{\beta} \left(\frac{x}{\beta} \right)^{\alpha-1} e^{-(x/\beta)^\alpha}, \quad \alpha > 0, \quad \beta > 0, \quad x \geq 0, \quad (\text{A.29})$$

with a scale parameter β and a shape parameter α . Often, x is the time to some event. The Weibull distribution is often used when considering the distribution of lifetimes, and so finds use, for instance, in fatigue analysis. It is also used to describe the probability that large acute radiation doses ($x = \text{dose}$) lead to deterministic effects, such as cataract formation [[Faw and Shultis 1999](#)]. The Weibull CDF is readily found to be

$$F(x) = 1 - e^{-(x/\beta)^\alpha}. \quad (\text{A.30})$$

For the special case that $\alpha = 2$ and $\beta = \sqrt{2}\sigma$, the Weibull distribution becomes the *Rayleigh distribution* with a PDF of

$$f(x) = \frac{x}{\sigma^2} \exp \left[-\frac{x^2}{2\sigma^2} \right]. \quad (\text{A.31})$$

The moments of the Weibull distribution are

$$\langle x^n \rangle \equiv \int_0^\infty x^n f(x) dx = \alpha^n \Gamma \left(1 + \frac{n}{\beta} \right). \quad (\text{A.32})$$

The mean value is thus given by

$$\mu \equiv \langle x \rangle = \beta \Gamma \left(1 + \frac{1}{\alpha} \right), \quad (\text{A.33})$$

and the variance can be written as

$$\sigma^2 \equiv \langle x^2 \rangle - \mu^2 = \beta^2 \left[\Gamma \left(1 + \frac{2}{\alpha} \right) - \Gamma^2 \left(1 + \frac{1}{\alpha} \right) \right]. \quad (\text{A.34})$$

Examples of the Weibull PDF are shown in [Figs. A.8 and A.9](#). For $\alpha < 1$, the Weibull PDF is a monotonic decreasing function of x that is unbounded at $x = 1$, while for $\alpha > 1$ the Weibull distribution is unimodal with a maximum at $x = \alpha[1 - (1/\beta)]^{1/\beta}$.

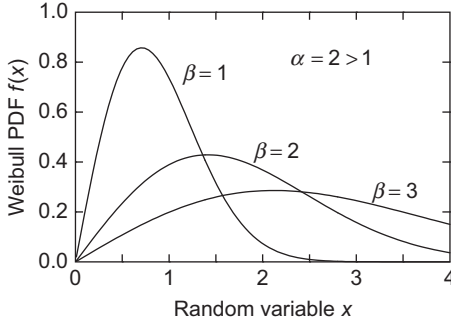


Figure A.8 The PDF for the Weibull distribution for $\alpha > 1$.

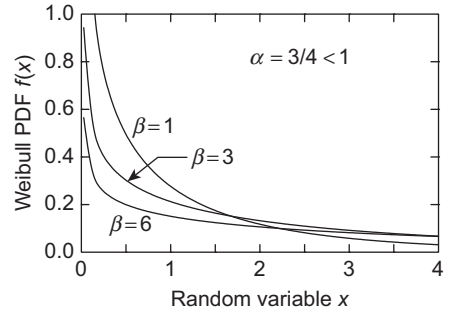


Figure A.9 The PDF for the Weibull distribution for $\alpha < 1$.

Sampling from a Weibull distribution

To generate samples from a Weibull PDF, the inverse CDF method is quite efficient, namely,

$$x_i = -\beta[\ln \rho_i]^{1/\alpha}, \quad (\text{A.35})$$

where the ρ_i are random samples from $\mathcal{U}(0, 1)$.

A.1.6 Normal Distribution

The *normal distribution*, also often called the *Gaussian distribution*, has the PDF

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(\mu-x)^2/(2\sigma^2)}, \quad -\infty < x < \infty \quad (\text{A.36})$$

and is denoted as $\mathcal{N}(\mu, \sigma)$. The mean μ and variance σ^2 are the parameters of this distribution. A special case is the *unit normal* distribution $\mathcal{N}(0, 1)$. Numerical integration of the PDF must be used to obtain the CDF. Thus,

$$F(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-(\mu-u)^2/(2\sigma^2)} du = \frac{1}{2} \left[1 - \text{sign}(x - \mu) \text{erf} \left(\left| \frac{x - \mu}{\sqrt{2}\sigma} \right| \right) \right], \quad (\text{A.37})$$

where $\text{erf}(x)$ is the *error function*

$$\text{erf}(x) \equiv \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt, \quad x > 0. \quad (\text{A.38})$$

The error function is related to the incomplete gamma function of Eq. (A.21) by $\text{erf}(x) = P(1/2, x^2)$, $x \geq 0$. The central moments of the normal distribution are

$$\begin{aligned} \langle (x - \mu)^n \rangle &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} u^n e^{-u^2/(2\sigma^2)} dx \\ &= \begin{cases} 0, & n \text{ odd} \\ 1 \cdot 3 \cdot 5 \cdots (n-1)\sigma^n, & n \text{ even} \end{cases}. \end{aligned} \quad (\text{A.39})$$

A normal distribution in which $\sigma = \sqrt{\mu}$ is a very good approximation to the Poisson distribution with mean μ , especially for large μ .¹ Hence, we often treat processes (like radioactive decay) that behave according to Poisson statistics with the continuous normal distribution *for the special case that* $\sigma = \sqrt{\mu}$.

Sampling from a normal distribution

To generate variates from a normal distribution, first sample y_i from the unit normal $\mathcal{N}(0, 1)$ and then take $x_i = \sigma y_i + \mu$ as a sample from $\mathcal{N}(\mu, \sigma)$. The **Box-Muller** [1958] scheme can be used to sample from the unit normal as follows. Select two random numbers ρ_1 and ρ_2 distributed uniformly on $[0, 1]$. Then *two* unit normal samples are

$$\begin{aligned} y_1 &= \sqrt{-2 \ln \rho_1} \cos 2\pi \rho_2 \\ y_2 &= \sqrt{-2 \ln \rho_1} \sin 2\pi \rho_2 \end{aligned} \quad (\text{A.40})$$

To avoid the computer expense of evaluating trigonometric functions this scheme can be modified. In the so-called *rejection polar method*, two uniform samples ξ_1 and ξ_2 are sampled from the unit disk using the rejection technique of Example 4.4. Then set $r^2 = \xi_1^2 + \xi_2^2$, which is a uniform sample from the unit disk. The unit normal samples are then given by

$$\begin{aligned} y_1 &= \xi_1 \sqrt{-2 \ln(r^2)/r^2} \\ y_2 &= \xi_2 \sqrt{-2 \ln(r^2)/r^2} \end{aligned} \quad (\text{A.41})$$

This polar-rejection method is the same as the Box-Muller method since r^2 can be used as x_1 and the quantities $\xi_1/\sqrt{r^2}$ and $\xi_2/\sqrt{r^2}$ are the cosine and sine in Eq. (A.40). Notice that this Box-Muller algorithm returns two samples from the unit normal.

Finally, to obtain a sample from $\mathcal{N}(\mu, \sigma)$ simply transform one of the unit normal samples y_i to obtain $x_i = \sigma y_i + \mu$ as the sample. Although faster sampling algorithms are available (see, for example, **Ahrens and Dieter** [1988]), that of Eq. (A.41) is sufficient for most purposes.

¹ For most purposes, this approximation is quite good if $\mu > 25$.

A.1.7 Lognormal Distribution

In many physical systems, the logarithm of some quantity of interest is found to be distributed normally, i.e., $\ln x \sim \mathcal{N}(\alpha, \beta)$. For example, the radon concentration in U.S. homes is observed to have a lognormal distribution [Faw and Shultis 1999]. The lognormal distribution has the PDF

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-(\ln x - \alpha)^2/(2\beta^2)}, \quad 0 \leq x < \infty \quad (\text{A.42})$$

and is denoted by $\mathcal{LN}(\alpha, \beta)$. Its CDF can be expressed as

$$F(x) = \frac{1}{2} \left[1 + \text{sign}(\ln x - \alpha) \text{erf} \left(\frac{\ln x - \alpha}{\sqrt{2}} \beta \right) \right]. \quad (\text{A.43})$$

Examples of the lognormal distribution are shown in Fig. A.10.

The moments of the lognormal distribution about zero are

$$\langle x^n \rangle = e^{(n\alpha + n^2\beta^2)/2}, \quad (\text{A.44})$$

from which the mean and variance are found to be

$$\mu = e^{(\alpha + \beta^2)/2} \quad (\text{A.45})$$

and

$$\sigma^2 = \langle x^2 \rangle - \mu^2 = (e^{\beta^2} - 1)e^{2\alpha + \beta^2}, \quad (\text{A.46})$$

respectively.

Sampling from the lognormal distribution

To generate samples from a lognormal distribution $\mathcal{LN}(\alpha, \beta)$, first generate x_i from the unit normal $\mathcal{N}(0, 1)$ and set $y_i = \alpha x_i + \beta$. Finally, use $z_i = e^{y_i}$ as the desired sample.

A.1.8 Cauchy Distribution

The *Cauchy distribution* $\mathcal{C}(\alpha, \beta)$, also sometimes called the *Lorentzian distribution*, is often used to describe the behavior of resonance phenomena. The Lorentzian PDF can be written as

$$f(x) = \frac{1}{\pi} \frac{\beta}{(x - \alpha)^2 + \beta^2}, \quad -\infty < \alpha < \infty, \quad \beta > 0, \quad -\infty < x < \infty. \quad (\text{A.47})$$

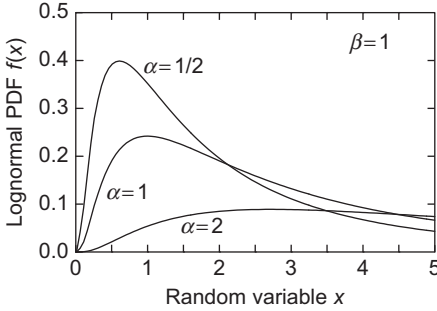


Figure A.10 The PDF for the lognormal distribution.

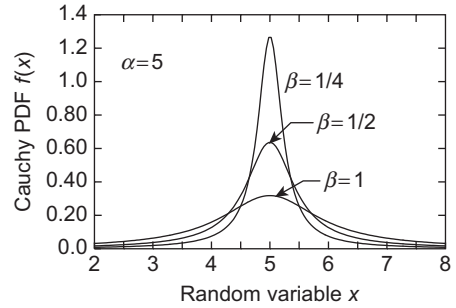


Figure A.11 The PDF for the Cauchy distribution.

The corresponding CDF is

$$F(x) = \frac{1}{2} + \frac{1}{\pi} \tan^{-1} \left(\frac{x - \alpha}{\beta} \right). \quad (\text{A.48})$$

Examples of the Cauchy distribution are shown in [Fig. A.11](#). The α parameter determines the location of the resonance peak, and the width parameter β is twice the peak's full-width-at-half maximum (FWHM).

Unlike most continuous distributions, the mean and variance are undefined, since the moments

$$\langle x^n \rangle = \frac{\beta}{\pi} \int_{-\infty}^{\infty} \frac{x^n}{(x - \alpha)^2 + \beta^2} dx \quad (\text{A.49})$$

diverge for all $n \geq 1$.

Sampling from the Cauchy distribution

The inverse CDF method can be used to generate samples from the Cauchy distribution, namely,

$$x_i = \alpha + \beta \tan \pi(\rho_i - 1/2), \quad (\text{A.50})$$

where ρ_i is from $\mathcal{U}(0, 1)$. To avoid the expense of the tangent evaluation, take two samples y_1 and y_2 from the unit normal $\mathcal{N}(0, 1)$. The ratio $z = y_1/y_2$ is then from the unit Cauchy distribution $\mathcal{C}(0, 1)$, so that $x = \alpha + \beta z$ is from $\mathcal{C}(\alpha, \beta)$. Alternatively, sample y_1 and y_2 uniformly from a disk with diameter one (see Example 4.4), and the ratio y_1/y_2 is from $\mathcal{C}(0, 1)$. Another fast sampling scheme is offered by [Kronmal and Peterson \[1981\]](#).

A.1.9 Chi-Squared Distribution

The chi-squared statistic can be defined as

$$\chi^2 = \sum_{i=1}^N \frac{(x_i - \mu)^2}{\sigma^2}, \quad (\text{A.51})$$

where x_i is the i th sample of x from the distribution that has mean μ and variance σ^2 and N is the number of samples. Use of this statistic in data analysis is widespread because the “reduced” chi-square, defined as

$$\chi_v^2 = \frac{\chi^2}{\nu}, \quad (\text{A.52})$$

where ν is the number of degrees of freedom, approaches unity if the “data” are drawn from the correct distribution with mean μ and variance σ^2 . For instance, if data d_i are fit to a model $m(x_i)$ and the square of the deviations of the data from the model value (the numerator $[d_i - m(x)]^2$) is about the same as the variance in the data (the denominator $\sigma^2(d_i)$), normalized by the number of degrees of freedom (the factor $1/\nu$), then χ_v^2 will be approximately unity and one can conclude that the model is a good fit to the data. If the model is not a good fit, the numerator will generally exceed the denominator and χ_v^2 will be greater than unity. This concept has been used widely in data analysis as, for instance, in the rolling-window template-matching procedure of Dunn [2004].

The degrees of freedom ν is simply a measure of the number of samples, N , less the number of constraints c or

$$\nu = N - c. \quad (\text{A.53})$$

There are no constraints when only the data themselves are used to determine a quantity, but one constraint when a value depending on the data is used to compute a quantity. This accounts for why the sample mean is given by

$$\langle x \rangle = \frac{1}{N} \sum_{i=1}^N x_i, \quad (\text{A.54})$$

since only the N data points are used, whereas the sample variance is given by

$$\sigma^2 = \frac{1}{N-1} \sum_{i=1}^N [x_i - \hat{x}]^2. \quad (\text{A.55})$$

Both of these quantities can be written as sums divided by ν , where in the former case $\nu = N$ ($c = 0$), and in the latter $\nu = N - 1$ ($c = 1$). See Section 2.1.5 for a more complete discussion of degrees of freedom.

For the chi-squared statistic defined by Eq. (A.51), the *chi-squared* PDF is

$$f(\chi^2|\nu) = \left[2^{\nu/2}\Gamma\left(\frac{\nu}{2}\right)\right]^{-1} (\chi^2)^{\nu/2-1} e^{-\chi^2/2}, \quad \nu > 0 \quad x \geq 0 \quad (\text{A.56})$$

and is denoted by $\chi^2(\nu)$. This is simply a special case of the gamma distribution $\mathcal{G}(\alpha, \beta)$, namely,

$$f(\chi^2|\nu) = \mathcal{G}(\nu/2, 2). \quad (\text{A.57})$$

Hence, the results obtained for the gamma distribution in Section A.1.3 can be used to obtain the moments and CDF for $\chi^2(\nu)$. Also, sampling methods for the gamma distribution can be used to obtain samples from the chi-squared distribution.

The mean value of the chi-squared distribution is $\mu = \nu$, and the variance is $\sigma^2 = 2\nu$. For large ν (> 30), the chi-squared distribution is well approximated by a normal distribution with mean ν and variance 2ν . See Fig. A.12 for examples of $\chi^2(\nu)$ and the approximating normal distribution for $\nu = 40$.

A.1.10 Student's *t* Distribution

The central limit theorem states that, for very large samples, the distribution of the deviations between the sample and population means is described by the unit normal distribution. However, the behavior of smaller samples is only “semi-normal.” William Gosset, who worked for the Guinness brewery, developed an interesting alternative to the normal distribution. His employer did not want him publishing his findings as an identified employee and so he published several seminal statistical works under the pseudonym “Student.” As a result, a distribution he developed is known today as Student's *t* distribution.

Gosset recognized that for small populations the normal distribution overestimates the achieved results near the mean and underestimates them far from the mean. He thus devised a distribution that has the desirable properties of (1) more closely fitting the results of finite samples and (2) containing an adjustable parameter.

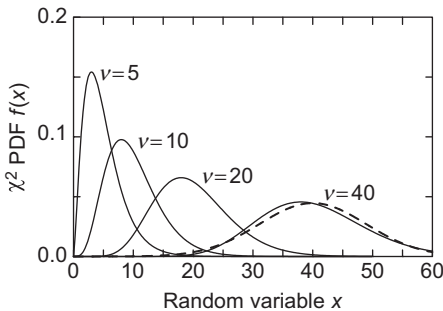


Figure A.12 The PDF for the chi-squared distribution. Shown by the heavy dashed line is the approximating normal distribution.

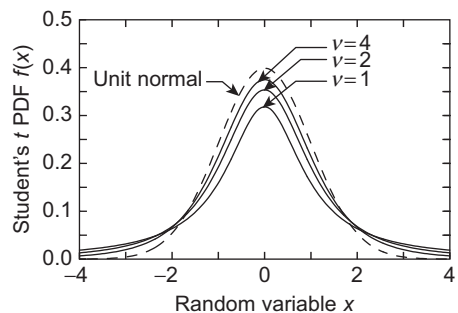


Figure A.13 The PDF for the Student's *t* distribution. The heavy dashed line is the limiting unit normal distribution.

Student's t distribution, denoted by $\mathcal{S}_t(\nu)$, has a PDF of the form

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\nu}} \left(1 + \frac{x^2}{\nu}\right)^{-(\nu+1)/2}, \quad \nu > 0 \quad -\infty < x < \infty, \quad (\text{A.58})$$

where the parameter ν is known as the number of degrees of freedom. This function is symmetric about $x=0$, unimodal, and extends to infinity in both directions. The CDF can be written in terms of the incomplete beta function defined by Eq. (A.25) as [Kendall and Stuart 1977]

$$F(x) = 1 - \frac{1}{2}I_{x^*}\left(\frac{\nu}{2}, \frac{1}{2}\right), \quad (\text{A.59})$$

where $x^* = [1 + (x/\nu)]^{-1}$. Moments $\langle x^n \rangle$ about zero exist only for $n < \nu$. Because of the symmetry of $f(x)$, all odd-order moments vanish. The even moments are

$$\langle x^{2n} \rangle = \nu^n \frac{\Gamma(n + \frac{1}{2})\Gamma(\frac{\nu}{2} - n)}{\Gamma(\frac{1}{2})\Gamma(\frac{\nu}{2})}, \quad 2n < \nu. \quad (\text{A.60})$$

When $\nu = 1$, Student's t distribution is the same as the unit Cauchy distribution $\mathcal{C}(0, 1)$. Also, as $\nu \rightarrow \infty$, Student's t distribution becomes the unit normal $\mathcal{N}(0, 1)$. In Fig. A.13, examples of Student's t distribution are shown. For $\nu \gtrsim 10$, the unit normal is a good approximation for $\mathcal{S}_t(\nu)$.

Sampling from student's t distribution

If generators are available for the unit normal and the gamma distributions, a convenient way to generate samples from Student's t distribution is as follows [Fishman 2003]: sample y_1 and y_2 from $\mathcal{N}(0, 1)$ and $\mathcal{G}(\nu/2, 1/2)$, respectively; then $x = y_1/\sqrt{y_2}$ is a sample from $\mathcal{S}_t(\nu)$. Other faster, but more complex, generators are available (see Kinderman and Monahan [1980], Marsaglia [1984], and Stadlober [1982]).

A.1.11 Pareto Distribution

The PDF of the Pareto distribution $\mathcal{P}_a(\alpha, \beta)$ is given by

$$f(x) = \frac{\alpha\beta^\alpha}{x^{\alpha+1}}, \quad \alpha > 0, \quad x \geq \beta > 0, \quad (\text{A.61})$$

and the CDF has the form

$$F(x) = 1 - \left[\frac{\beta}{x}\right]^\alpha. \quad (\text{A.62})$$

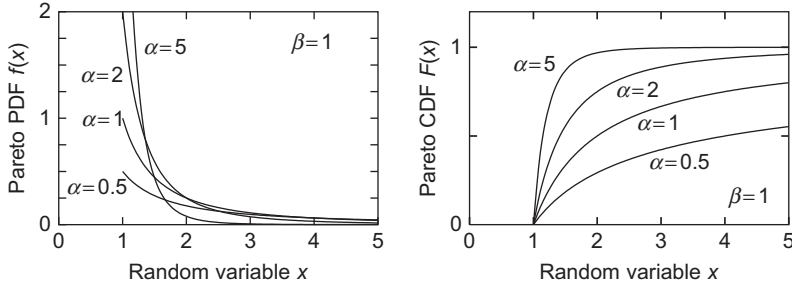


Figure A.14 The PDF and CDF of the Pareto distribution for $\beta = 1$.

This distribution is a power-law distribution, which has been applied in many fields of the physical and social sciences. Sometimes it is referred to as the *Bradford* distribution. The moments of the Pareto distribution for $\alpha > n$ can be expressed as

$$\overline{x^n} = \frac{\alpha \beta^n}{\alpha - n}, \quad (\text{A.63})$$

from which it is easy to show that the mean value is given by

$$\mu = \frac{\alpha \beta}{\alpha - 1}, \quad (\text{A.64})$$

and the variance has the form

$$\sigma^2(x) = \frac{\alpha \beta^2}{(\alpha - 1)^2(\alpha - 2)}. \quad (\text{A.65})$$

The PDF and CDF of the Pareto distribution for $\beta = 1$ are shown in Fig. A.14 for four values of α .

Sampling from the Pareto distribution

Straightforward sampling can be used to obtain samples from the Pareto distribution. From Table A.1, it follows that

$$x_i = \frac{\beta}{(1 - \rho_i)^{1/\alpha}}, \quad (\text{A.66})$$

where ρ_i is a pseudorandom number selected uniformly from the unit interval.

A.2 Discrete Distributions

Five common discrete distributions are presented: the Bernoulli, the binomial, the geometric, the negative binomial, and the Poisson. More thorough coverage of these and

Table A.2 Summary of properties of the inverse CDF method for four discrete distributions (after Fishman [2003]).

Distribution	PDF $f(x)$	$f(0)$	$R(x) =$ $f(x+1)/f(x)$	Mean No. of Trials
binomial $\mathcal{B}(n, p)$ $x=0, 1, \dots, n; 0 < p < 1$	$\binom{n}{x} p^x (1-p)^{n-x}$	$(1-p)^n$	$\frac{(n-x)p}{(x+1)(1-p)}$	$1 + np$
geometric $\mathcal{G}_e(p)$ $x=0, 1, \dots; 0 < p < 1$	$p(1-p)^x$	p	$1-p$	$1 + \frac{1-p}{p}$
neg. binomial $\mathcal{NB}(s, p)$ $x=0, 1, \dots; 0 < p < 1$	$\binom{x+s-1}{x} p^s (1-p)^x$	p^s	$\frac{s+x}{x+1} (1-p)$	$1 + \frac{s(1-p)}{p}$
Poisson $\mathcal{P}(\mu)$ $x=0, 1, \dots; \mu > 0$	$\frac{\mu^x}{x!} e^{-\mu}$	$e^{-\mu}$	$\frac{\mu}{x+1}$	$1 + \mu$

other discrete distributions is provided by Johnson et al. [1992]. For many discrete distributions the inverse CDF method is a simple, although not always the most efficient, method for sampling from these distributions. In this method (see Section 4.1.1), one needs the first density function $f(0)$ in the sequence $f(x)$, $x = 0, 1, 2, \dots$ and the ratio between successive values, i.e., $R(x) = f(x+1)/f(x)$. In Table A.2, the results of the inverse CDF method are summarized for the five discrete distributions considered here.

A.2.1 Bernoulli Distribution

The Bernoulli distribution applies to a variable that can assume either of two values. The value $x = 1$ is assigned to one outcome that is called a “success” and the value $x = 0$ is given to the other outcome that is called a “failure.” Let $0 < p < 1$ be the probability of success and $q = 1 - p$ the probability of failure. An obvious example is flipping a (fair) coin, for which $p = q = 0.5$, and success can be assigned to either “heads” or “tails.” Another example might be a baseball player getting a hit, given that his or her average is 300 (or $p = 0.30$). In any given at-bat, a hit is a success, $x = 1$, and an out is a failure, $x = 0$ (a walk doesn’t count).

The Bernoulli PDF $f(x)$ is defined only for $x = 0$ and $x = 1$ and is given by

$$\begin{aligned} f_1 &= 1 - p, & \text{associated with } x = 0 \\ f_2 &= p, & \text{associated with } x = 1 \end{aligned} \quad (\text{A.67})$$

Another way to express this is

$$f(x) = (1-p)\delta(x) + p\delta(x-1). \quad (\text{A.68})$$

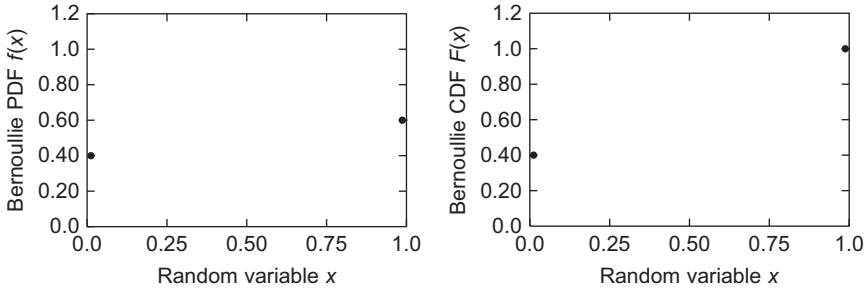


Figure A.15 The PDF and the CDF for the Bernoulli distribution with parameter $p = 0.6$. There are only two outcomes, $x = 0$ (with probability $1 - p$) and $x = 1$ (with probability p).

The CDF is simply

$$\begin{aligned} F_1 &= 1 - p, & \text{associated with } x = 0 \\ F_2 &= 1, & \text{associated with } x = 1 \end{aligned} \quad (\text{A.69})$$

The mean value is

$$\langle x \rangle \equiv \mu \equiv \sum_{i=1}^2 x_i f_i = 0(1 - p) + 1(p) = p, \quad (\text{A.70})$$

and the variance is

$$\sigma^2(x) \equiv \sum_{i=1}^2 (x_i - \mu)^2 f_i = p^2(1 - p) + (1 - p)^2 p = p(1 - p). \quad (\text{A.71})$$

The PDF and CDF for a Bernoulli distribution are shown in [Fig. A.15](#).

The Bernoulli distribution is important because it gives rise to the concept of “Bernoulli trials,” in which independent experiments are run, each presuming a constant success rate, p . The baseball example is interesting because the value of p is not known until after the season (or even career). However, one would expect a hitter with a cumulative average of 243 to get a hit about once every four times at-bat and a 330-hitter to get a hit about once every three times at-bat. Of course, the result of any given at-bat cannot be predicted with certainty.

A.2.2 Binomial Distribution

Another basic probability distribution is the binomial distribution. If one conducts n trials of an experiment that has probability p of success (such as rolling a 3 with a pair of dice or getting 100 counts in a minute with a radiation detector), then the probability density function that x successes are obtained in n trials is given by the

binomial distribution

$$f(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}, \quad x = 0, 1, 2, \dots, n, \quad (\text{A.72})$$

denoted by $\mathcal{B}(n, p)$. The fraction in this distribution is sometimes called the *binomial coefficient* and is often given the shorthand notation

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}. \quad (\text{A.73})$$

Note that the binomial distribution is not defined for noninteger or negative values of x . The binomial CDF is given by

$$F(x) = \sum_{m=0}^x \frac{n!}{m!(n-m)!} p^m (1-p)^{n-m}, \quad x = 0, 1, 2, \dots, n, \quad (\text{A.74})$$

with $F(n) = 1$.

The mean or average value is defined as

$$\langle x \rangle = \mu = \sum_{x=0}^n x f(x|n, p) = \sum_{x=0}^n \frac{xn!}{x!(n-x)!} p^x (1-p)^{n-x}. \quad (\text{A.75})$$

Let $y = x - 1$ and $m = n - 1$; then $x = y + 1$, $n = m + 1$, and it is easy to show that $n - x = m - y$. Thus,

$$\begin{aligned} \mu &= \sum_{y=0}^m \frac{(y+1)(m+1)!}{(y+1)!(m-y)!} p^{y+1} (1-p)^{m-y} \\ &= \sum_{y=0}^m \frac{(m+1)m!}{y!(m-y)!} p p^y (1-p)^{m-y} \\ &= (m+1)p \sum_{y=0}^m \frac{m!}{y!(m-y)!} p^y (1-p)^{m-y} \\ &= np \sum_{y=0}^m f(y|m, p). \end{aligned} \quad (\text{A.76})$$

Note that the summations on y begin at $y = 0$ because the binomial PDF is not defined for negative values of the argument. Because the summation over all outcomes of a discrete PDF is unity, we see that the mean is given simply as

$$\mu = np. \quad (\text{A.77})$$

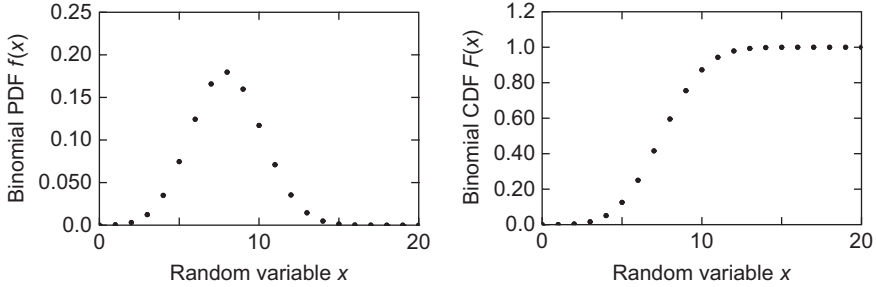


Figure A.16 The PDF and the CDF for the binomial distribution with parameters $n = 20$ and $p = 0.4$.

In a similar manner the variance of the binomial distribution is found to be

$$\sigma^2 = np(1 - p) = npq. \quad (\text{A.78})$$

An example of the binomial distribution is shown in [Fig. A.16](#).

Application to radioactive decay

One reason the binomial distribution is of importance is that it applies to radioactive decay and to radiation counting. If a detector is placed in a field of radiation particles, such as in the core of a fission reactor, the stochastic processes of radiation emission, transport, and detection cause the number of “counts” registered by the detector over some time period to be governed by binomial statistics. This statement means that if one were to repeat the same measurement under the same conditions over and over, the results of the various trials would be distributed according to the binomial distribution. Later, in [Section A.2.5](#), it is shown that, under certain circumstances, the binomial distribution is well approximated by the Poisson distribution. Consequently, radiation counting is almost always treated by Poisson statistics because it has the convenient property that the variance is equal to the mean.

Sampling from a binomial distribution

The simplest way to generate a random sample from a binomial distribution is to conduct n independent Bernoulli trials, each with probability of success p , and let x be the number of successes. By definition, the x is a random sample from the binomial distribution. This method works well provided n is not very large, since the time to generate one sample is clearly proportional to n .

Alternatively, one could use the inverse CDF method (see [Section 4.1.1](#)). For the binomial $f(0) = (1 - p)^n$, the ratio of successive values of the PDF is

$$R(x) \equiv \frac{f(x+1)}{f(x)} = \frac{n-x}{x+1} \frac{p}{1-p}. \quad (\text{A.79})$$

Because the inverse CDF method is essentially a table search, for large n , starting the search at the mode (maximum) of the binomial reduces the time by more than half compared to a sequential search starting from $x = 0$ [Kemp 1986].

For values of $n(1-p) \gtrsim 500$, other methods are available that are not so dependent on n [see Stadlober 1991; Kachitvichyanukul and Schmeiser 1988].

A.2.3 Geometric Distribution

Consider a sequence of Bernoulli trials in which p is the probability of success. Then the probability of obtaining x failures before getting the first success on trial $x + 1$ is

$$f(x) = p(1-p)^x, \quad 0 < p < 1, \quad x = 0, 1, 2, \dots \quad (\text{A.80})$$

This is known as the *geometric distribution* and is denoted by $\mathcal{G}_e(p)$. The corresponding CDF is

$$F(x) = p \sum_{y=0}^x (1-p)^y = 1 - (1-p)^{x+1}, \quad (\text{A.81})$$

and the mean and variance are $\mu = 1/p$ and $\sigma^2 = (1-p)/p^2$, respectively.

Sampling from a geometric distribution

To generate samples from $\mathcal{G}_e(p)$ one of the best ways is to use the property that if x is a random variate from the exponential distribution $\mathcal{E}(\lambda)$ (see Section A.1.2) with $\lambda \equiv -1/\ln p$, then x is distributed as $\mathcal{G}_e(p)$. Thus, generate ρ_i from $\mathcal{U}(0, 1)$ and use as the exponential variate²

$$x_i = \left\lfloor \frac{\ln \rho_i}{\ln p} \right\rfloor. \quad (\text{A.82})$$

Alternatively, one could use the inverse CDF method (see Section 4.1.1) in which the starting density value is $f(0) = p$ and the ratio of successive values is $R(x) = f(x+1)/f(x) = (1-p)$. Although this is simple to implement it can be very inefficient if p is very small.

A.2.4 Negative Binomial Distribution

Closely related to the geometric distribution is the *negative binomial* distribution, which gives the probability of x failures prior to obtaining s successes in a sequence of Bernoulli trials. Suppose that p is the probability of success. Then the probability of

² Here, the *floor function* $\lfloor x \rfloor$ is used to indicate the largest integer $\leq x$. Thus, $\lfloor 5.2 \rfloor = 5$. By contrast, the *ceiling function* $\lceil x \rceil$ gives the smallest integer $\geq x$.

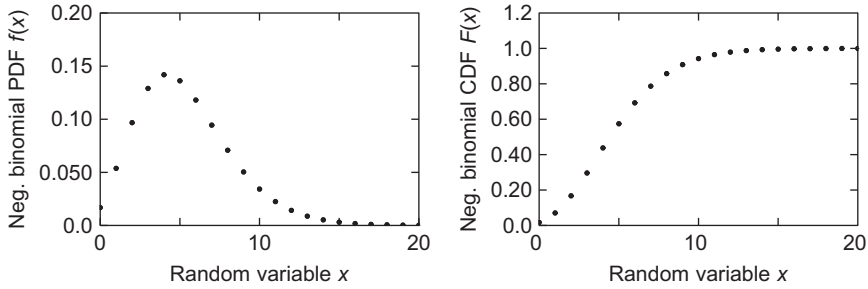


Figure A.17 The PDF and the CDF for the negative binomial distribution with parameters $p = 0.6$ and $s = 8$.

obtaining x failures before obtaining the s th success on the $x + s$ trial is

$$f(x) = \frac{(s+x-1)!}{x!(s-1)!} p^s (1-p)^x, \quad x = 0, 1, 2, \dots \quad (\text{A.83})$$

This distribution is denoted by $\mathcal{NB}(p)$. Its CDF is

$$F(x) = \sum_{y=0}^x \frac{s+y-1}{y!(s-1)!} p^s (1-p)^y, \quad x = 0, 1, 2, \dots \quad (\text{A.84})$$

Note that when $s = 1$, the negative binomial is just the geometric distribution. However, the parameter s in Eq. (A.83) need not be a positive integer, and, in general, can be any real positive number. When s is integral, then Eq. (A.83) is also known as the *Pascal* distribution. It can be shown that the mean value is $\mu = \langle x \rangle = s(1-p)/p$, and the variance is $\sigma^2 = \langle x^2 - \mu^2 \rangle = s(1-p)/p^2$. An example of the negative binomial distribution is shown in Fig. A.17.

Sampling from a negative binomial

The inverse CDF method of Section 4.1.1, with $f(0) = p^s$ and the ratio $R(x) \equiv f(x+1)/f(x) = (s+x)(1-p)/(x+1)$ can be used. However, this approach becomes very inefficient as p becomes very small. Alternatively, if generators for the Poisson distribution $\mathcal{P}(\mu)$ (see next section) and for the gamma distribution $\mathcal{G}(\alpha, \beta)$ (see Section A.1.3) are available, proceed as follows. First, generate x_i from $\mathcal{G}(s, 1)$ and then from $\mathcal{P}(x_i(1-p)/p)$ generate y_i , which is distributed according to the negative binomial $\mathcal{NB}(p)$.

A.2.5 Poisson Distribution

When n is large, the evaluation of the binomial distribution of Eq. (A.72) becomes computationally difficult because the various factorials in the binomial coefficient become enormous. However, when p is very small and n very large, the binomial

distribution can be well approximated by the Poisson distribution, which is computationally much simpler. To see this simplification, consider Eq. (A.72) for a fixed x and let $n \rightarrow \infty$ and $p \rightarrow 0$ in such a manner that np remains constant, say, $np = \mu$. Then with $p = \mu/n$, the limit of Eq. (A.72) yields

$$\begin{aligned} \lim_{n \rightarrow \infty} f_B(x) &= \lim_{n \rightarrow \infty} \frac{n!}{x!(n-x)!} \left(\frac{\mu}{n}\right)^x \left(1 - \frac{\mu}{n}\right)^{n-x} \\ &= \lim_{n \rightarrow \infty} \frac{n(n-1) \cdots (n-x+1)}{n \cdots n} \frac{\mu^x}{x!} \left(1 - \frac{\mu}{n}\right)^n \left(1 - \frac{\mu}{n}\right)^{-x} \\ &= \lim_{n \rightarrow \infty} \frac{\left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{x-1}{n}\right)}{\left(1 - \frac{\mu}{n}\right)^x} \frac{\mu^x}{x!} \left(1 - \frac{\mu}{n}\right)^n. \end{aligned} \quad (\text{A.85})$$

As $n \rightarrow \infty$, the terms $(1 - i/n)$ and $(1 - \mu/n)$ approach unity. Also, from the properties of the base of the natural logarithms e , it is known that

$$\lim_{n \rightarrow \infty} \left(1 - \frac{\mu}{n}\right)^n = e^{-\mu}. \quad (\text{A.86})$$

Hence, Eq. (A.85), in the limit, yields

$$\lim_{n \rightarrow \infty} f_B(x) = \frac{\mu^x}{x!} e^{-\mu}, \quad x = 0, 1, 2, \dots \quad (\text{A.87})$$

This is the PDF $f(x)$ of the *Poisson distribution*, denoted by $\mathcal{P}(\mu)$, which has only a single parameter μ . In Fig. A.18, a comparison of the binomial PDF and its approximating Poisson PDF is shown. It is seen that the two distributions are very nearly the same, even for this rather large value of $p = 0.1$. For many applications the Poisson approximation is adequate with n as small as 20 provided $\mu \equiv np < 5$.

The CDF for the Poisson distribution can be written as

$$F(x) = \sum_{k=0}^x \frac{\mu^k}{k!} e^{-\mu}, \quad x \geq 0, \quad (\text{A.88})$$

which has the limiting value $F(\infty) = 1$.

For a large number of trials, p represents the proportion of time an event (success) occurs. Then for a large number of trials n , the *average* or *expected* number of successes would be $np \equiv \mu$. Thus, the expected value $\langle x \rangle = \mu$, which is the single parameter of the Poisson distribution. This result can be derived more formally as follows:

$$\langle x \rangle = \sum_{x=0}^{\infty} x \frac{\mu^x}{x!} e^{-\mu} = \sum_{x=1}^{\infty} \frac{\mu^x}{(x-1)!} e^{-\mu} = \mu \sum_{y=0}^{\infty} \frac{\mu^y}{y!} e^{-\mu} = \mu F(\infty) = \mu. \quad (\text{A.89})$$

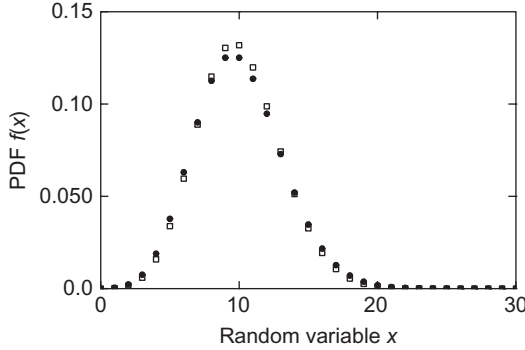


Figure A.18 Comparison of PDFs for the Poisson distribution (circles) with $\mu = 10$ and the binomial PDF (open squares) with $n = 100$ and $p = 0.1$ (and hence $\mu = np = 10$).

In a similar fashion, the variance of the Poisson distribution is found to be

$$\sigma^2(x) = \mu. \quad (\text{A.90})$$

This is a very important property of Poisson statistics, namely, that the variance is equal to its mean, i.e., $\sigma^2 = \mu$. As is necessary from the Poisson being the limiting distribution for the binomial distribution, this result is in agreement with the limiting value of the binomial variance for large n , p small, and $np = \mu$. From Eq. (A.78) one has

$$\sigma_B^2 = np(1 - p) = \mu(1 - p) = \lim_{p \rightarrow 0} \mu(1 - p) = \sigma_P^2. \quad (\text{A.91})$$

The Poisson distribution has several interesting properties [Fishman 2003]:

1. If x_i are independent samples from $\mathcal{E}(1)$ and y is the smallest integer such that $\sum_{i=1}^{y+1} x_i > \mu$, then y is distributed as $\mathcal{P}(\mu)$.
2. If ρ_i are independent samples from $\mathcal{U}(0, 1)$ and y is the smallest integer such that $\prod_{i=1}^{y+1} \rho_i < e^{-\mu}$, then y is distributed according to the Poisson distribution $\mathcal{P}(\mu)$.
3. If x is from $\mathcal{P}(\mu)$, then $(x - \mu)/\sqrt{\mu}$ has a distribution that converges to $\mathcal{N}(0, 1)$ as $\mu \rightarrow \infty$.

Sampling from the Poisson distribution

Often property 2 of the properties of the Poisson distribution has been used to generate Poisson deviates. However, the inverse CDF method has comparable speed because of the large number of samples from $\mathcal{U}(0, 1)$ required by property 2 [Schmeiser and Kachitvichyanukul 1981]. In the inverse CDF method (see Section 4.1.1), the initial value $f(0) = e^{-\mu}$ and the ratio of successive PDF values is $R(x) \equiv f(x + 1)/f(x) = \mu/(x + 1)$. The inverse CDF method has a computational time proportional to μ .

For large values of μ , there are generating schemes that are relatively insensitive to μ , all of which are rather complex [Atkinson 1979; Devroye 1981; Schmeiser and Kachitvichyanukul 1981; and Ahrens and Dieter [1991]].

A.3 Joint Distributions

Finally, two examples of joint or multivariate distributions are presented, the multivariate normal distribution and the multinomial distribution. Other joint distributions are covered extensively in, for example, Johnson and Kotz [1987].

Sampling from multivariate distributions is often done by writing the joint distribution as a product of the marginal distribution of one of the variates and conditional distributions for the other variates. Each univariate distribution is then sampled sequentially to create a sample. This is the approach discussed earlier in Section 4.1.9. This approach is especially popular for analyses based on Markov chain Monte Carlo methods. An alternative sampling approach, as illustrated in the following, is to build a sample by a direct transformation of a vector of independent identically distributed scalar variates, as illustrated.

A.3.1 Multivariate Normal Distribution

The n -dimensional multinormal distribution with mean vector $\boldsymbol{\mu}$ and a nonsingular symmetric covariance matrix $\boldsymbol{\Sigma}$ has the probability density function

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} \sqrt{|\boldsymbol{\Sigma}|}} \exp \left(-[\mathbf{x} - \boldsymbol{\mu}]^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) / 2 \right). \quad (\text{A.92})$$

A direct way of generating random vectors $\mathbf{x}$ from this PDF is to generate a vector $\mathbf{y} = (y_1, y_2, \dots, y_n)$, where each y_i is a sample from the unit normal distribution $\mathcal{N}(0, 1)$. A sample of $\mathbf{x}$ is then formed by the transformation

$$\mathbf{x} = \mathbf{T}^T \mathbf{y} + \boldsymbol{\Sigma}, \quad (\text{A.93})$$

where $\mathbf{T}$ is an $n \times n$ matrix such that $\mathbf{T}^T \mathbf{T} = \boldsymbol{\Sigma}$ and can be calculated by a Cholesky decomposition of $\boldsymbol{\Sigma}$ [Press et al. 1996].

Another approach for generating samples from a multinormal distribution is to first generate x_1 from $\mathcal{N}(\mu_1, \sigma_1)$, then generate x_2 conditionally on x_1 , after which, generate x_3 conditionally on x_1 and x_2 , and so on.

For the special case of $n = 2$, the bivariate normal distribution is obtained. The covariance matrix for this case is

$$\begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}, \quad (\text{A.94})$$

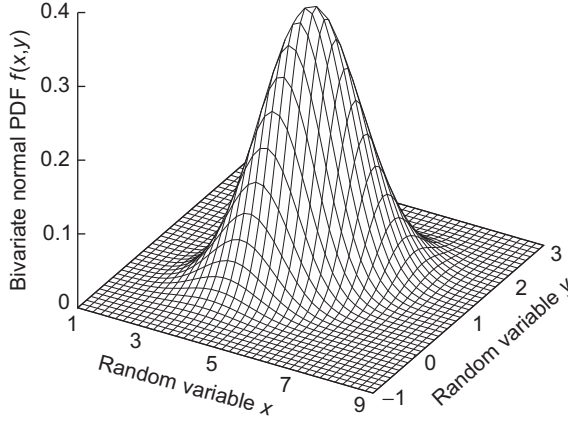


Figure A.19 The PDF for the bivariate normal distribution. For this example, $\mu_1 = 5$, $\mu_2 = 1$, $\sigma_1 = 1$, $\sigma_2 = 0.5$, and $\rho = 0.5$.

where ρ is the correlation coefficient defined as $\text{covar}(x_1, x_2)/\sigma_1\sigma_2$. Then Σ^{-1} is given by

$$\Sigma^{-1} = \begin{pmatrix} \frac{1}{\sigma_1^2(1-\rho^2)} & \frac{-\rho}{\sigma_1\sigma_2(1-\rho^2)} \\ \frac{-\rho}{\sigma_1\sigma_2(1-\rho^2)} & \frac{1}{\sigma_2^2(1-\rho^2)} \end{pmatrix}. \quad (\text{A.95})$$

With this result, Eq. (A.92) reduces to

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{\xi}{2(1-\rho^2)}\right\}, \quad (\text{A.96})$$

where

$$\xi = \frac{(x_1 - \mu_1)^2}{\sigma_1^2} - \frac{2\rho(x_1 - \mu_1)(x_2 - \mu_2)}{\sigma_1\sigma_2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2}. \quad (\text{A.97})$$

An example of the bivariate normal is shown in Fig. A.19.

A.3.2 Multinomial Distribution

Consider a situation in which k mutually exclusive events $E_1, E_2, \dots, E_k$ must occur with p_i representing the probability event E_i occurs. Because an event must occur, $\sum_{i=1}^k p_i = 1$. The joint distribution of the random variables $\mathbf{x} = (x_1, x_2, \dots, x_k)$, where

x_i is the number of occurrences of event E_i in n independent trials with $\sum_{i=1}^k x_i = n$ is given by the multinomial distribution

$$f(\mathbf{x}) = n! \prod_{i=1}^k \frac{p_i^{x_i}}{x_i!}. \quad (\text{A.98})$$

This distribution is a generalization of the binomial distribution (see [Section A.2.2](#)) and reduces to [Eq. \(A.72\)](#) when $k = 2$.

An important property is that the marginal and conditional distributions of any x_i are binomial distributions. Thus, to generate a multinomial sample, one can simply sample, sequentially, from the binomial marginal and conditional distributions. For the best efficiency, the first marginal considered should be that for the x_i with the largest p_i .

Bibliography

- AHRENS, J.H. AND U. DIETER 1974, "Computer Methods for Sampling from Gamma, Beta, Poisson, and Binomial Distributions," *Computing*, **25**, 193–208.
- AHRENS, J.H. AND U. DIETER 1988, "Efficient Table-Free Sampling Methods for the Exponential, Cauchy and Normal Distributions," *Commun. ACM*, **31**, 1330–1337.
- AHRENS, J.H. AND U. DIETER 1991, "A Convenient Sampling Method with Bounded Computing Times for Poisson Distributions," *The Frontiers of Statistical Computation, Simulation and Modeling*, P.R. Nelson, E.J. Dudewicz, A. Öztürk, and E.C. van der Meulen, eds., American Science Press, Columbus, OH, 137–149.
- ATKINSON, A.C. 1979, "The Computer Generation of Poisson Random Variables," *Appl. Stat.*, **28**, 29–35.
- ATKINSON, A.C. AND J. WHITTAKER 1976, "A Switching Algorithm for the Computer Generation of Beta Random Variables with at Least One Parameter Less than One," *J. R. Stat. Soc., Ser. A-G*, **139**, 462–467.
- BEST, D.J. 1983, "A Note on Gamma Variate Generators with Shape Parameter Less Than Unity," *Computing*, **30**, 152–157.
- BOX, G.E.P. AND M.E. MULLER 1958, "A Note on the Generation of Normal Deviates," *Ann. Math. Stat.*, **29**(2), 610–611.
- CHENG, R.C.H. 1978, "Generating Beta Variates with Non Integral Shape Parameters," *Commun. ACM*, **21**, 290–295.
- CHENG, R.C.H. AND G.M. FEAST 1979, "Some Simple Gamma Variate Generators," *Appl. Stat.*, **28**, 290–295.
- DEVROYE, L. 1981, "The Computer Generation of Poisson Random Variables," *Computing*, **26**, 197–207.
- DEVROYE, L. 1986, *Non-Uniform Random Variate Generation*, Springer-Verlag, New York.
- DUNN, W.L. 2004, "Flaw Detection by X-ray Backscattering Using the Rolling-window Template-matching Procedure," *Appl. Radiat. Isot.*, **61**(6), 217–225.
- FAW, R.E. AND J.K. SHULTIS 1999, *Radiological Assessment: Sources and Doses*, American Nuclear Society, LaGrange Park, IL.
- FISHMAN, G.S. 2003, *Monte Carlo Concepts, Algorithms, and Applications*, Corrected Printing, Springer, New York.

- GENTLE, J.E. 1998, *Random Number Generation and Monte Carlo Methods*, Springer-Verlag, New York.
- HASTINGS, C. 1955, *Approximations for Digital Computers*, p. 192, Princeton University Press, Princeton, NJ.
- JÖHNK, M.D. 1964, "Erzeugung von Betaverteilten und Gammaverteilten Zufallszahlen," *Metrika*, **8**, 5–15.
- JOHNSON, N.L. AND S. KOTZ 1987, *Distributions in Statistics, Continuous Multivariate Distributions*, John Wiley & Sons, New York.
- JOHNSON, N.L., S. KOTZ, AND N. BALAKRISHNAN 1992, *Univariate Discrete Distributions*, 2nd ed., Wiley, New York.
- JOHNSON, N.L., S. KOTZ, AND N. BALAKRISHNAN 1994, *Continuous Univariate Distributions*, Vol. 1, 2nd ed., Wiley, New York.
- JOHNSON, N.L., S. KOTZ, AND N. BALAKRISHNAN 1995, *Continuous Univariate Distributions*, Vol. 2, 2nd ed., Wiley, New York.
- KACHITVICHYANUKUL, V. AND B.C. SCHMEISER 1988, "Binomial Random Variate Generation," *Commun. ACM*, **31**, 216–223.
- KEMP, C.D. 1986, "A Modal Method for Generating Binomial Variables," *Commun. Stat.*, **15**, 805–813.
- KENDALL, M. AND A. STUART 1977, *The Advanced Theory of Statistics*, Vol. 1, 3rd ed., MacMillan, New York.
- KINDERMAN, A.J. AND J.F. MONAHAN 1980, "New Methods for Generating Student's t and Gamma Variables," *Computing*, **25**, 367–377.
- KRONMAL, J.C. AND A.V. PETERSON 1981, "A Variant of the Acceptance/Rejection Method for Generating Random Variables," *J. Am. Stat. Assoc.*, **76**, 446–451.
- MARSAGLIA, G. 1984, "The Exact-Approximation Method for Generating Random Variables in a Computer," *J. Am. Stat. Assoc.*, **79**, 218–221.
- MCGRATH, E.J. AND D.C. IRVING 1975, *Techniques for Efficient Monte Carlo Simulations, Vol. II, Random Number Generation for Selected Probability Distributions*, Report RSIC-38, Radiation Shielding Information Center, Oak Ridge National Lab, Oak Ridge, TN.
- PRESS, W.H., S.A. TEUKOLSKY, W.T. VETTERLING, AND B.P. FLANNERY 1996, *Numerical Recipes*, 2nd ed., Cambridge University Press, Cambridge, England.
- RUBINSTEIN, R.Y. 1981, *Simulation and the Monte Carlo Method*, Wiley, New York.
- SCHMEISER, B.W. AND A.J.G. BABU 1980, "Beta Variate Generation via Exponential Majorizing Functions," *Oper. Res.*, **28**, 719–926. Corrected *Oper. Res.*, **31**, 802.
- SCHMEISER, B.W. AND V. KACHITVICHYANUKUL 1981, *Poisson Random Variate Generation*, Res. Memo., 81–84, School of Industrial Engineering, Purdue Univ., Lafayette, IN.
- SNEDECOR, G.W. AND W.G. COCHRAN 1971, *Statistical Methods*, 6th ed., The Iowa State University Press, Ames.
- STADLOBER, E. 1982, "Generating Student's t Variates by a Modified Rejection Method," *Probability and Statistical Inference*, W. Grossmann, G. Plug and W. Wertz, eds., 349–360, Reidel, Dordrecht, The Netherlands.
- STADLOBER, E. 1991, "Binomial Variate Generation: A Method Based on Ratios of Uniforms," *The Frontiers of Statistical Computation, Simulation and Modeling*, P.R. Nelson, E.J. Dudewicz, A. Öztürk, and E.C. van der Meulen, eds., American Science Press, Columbus, OH, 93–112.
- ZECHNER, H. AND E. STADLOBER 1993, "Generating Beta Variates via Patchwork Rejection," *Computing*, **50**, 1–18.

B The Weak and Strong Laws of Large Numbers

In Chapter 1, it was stated that the history of Monte Carlo dates, in some sense, to 1689 when Jacob Bernoulli first enunciated the law of large numbers [Bernoulli 1713]. Bernoulli spent almost 20 years devising a rigorous proof of what he called his “golden theorem” but which was generally referred to as “Bernoulli’s theorem.” Simèon-Denis Poisson referred to this theorem, in 1835, as “la loi des grands nombres,” and this label “the law of large numbers” is commonly used today.

Two forms of this law, weak and strong, are now distinguished. The differences between these forms are unimportant for Monte Carlo applications, as both state that the sample mean $\bar{x}_N$ approaches the population mean μ in the limit of large N , i.e., $\lim_{N \rightarrow \infty} \bar{x}_N = \mu$. It almost seems intuitive that, as long as the process being simulated has a finite population mean, then the sample mean should approach the true mean as the number of samples increases. The distinction between the two forms of the law has to do with the manner in which this convergence takes place. For the interested reader the distinction between the two is briefly discussed here.

B.1 The Weak Law of Large Numbers

Consider a sequence of N random samples, denoted by $\{x_i\}$, $i = 1, 2, \dots, N$, that are uncorrelated so that $\langle x_k x_j \rangle = \langle x_k \rangle \langle x_j \rangle$ and $\sigma^2(x_k + x_j) = \sigma^2(x_k) + \sigma^2(x_j)$, i.e., the covariance between any two samples x_k and x_j is zero. Another way to state this restriction is that all members of the random sequence are independent and identically distributed. The population mean is

$$\mu = \langle x_i \rangle \equiv E(x_i), \quad (\text{B.1})$$

while the sample mean is given by

$$\bar{x}_N = \frac{1}{N} \sum_{i=1}^N x_i. \quad (\text{B.2})$$

The weak law states that, for all positive values of ϵ , no matter how small,

$$\lim_{N \rightarrow \infty} \text{Prob}\{|\bar{x}_N - \mu| \geq \epsilon\} = 0. \quad (\text{B.3})$$

Proof

To prove the weak law, consider a PDF $f(x)$ with finite mean μ and finite variance σ^2 . If $g(x)$ is any non-negative function and k is any positive constant, then

$$\begin{aligned}\langle g(x) \rangle &= \int_{-\infty}^{\infty} g(x)f(x) dx = \int_{\{x:g(x) \geq k\}} g(x)f(x) dx + \int_{\{x:g(x) < k\}} g(x)f(x) dx \\ &\geq \int_{\{x:g(x) \geq k\}} kf(x) dx \geq k \text{Prob}\{g(x) \geq k\}\end{aligned}$$

or

$$\text{Prob}\{g(x) \geq k\} \leq \langle g(x) \rangle / k. \quad (\text{B.4})$$

In particular, let $g(x) = (x - \mu)^2$ and $k = \lambda^2 \sigma^2$ so that Eq. (B.4) gives

$$\text{Prob}\{|x - \mu| \geq \lambda \sigma(x)\} = \text{Prob}\{(x - \mu)^2 \geq \lambda^2 \sigma^2(x)\} \leq \frac{\langle (x - \mu)^2 \rangle}{\lambda^2 \sigma^2(x)} = \frac{1}{\lambda^2} \quad (\text{B.5})$$

for all $\lambda > 0$. This result is known as the Bienaymé-Chebyshev inequality and leads directly to the weak law. The quantity $\bar{x}_N$ is itself a random variable with mean μ and a variance $\sigma_{\bar{x}}^2 = \sigma^2(x)/N$. Set $\lambda \sigma_{\bar{x}} = \epsilon$ and use Eq. (B.5) to obtain

$$\text{Prob}\{|\bar{x}_N - \mu| \geq \epsilon\} \leq \frac{\sigma^2(x)}{N\epsilon^2}. \quad (\text{B.6})$$

Because it is implicitly assumed that $\sigma^2(x)$ is bounded, the right side approaches zero as N approaches infinity, thus proving the weak law. This law can be written as

$$\boxed{\lim_{N \rightarrow \infty} \text{Prob}\{|\bar{x}_N - \mu| \geq \epsilon\} = 0} \quad (\text{B.7})$$

or, equivalently,

$$\boxed{\lim_{N \rightarrow \infty} \text{Prob}\{|\bar{x}_N - \mu| < \epsilon\} = 1.} \quad (\text{B.8})$$

In words, the weak law states that for any nonzero tolerance ϵ , no matter how small, with a sufficiently large sample there is a high probability that the sample average will be close to the expected value, i.e., within the tolerance margin. This type of convergence for $\bar{x}_N$ is technically called *convergence in probability* or *weak convergence* of a random variable.

Example B.1 Use of the Weak Law

What should the sample size N be to ensure with 99% probability that $\bar{x}_N$ is within 0.2σ of the population mean μ ?

From Eq. (B.6) one immediately obtains

$$\text{Prob}\{|\bar{x}_N - \mu| < \epsilon\} \geq 1 - \frac{\sigma^2}{N\epsilon^2(x)} \equiv 1 - \delta$$

for $\delta > \sigma^2/(N\epsilon^2)$ or $N > \sigma^2(x)/(\epsilon^2\delta)$. Here, one has $\epsilon = 0.2\sigma$ and $\delta = 0.01$, so that

$$N > \frac{\sigma^2}{\delta\epsilon^2} = \frac{\sigma^2}{(0.01)(0.2\sigma)^2} = 2500.$$

B.2 The Strong Law of Large Numbers

The weak law is a statement about the *limiting* behavior of sums of random variables. It is possible, however, to make a stronger statement and say something about the behavior of the sum $\sum_{i=1}^N x_i$ on the way to the limit. The derivation of the strong law is much more complex than that for the weak law and is simply stated here as follows [Kendall and Stuart 1977]. For any small $\epsilon > 0$ and $0 < \delta < 1$, a value of n exists such that, for any specified $m > 0$,

$$\text{Prob}\{|\bar{x}_N - \mu| \geq \epsilon\} \leq \delta, \quad (\text{B.9})$$

where $N = n, n + 1, \dots, n + m$. Equivalently, the strong law requires

$$\text{Prob}\{\lim_{N \rightarrow \infty} \bar{x}_N = \mu\} = 1. \quad (\text{B.10})$$

This theorem says more than does the weak law. In particular, it asserts that, for any specified $\epsilon > 0$ and $0 < \delta < 1$, one can identify a number of histories n such that Eq. (B.9) holds for a sequence of values of $N > n$ of arbitrary finite length m . Occasional deviations $|\bar{x}_N - \mu|$ may exceed ϵ but these are few. The strong law, thus, requires that the sample mean approach the population mean in a more demanding way. Technically, the sample average is said to *converge almost surely* or to *converge strongly*. A proof is given by Etemadi [1981].

B.2.1 Difference Between the Weak and Strong Laws

The weak form of the law of large numbers states that, for a specified large N , the sample mean $\bar{x}_N$ is likely to be near μ . But it leaves open the possibility that cases when $|\bar{x}_N - \mu| > \epsilon$, i.e., when the deviation of $\bar{x}_N$ from μ is outside some small tolerance ϵ , can occur an arbitrary (even infinite) number of times as N increases; however, such

occurrences happen at infrequent intervals. The strong law shows that such an infinite number of occurrences cannot happen, i.e., $|\bar{x}_N - \mu|$ will lie outside ϵ at most only a finite number of times.

In other words, the weak law states that $|\bar{x}_N - \mu|$ eventually becomes small but that not every value on the way is small. It may be that, for some N , the difference is relatively large. The strong law says that the probability of such a large event, however, is extremely small. Either form of the law is sufficient to form a basis for the Monte Carlo method because both indicate that $\lim_{N \rightarrow \infty} \bar{x}_N = \mu$.

B.2.2 Other Subtleties

It should be noted that the law of large numbers also holds even for an unbounded variance σ^2 [Cremér 1947]. However, for an infinite variance the proof of the law is considerably more complex than that given above for the weak law. A very large or infinite variance simply slows the convergence of $\bar{x}_N$ to μ .

Thus, both the strong and weak laws are valid for independent random variables that are identically distributed under the *sole* requirement that the mean exists. If the mean does not exist then the sample mean does not have a distribution that becomes narrower as the sample size N increases. An example of a distribution without a mean is the Cauchy distribution. The distribution of the sample mean is always the same for all N , namely, the Cauchy distribution itself.

Often, the law of large numbers, also known as the “law of averages,” is misused, as exemplified by the gambler’s fallacy. Gamblers often mistakenly believe that, in a fair game of chance, for example, a long series of losses must be followed by a series in which wins are more frequent in order to balance out the even win/loss ratio required by the law of averages for a large number of games. As another example, consider the tossing of a coin of which the probability of a head is p . The law of large numbers does *not* imply that the number of heads in N tosses necessarily deviates little from the expected number of heads Np ; rather, it states that the relative frequency of heads to flips is close to p . Nor does this law imply that, if the ratio of heads to flips is $> p$, after N flips, the probability of tails on subsequent flips becomes larger to compensate for the surplus of heads. Nature “averages out” by swamping (large N), rather than by fluctuating (keeping $\bar{x}_N$ close to μ within every subset of samples)!

Bibliography

- BERNOULLI, J. 1713, *Ars Conjectandi*, Impensis Thurnisiorum, Fratrū, Basilea. This posthumous Latin work was subsequently published in German as *Wahrscheinlichkeitsrechnung* in two volumes by Englemann, Leipzig, 1899.
- CREMÉR, H. 1947, *Mathematical Methods for Statistics*, Princeton University Press, Princeton, NJ.
- ETEMADI, N. 1981, “An Elementary Proof of the Strong Law of Large Numbers,” *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **55**, 119–122.
- KENDALL, M. AND A. STUART 1977, *The Advanced Theory of Statistics, Volume 1, Distribution Theory*, 4th ed., MacMillan Publishing Co., New York.

C Central Limit Theorem

In this appendix, the validity of the central limit theorem is demonstrated. Although the proof introduces some statistical concepts not previously discussed, the extreme importance that this theorem plays in Monte Carlo calculations merits examining the underpinnings of this amazing theorem. One, of course, may just accept the validity of the central limit theorem, but for those interested, the pertinent details are presented in this appendix.

C.1 Moment Generating Functions

The *moment generating function* (MGF) of the random variable x with a PDF $f(x)$ is defined by

$$M_x(t) \equiv \langle e^{tx} \rangle = \begin{cases} \sum_j e^{tx_j} f(x_j), & \text{for a discrete distribution} \\ \int e^{tx} f(x) dx, & \text{for a continuous distribution} \end{cases}. \quad (\text{C.1})$$

This function plays a central role in probability theory and its applications because it has many useful properties.

One of the more important properties of the MGF is that it readily yields all the moments of the PDF. From its definition

$$\begin{aligned} M_x(t) &= \langle e^{tx} \rangle = \left\langle 1 + tx + \frac{(tx)^2}{2!} + \frac{(tx)^3}{3!} + \dots \right\rangle \\ &= 1 + \langle tx \rangle + \langle (tx)^2/2! \rangle + \langle (tx)^3/3! \rangle + \dots \\ &= 1 + t \langle x \rangle + \frac{t^2}{2!} \langle x^2 \rangle + \frac{t^3}{3!} \langle x^3 \rangle + \dots \\ &= 1 + \mu_1 t + \mu_2 \frac{t^2}{2!} + \mu_3 \frac{t^3}{3!} + \dots, \end{aligned} \quad (\text{C.2})$$

where $\mu_n = \langle x^n \rangle$ is the n th moment of the PDF $f(x)$. Now differentiate this last expression n times and set $t = 0$ in the result to obtain

$$\left. \frac{d^n M_x(t)}{dt^n} \right|_{t=0} = \mu_n. \quad (\text{C.3})$$

Thus, if the characteristic function is known, it is a simple matter to compute all the moments of the PDF.

If all the moments are known, then $M_x(t)$ is also known because a Maclaurin series gives

$$\begin{aligned} M_x(t) &= M_x(0) + M'_x(0)t + M''_x(0)\frac{t^2}{2!} + M'''_x(0)\frac{t^3}{3!} + \cdots \\ &= 1 + \mu_1 t + \mu_2 \frac{t^2}{2!} + \mu_3 \frac{t^3}{3!} + \cdots \end{aligned} \quad (\text{C.4})$$

Thus, knowing all the moments of a PDF is equivalent to knowing the PDF itself.

C.1.1 Central Moments

In some applications, the n th *central* moment μ'_n is defined by $\langle (x - \mu)^n \rangle$, where $\mu = \mu_1$ is the first moment or mean of the distribution. These central moments can be related to the moments μ_n . For example, the first central moment is always zero ($\langle x - \mu \rangle = \langle x \rangle - \mu = 0$), and the second central moment is the variance given by

$$\begin{aligned} \text{var}(x) &\equiv \sigma^2 \equiv \mu'_2 \equiv \langle (x - \mu)^2 \rangle = \langle x^2 - 2\mu x + \mu^2 \rangle \\ &= \langle x^2 \rangle - 2\mu \langle x \rangle + \mu^2 = \langle x^2 \rangle - 2\mu^2 + \mu^2 = \mu_2 - \mu^2. \end{aligned} \quad (\text{C.5})$$

In general, the central moments can be related to the moments μ_n and vice versa. It can be shown [Kendall and Stuart 1977] that

$$\mu_n = \sum_{m=0}^n \binom{n}{m} \mu'_{n-m} \mu^m \quad (\text{C.6})$$

$$\mu'_n = \sum_{m=0}^n \binom{n}{m} \mu_{n-m} (-\mu)^m. \quad (\text{C.7})$$

C.1.2 Some Properties of the MGF

The MGF has many other useful properties. Here, two simple properties are derived. The MGF of cx , where c is a constant, is

$$M_{cx}(t) = \langle e^{cxt} \rangle = \langle e^{x(ct)} \rangle = M_x(ct). \quad (\text{C.8})$$

Similarly, the MGF for $x + c$ is

$$M_{x+c}(t) = \langle e^{(x+c)t} \rangle = e^{ct} \langle e^{xt} \rangle = e^{ct} M_x(t). \quad (\text{C.9})$$

Another property, needed for the proof of the central limit theorem, is the MGF for a sum of independent random variables. Let $S = \sum_{i=1}^N x_i$, where the x_i are independent random variables. Then

$$M_S(t) = \langle e^{tS} \rangle = \langle e^{t(x_1+x_2+\cdots+x_N)} \rangle = \left\langle \prod_{i=1}^N e^{tx_i} \right\rangle.$$

Because the x_i are independent,

$$M_S(t) = \prod_{i=1}^N \langle e^{tx_i} \rangle = \prod_{i=1}^N M_{x_i}(t). \quad (\text{C.10})$$

In words, the MGF of a sum of independent random variables is the product of their individual MGFs. From this result and Eq. (C.9), the MGF of $S = c_1x_1 + c_2x_2 + \cdots + c_Nx_N$, where the c_i are constants is

$$M_S(t) = \prod_{i=1}^N M_{x_i}(c_it). \quad (\text{C.11})$$

Example C.1 The MGF for a Normal Distribution

The PDF for a normal distribution with mean μ and variance σ^2 is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp[-(x - \mu)^2/\sigma^2].$$

The random variable $y = (x - \mu)/\sigma$ is distributed according to the unit normal

$$g(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2},$$

which has the MGF

$$\begin{aligned} M_y(t) &= \langle e^{yt} \rangle = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{yt} e^{-y^2/2} dy = e^{t^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(y-t)^2/2} d(y-t) \\ &= e^{t^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-z^2/2} dz = e^{t^2/2}. \end{aligned} \quad (\text{C.12})$$

Finally, use Eqs. (C.8) and (C.9) with $x = \sigma y + \mu$ to obtain

$$M_x(t) = e^{\mu t} M_{\sigma y}(t) = e^{\mu t} e^{\sigma^2 t^2/2} = \exp \left[\frac{\sigma^2 t^2}{2} + \mu t \right]. \quad (\text{C.13})$$

C.1.3 Uniqueness of the MGF

Does the MGF or the moments μ_n completely characterize the PDF $f(x)$? Could there be two PDFs that have the same MGF and the same moments? The answer is no. If the MGF for a random variable x_1 is the same as that for another random variable x_2 , then the PDFs for the two random variables are identical. Although, this is intuitively reasonable, a rigorous proof is rather complicated (see, for example, [Cramér \[1946\]](#)). Here, we quote two related theorems about the uniqueness of the MGF and a distribution's moments [[Cramér 1946](#)]:

- **Theorem 1:** If a MGF $M_x(t)$ exists in an interval that includes the origin $t = 0$, then there is only one distribution that has that MGF.
- **Theorem 2:** If the moments $\mu_1, \mu_2, \dots$ exist such that the right side of [Eq. \(C.4\)](#) converges in an interval containing the origin $t = 0$, there is only one distribution having those moments.

Thus, if two distributions have the same MGFs or the same moments, they must be the same. This is the key fact needed to prove the central limit theorem.

C.2 Central Limit Theorem

Pierre Simon de Laplace (1749–1827) is largely responsible for discovering and demonstrating the central role the normal distribution plays in the mathematical theory of probability. He also proved what is today called the central limit theorem [[Laplace 1810](#)]. This theorem was a generalization of the special case of [Moivre's \[1733\]](#) earlier result. A complete statement of the central limit theorem is as follows [[Riley et al. 2002](#)]:

Central limit theorem. *Let $x_i, i = 1, 2, \dots, N$ be independent random variables, each of which is described by a probability density function $f_i(x)$ (these may be all different) with a mean μ_i and variance σ_i^2 . The random variable $z = \sum_i x_i/N$, i.e., the average of the x_i , has the following properties:*

1. *its expected value is given by $\langle z \rangle = (\sum_i \mu_i)/N$;*
2. *its variance is given by $\text{var}(z) = (\sum_i \sigma_i^2)/N^2$;*
3. *and as $N \rightarrow \infty$ the probability density function of z tends to a normal distribution with the same mean and variance.*

It should be noted that, for this theorem to hold, the PDFs $f_i(x)$ must have well-defined means and variances. If, for example, the $f_i(x)$ were Cauchy distributions, which have no mean, then the theorem would not apply.

The first two properties are easily proved. The first is proved as follows:

$$\langle z \rangle = \frac{1}{N}(\langle x_1 \rangle + \langle x_2 \rangle + \dots + \langle x_N \rangle) = \frac{1}{N}(\mu_1 + \mu_2 + \dots + \mu_N) = \frac{\sum_i \mu_i}{N}. \quad (\text{C.14})$$

Note that this result does *not* require the x_i to be *independent*. Indeed, if $\mu_i = \mu$ for all i , then the above result reduces to $\langle z \rangle = N\mu/N = \mu$. The second property, however, does

require the x_i to be independent so that all covariances are zero. Then

$$\begin{aligned}\text{var}(z) &= \text{var}\left(\frac{1}{N}[x_1 + x_2 + \cdots + x_N]\right) \\ &= \frac{1}{N^2}[\text{var}(x_1) + \text{var}(x_2) + \cdots + \text{var}(x_N)] = \frac{\sum_i \sigma_i^2}{N^2}.\end{aligned}\quad (\text{C.15})$$

If x_i are all from the same distribution, then $\text{var}(z) = \sigma^2/N$.

Now examine property (3), a property that accounts for the ubiquity of the normal distribution throughout statistics. Begin by considering the MGF of z . From Eq. (C.11) this MGF is given by

$$M_z(t) = \prod_{i=1}^N M_{x_i}\left(\frac{t}{N}\right), \quad (\text{C.16})$$

where $M_{x_i}(t)$ is the MGF of $f_i(x)$. Now expand $M_{x_i}(t/N)$ as

$$M_{x_i}\left(\frac{t}{N}\right) = 1 + \frac{t}{N}\langle x_i \rangle + \frac{t^2}{2!N^2}\langle x_i^2 \rangle + \cdots = 1 + \mu_{1i}\frac{t}{N} + \mu_{2i}\frac{t^2}{2!N} + \cdots \quad (\text{C.17})$$

Here, $\mu_{1i}, \mu_{2i}, \dots$ are the moments of $f_i(x)$. With Eq. (C.5) this result can be written in terms of the central moments as

$$M_{x_i}\left(\frac{t}{N}\right) = 1 + \mu_i\frac{t}{N} + \frac{1}{2}(\sigma_i^2 + \mu_i^2)\frac{t^2}{N^2} + \cdots \quad (\text{C.18})$$

As N becomes large

$$M_{x_i}\left(\frac{t}{N}\right) \simeq \exp\left[\frac{\mu_i t}{N} + \frac{1}{2}\sigma_i^2\frac{t^2}{N^2}\right],$$

which can be verified by expanding the exponential up to terms of order $(t/N)^2$ to obtain Eq. (C.18). Then from Eq. (C.16) one obtains, for large N ,

$$M_z(t) \simeq \prod_{i=1}^N \exp\left(\frac{\mu_i t}{N} + \frac{1}{2}\sigma_i^2\frac{t^2}{N^2}\right) = \exp\left(\frac{\sum_i \mu_i}{N}t + \frac{1}{2}\frac{\sum_i \sigma_i^2}{N^2}t^2\right). \quad (\text{C.19})$$

Comparing this MGF to that of the normal distribution (see Eq. (C.13)), it is observed that they are the same. Thus, the PDF of the random variable z , as N becomes large, approaches a normal distribution with mean $\sum_i \mu_i/N$ and variance $\sum_i \sigma_i^2/N^2$.

In particular, if z is the sum of N independent measurements of the same random variable, i.e., from the same PDF $f(x)$, then, as $N \rightarrow \infty$, z has a normal distribution with mean μ and variance σ^2 .

Bibliography

- CRAMÉR, H. 1946, *Mathematical Methods of Statistics*, Princeton University Press, Princeton, NJ.
- KENDALL, M. AND A. STUART 1977, *The Advanced Theory of Statistics, Volume 1, Distribution Theory*, 4th ed., MacMillan Publishing Co., New York.
- LAPLACE, P.S. 1810, 1898 *Mémoire sur Les Approximations des Formules qui Sont Fonctions de Très Grand Nombres et sur Leur Application aux Probabilités*, Vol. 12, 301–345, Gauthier-Villars, Paris.
- MOIVRE, Abraham De 1733, *Approximatio ad Summam Terminorum Binomii $(a+b)^n$ in Seriem Expansi*. Republished: SMITH, D.E. 1959, “A Method of Approximating the Sum of Terms of the Binomial $(a+b)^n$ Expanded into a Series, from Whence Are Deduced Some Practical Rules to Estimate the Degree of Assent Which Is to be Given to Experiments,” *A Source Book in Mathematics*, Vol. 2, 566–575, Dover, New York.
- RILEY K.F., M.P. HOBSON, AND S.J. BENICE 2002, *Mathematical Methods for Physics and Engineering*, 2nd ed., Cambridge University Press, Cambridge, England.

D Some Popular Monte Carlo Codes for Particle Transport

There are many general-purpose Monte Carlo codes for radiation transport simulation. Some of these are well documented and widely used, others less well so, and still others proprietary. Nine popular nonproprietary codes are briefly described in this appendix. The code descriptions are based on material provided by representatives of the code collaborations. The principal authors of this material are identified within each code description. We thank these authors for providing information about their codes.

Not all the codes are similar in complexity nor are they similar in the range of problems that they address. Nevertheless, we have sought to limit each code description to about two pages in length. In each case, references to more complete documentation are provided in the bibliography at the end of this appendix. The codes are considered in alphabetical order.

D.1 COG

COG is a multiparticle Monte Carlo code for shielding and criticality use.

D.1.1 Current Version

As of June 2010, the currently distributed version is V10.0. V11.0 is being readied for release.

D.1.2 Principal Authors

The principal authors are Richard Buck, Lawrence Livermore National Laboratory (LLNL), Edward Lent (LLNL), and Thomas Wilcox, Jr., (formerly LLNL).

D.1.3 Operating Systems

Releases are available for SUN Solaris, Windows PC, and Intel/Linux.

D.1.4 Availability

Software and supporting library files may be obtained from the RSICC software distribution center at Oak Ridge National Laboratory or from the Nuclear Energy Agency

(CCC-0724). The COG Manual and libraries may be found at the COG web site at <http://cog.llnl.gov/>.

D.1.5 Applications

COG is a modern, full-featured Monte Carlo radiation transport code that provides accurate answers to complex shielding, criticality, and activation problems. Neutrons, photons, electrons, and protons may be transported. COG was written to be state-of-the-art and is free of physics approximations and compromises found in earlier codes. COG is fully three-dimensional, uses pointwise cross sections and exact angular scattering, and allows for a full range of variance reduction options to speed up solutions for deep penetration problems. Additionally, a criticality option is available for computing k_{eff} for assemblies of fissile materials.

Neutron cross-section libraries available with V10.0 include the LLNL ENDL-90 cross-section set and the Brookhaven ENDFB/VI Release 7 set. Neutron libraries used with V11.0 include ENDL-90, ENDFB/VI Release 7, ENDFB/VII Release 0, JEFF2.2, JENDL3.3, MCNP.66c, and RED2002. Several choices of $S(\alpha, \beta)$ thermal scattering libraries are also available, including ENDFB/VI Release 2 and ENDFB/VII Release 0.

Analytic surfaces are used to describe geometric boundaries. *Parts* (volumes) are described by a method of constructive solid geometry. Surface types include surfaces of up to fourth order, and psuedosurfaces such as boxes, finite cylinders, and figures of revolution. Repeated assemblies need be defined only once. Parts are visualized in cross-section and perspective picture views.

Source and random-walk biasing techniques may be selected to improve solution statistics. These include source angular biasing, importance weighting, particle splitting and Russian roulette, path length stretching, point detectors, scattered direction biasing, and forced collisions.

For a fissioning system, COG can compute the effective multiplication factor k_{eff} by transporting batches of neutrons through the system. COG can also compute gamma-ray doses produced by neutron-activated materials, starting with just a neutron source.

D.1.6 Significant Features

Special geometry features include: (1) easy description of complex surfaces by surface-of-revolution, (2) use of the *Fill* feature obviates the need to define all problem space, and (3) the *Unit* feature allows a part to be defined once and used many times in a problem.

Other features include: (1) a wide choice of neutron libraries (V11) and the ability to use multiple libraries, (2) a probability table treatment of the unresolved resonance region, (3) user-defined sources and detectors, which allow the user to write specialized source and/or detector routines and have them linked in the COG executable at runtime, (4) solution of coupled problems involving neutrons, photons, and electrons, (5) running on multinode parallel processors under the MPI system, and (6) the inclusion of photonuclear reactions.

D.2 EGSnrc

D.2.1 Acronym and Current Version

Electron Gamma Shower, **NRC** version [Kawrakow 2000a; Kawrakow and Rogers 2003]. The current version of EGSnrc is V4-r2.3.1. Official releases are typically published once a year. There is a separate distribution area for development snapshots that are produced every 2–4 weeks.

D.2.2 Principal Authors

The principal authors are I. Kawrakow (National Research Council of Canada) and D.W.O. Rogers (Carleton University), although many others have contributed to the system in a significant way, including W.R. Nelson and H. Hirayama (together with D.W.O. Rogers are the original authors of the EGSnrc predecessor code EGS4), A.F. Bielajew, C.M. Ma, J.P. Seuntjens, B.R.B. Walters, E. Mainegra-Hing, F. Tessier, to name a few.

D.2.3 Operating Systems

The core EGSnrc system works on all major modern operating systems (OSs). There are certain portions that are available only for a limited set of OSs (e.g., the graphical installer is available only for Linux and Windows NT/2000/XP/Vista with an installation script provided for other Unix systems, and the EGS_Windows package for visualizing linear accelerators and particle tracks runs only on Linux).

D.2.4 Availability

The EGSnrc system is available for download at <http://www.irs.inms.nrc.ca/EGSnrc/EGSnrc.html> and snapshots from the development version of the code can be found at <http://www.irs.inms.nrc.ca/EGSnrc/snapshot/>. The codes BEAMnrc [Rogers et al. 1995; Rogers et al. 2005] and DOSXYZnrc [Walters et al. 2005] are distributed separately (follow links on the main EGSnrc distribution site) and must be installed after completing the EGSnrc installation. The EGSnrc license permits free use for research and educational purposes, and a separate license from the National Research Council of Canada is required for commercial use. Portions of the system are licensed under version 2 of the GPL.

D.2.5 Applications

The EGSnrc package is not a monolithic application. Instead, it provides a set of functions and subroutines for the simulation of coupled electron and photon transport. To obtain a complete application the user must supply a so-called “user code” that provides a description of the simulation geometry, the particle source, and a function for extracting information from the simulation. User codes can be written in Mortran, Fortran, and, since the release of the EGSnrc C++ class library [Kawrakow 2005] in

2005, in C or C++. Included in the distribution are applications for computing dose distributions, ionization chamber correction factors, particle fluences, pulse height distributions [Rogers et al. 2003], and various simple tutorial codes [Kawrakow and Rogers 2003]. Two major applications based on EGSnrc are BEAMnrc [Rogers et al. 1995, 2005] and DOSXYZnrc [Walters et al. 2005]. BEAMnrc has been and is still being used extensively to model medical linear accelerators (408 citations to Rogers et al. [1995] in the scientific literature). DOSXYZnrc specializes in computing dose distributions in rectilinear geometries, including geometries defined via CT images and has been incorporated into a number of clinical implementations of Monte Carlo-based radiation treatment planning and/or treatment planning verification. The group at NRC is currently developing user codes for application in cone beam CT and in positron emission tomography as well as improving and refining the user codes for radiation dosimetry and radiation therapy. The group at Carleton University has developed a code for fast calculation of dose distributions in brachytherapy and the capability of modeling off-focal radiation and voltage ripple when simulating X-ray units with BEAMnrc.

The main application area of EGSnrc has been medical physics where it has been employed to study problems related to radiation dosimetry, radiation therapy, X-ray-based medical imaging, design and characterization of medical devices, radiation protection, and so forth. EGSnrc has also been applied in radiation detection and shielding, high-energy physics, atmospheric sciences, radiation chemistry, and many other areas. As of October 2007, the main EGSnrc paper [Kawrakow 2000a] has been cited 201 times and the EGSnrc manual [Kawrakow and Rogers 2003] 243 times in the scientific literature.

An incomplete and not regularly maintained list of EGSnrc- and EGS4-related publications can be found at http://www.irs.inms.nrc.ca/papers/egs.biblio/egs_papers.html. A separate list of BEAMnrc-/DOSXYZnrc-related papers is maintained at http://www.irs.inms.nrc.ca/BEAM/bibliog/omega_pubs.html.

D.2.6 Significant Features

EGSnrc is a package for the simulation of coupled electron and photon transport in the energy range between 1 keV and several tens of GeV. The most outstanding feature of EGSnrc is its accurate implementation of the condensed history technique for charged particle transport that has been demonstrated to be step-size independent and artifact-free at the 0.1% level [Kawrakow 2000a,b]. Such accuracy is a result of an exact multiple scattering theory [Kawrakow and Bielajew 1998a; Kawrakow 2000a], an improved electron-step algorithm [Kawrakow and Bielajew 1998b; Kawrakow 2000a], an exact boundary-crossing algorithm via single scattering simulation in the vicinity of region interfaces [Kawrakow 2000a], and various other details [Kawrakow 2000a]. Compared to its predecessor EGS4, EGSnrc includes numerous improvements in the modeling of photon and electron interactions: (1) binding effects and Doppler broadening in the relativistic impulse approximation for Compton scattering, (2) full relaxation cascade after innershell ionization taking into account radiative and non-radiative transitions, (3) an ability to utilize user-supplied molecular form factors for

Rayleigh scattering, (4) an ability to use total photon cross-section tabulations from XCOM, EPDL97, Storm and Israel, or any other user-supplied tabulation, (5) use of radiative corrections for Compton scattering, (6) use of exact differential pair production cross sections in the energy range up to 85 MeV, (7) an explicit simulation of pair production in photon-electron interactions (*triplet production*), (8) explicit simulation of innershell ionizations by electron impact, (9) improved bremsstrahlung differential cross-section tabulations from NIST that are the basis for ICRU-recommended radiative stopping powers, and (10) inclusion of spin and relativistic effects in electron and positron elastic scattering. Another important feature of EGSnrc is that it tends to be significantly faster than most other general-purpose Monte Carlo packages when charged particle transport is involved.

An important advance for the usability of the code was the release of the EGSnrc C++ class library [Kawrakow 2005], which provides a powerful general-purpose geometry package that can be used to model complex geometries, several particle source classes, and many utility classes for random number generation, for efficient sampling of probability distribution functions, for interpolation and integration, for tallying of quantities of interest, and so forth. The C++ class library has permitted the development of several new C++ user codes for EGSnrc in a short period of time that have been included in the general distribution or will be released in the near future.

D.3 GEANT4

D.3.1 Current Version

Geant4 Release 9.3, patch 01 (abbreviation: 9.3.p01) as of June 2010.

D.3.2 Principal Author

The principal author is the Geant4 Collaboration. This summary was kindly provided by J. Apostolakis and V.N. Ivanchenko.

D.3.3 Operating Systems

The Geant4 tool kit is verified and supported on specific configurations, on Linux variants, on Windows, and on Mac OS X. Testing was undertaken and precompiled libraries are provided for (1) gcc 3.4.6, tested on Scientific Linux 4 (based on Redhat Linux Enterprise 4); (2) gcc 4.0, tested on Mac OS X, 10.4; (3) Sun CC 5.5, tested on SUN Solaris (SunOS 5.8); and (4) Visual C++ 8.0, tested on Windows/XP.

D.3.4 Availability

The Geant4 source code is available openly for download on the Web, subject to the open-source conditions of the Geant4 license (which can be viewed at <http://geant4.web.cern.ch/geant4/license/index.shtml>). All data files required are part of the distribution. Optional external packages can provide more flexible visualization, storage of setup geometry, and analysis of the simulation results.

Geant4 is distributed as source code, enabling users to compile it for different compiler or OS versions. Configuration files are provided to enable a user to build Geant4 libraries on other systems (including Unix variants), other compilers, or compiler versions. Only one external package, CLHEP (<http://cern.ch/clhep/>), is necessary for the installation and running of a Geant4 application.

D.3.5 Applications

Geant4 is a tool kit for Monte Carlo simulation of the interaction and transport of particles in matter [Agostinelli et al. 2003; Allison et al. 2006]. Geant4 was created primarily for use in high-energy physics (HEP) and medical physics and for studying the effects of radiation on spacecraft. It has seen widespread use for simulations of HEP events in accelerator experiments. Since 2002, a Geant4-based program has been used in simulations for the BaBar experiment at SLAC and has produced over 10^{10} events [Banerjee 2008a]. Geant4 entered production in early 2004 for the ATLAS [Ueda 2010], CMS [Banerjee 2008b], and LHCb experiments at the Large Hadron Collider at CERN. Geant4 is also used for many other HEP experiments for detector simulations and beam transport calculations.

Geant4 applications also span space research, radiation assessment, and studies for medical physics, including imaging and therapy. In addition to a large variety of individual applications, a number of specialized software packages have been created, based on Geant4. For example, the MULASSES software (<http://reat.space.qinetiq.com/mulassis/>) is used to analyze radiation shielding for space missions, and the GRAS software package (http://space-env.esa.int/R_and_D/gras/) is used to simulate radiation transport in complex geometries such as a detailed model of a space module. The OpenGATE collaboration has used Geant4 to create the openly distributed GATE software package (<http://opengatecollaboration.healthgrid.org/>), which is widely used for tomography applications, such as positron emission tomography (PET) and single photon emission computed tomography (SPECT), and has a wide user base. A large number of other applications utilize Geant4 for the simulation of different medical applications, including brachytherapy, proton therapy [Aso et al. 2005], heavy ion therapy [Kase et al. 2006], and time-varying geometries for radiation treatment planning.

D.3.6 Significant Features

Geant4 provides comprehensive geometry and physics modeling capabilities embedded in a flexible structure base using object-oriented technology and C++. Because of the abstract interfaces, users can choose existing components, tailor and adapt them, or create their own plug-in components.

Models of the interaction of particles with matter are implemented in the Geant4 tool kit to describe electron, ion, muon, gamma ray, electromagnetic (EM), hadronic, and optical photon processes. For many physics processes [e.g., Apostolakis et al. 2008; Wright et al. 2006, 2007] a choice of physics models is offered, with different trade-offs between accuracy and CPU cost. Tools for optimization of configurations of geometry and physics are also provided. The combinations of physical processes

for typical applications are recommended in physics lists. Users can thus choose an existing physics list/configuration to meet their needs, revise or customize an existing one, or create a new one.

The Geant4 geometry module enables users to describe their setup, and then locates and moves tracks inside it. It contains a variety of shapes and allows Boolean operations of these shapes [Anninos et al. 2006] as well as the ability to create a hierarchy of volumes. In complex setups, volumes can be repeated, sharing key characteristics. Volumes can also be grouped into regions. A particular geometry can be described in C++ code, or read in from files in the GDML markup language (<http://cern.ch/GDML/>). Other tools can convert the geometry from a computer-aided design (CAD) system into a GDML file readable by Geant4. The tool kit also enables users to simulate motion of volumes in the geometrical setup.

Production thresholds (cuts) used for some EM processes can be varied by geometrical region to allow different precisions for different parts of the model. Other event-biasing techniques offered in Geant4 include geometrical importance biasing (splitting and Russian roulette), cross-section enhancement, options for leading particle biasing, and weight window methods. Parallel geometry descriptions, typically simpler, can be used simultaneously for scoring, importance biasing, and/or shower parameterization.

Electromagnetic fields described by Geant4 affect the transport of charged particles. A global field can be used and can be overridden locally for any volume. Typical scoring, including radiation dose deposition, is a part of the tool kit. Specific “hooks,” such as the optional user action classes, provide the user with complete control of the production of hits or scoring of the results. A powerful stacking mechanism enables users, at almost no CPU cost, to control the prioritization of the processing of tracks in an event, to abort particular tracks or events early in the simulation, and to postpone the propagation of certain parts of tracks.

Visualization options range from the fast OpenGL to high-quality rendered pictures. A comprehensive system is available for the visualization of geometries, tracks, and detector hits through a choice of drivers and a variety of user interface options. Time-dependent visualization of particles propagation is also possible.

The Geant4 tool kit can be used for stand-alone simulations or can be integrated in larger software frameworks, such as discussed above for large HEP experiments. Users can also customize and further extend the Geant4 software based on requirements for a particular application. For example, changing the run configuration allows for users to vary the geometry or materials composition during execution.

D.4 MCNP and MCNPX

MCNP5 and MCNPX are general-purpose Monte Carlo radiation transport codes where MCNP is an acronym for **M**onte **C**arlo **N**-**P**article and the “X” in MCNPX is for **e**Xtended. The two codes differ in the particles that they are capable of transporting and in their primary user bases. MCNP5 can be used for neutron, photon, electron, or coupled neutron/photon/electron transport, including the capability to calculate eigenvalues for critical systems. The primary user base for MCNP5 are those working more in the nuclear engineering fields. MCNPX by contrast can transport all

the particles that MCNP5 is capable of as well as protons, alpha particles, and many more, and is capable of linked transport between many of these particles. MCNPX's user base is primarily in high-energy physics though it too is capable of the same calculations as MCNP5. Both codes are discussed in the same section because they were at one time the same code that branched to pursue different emphases, and "MCNP" is used to refer to both MCNP5 and MCNPX.

D.4.1 Current Version and Availability

As of June 2010, the current release of MCNP5 is v1.51 and the current release of MCNPX is v2.6.0. The MCNP code is developed by Los Alamos National Laboratory's Applied Physics (X) Division's Computational Analysis and Simulations Group, X-3.

As of June 2010 both codes are currently available as a single package. MCNP5 and MCNPX are available in the United States through the Radiation Safety Information Computational Center (RSICC) in Oak Ridge, TN (see <http://www-rsicc.ornl.gov/>). Both codes are also distributed in Europe to members of the Nuclear Energy Agency (NEA) Data Bank and in Japan through the Research Organization for Information Science and Technology (RIST).

D.4.2 Principal Authors

As of June 2010, the principal authors of MCNP5 are Thomas E. Booth, Forrest B. Brown, Jeffrey S. Bull, J. Tim Goorley, H. Grady Hughes, Roger L. Martz, Avneet Sood, Jeremy E. Sweezy, and Anthony J. Zukaitis.

As of June 2010, the principal authors of MCNPX are John S. Hendricks, Gregg W. McKinney, Michael L. Fensin, Franz X. Gallmeie, Michael R. James, Russell C. Johns, Joe W. Durkee, Joshua P. Finch, Denise B. Pelowitz, Laurie S. Waters, and M. William Johnson.

D.4.3 Operating Systems

MCNP is supported on most of the available operating systems for both 32-bit integers and 64-bit integers including Microsoft Windows, Mac OS X, Linux, and other UNIX-like operating systems. Additionally, MCNP can be compiled for multiprocessing through either the message passing interface (MPI) or the application programming interface (API). MCNP5 also has the additional capability of parallelization through the open multiprocessing (OpenMP) API and can use both MPI and OpenMP together. MCNP's graphics routines are supported through the X Windows system.

D.4.4 Applications

Because of MCNP's general-purpose nature, it can be used in most fields of study where knowledge of how radiation interacts in matter is needed. Common applications include, but are not limited to, medical physics, space reactors, nuclear fuel

cycles, threat reduction, nuclear criticality (k_{eff}) calculations, radiography, oil well logging, irradiation facilities, isotope production, detector development, and environmental analysis. MCNPX has additional applications in the realm of high-energy physics and particle accelerator applications.

MCNP tracks neutrons from 0 to 20 MeV for all isotopes and up to 150 MeV for some isotopes using pointwise (continuous) cross-section data. Photons are tracked from 1 keV to 100 GeV using similar pointwise cross-section data in MCNP and electrons from 1 keV to 1 GeV. MCNPX extends the energy range of the above particles using physical cross-section models. MCNPX also extends transport capability to a myriad of other particles by means of physics models. Currently, the available models are the Bertini, ISABEL, CEM03, INCL4, FLUKA, and LAQGSM models.

D.4.5 Significant Features

The code treats an arbitrary three-dimensional configuration of materials in geometric cells bounded by first- and second-degree surfaces and fourth-degree elliptical tori. Pointwise cross-section data are used. For neutrons, all reactions given in a particular cross-section evaluation, most recently ENDF/B-VII, are accounted for in the transport and the user may select which cross-section evaluation to use. Thermal neutrons are described by both the free gas and $S(\alpha, \beta)$ models. For photons, the code accounts for incoherent and coherent scattering, the possibility of fluorescent emission after photoelectric absorption, and absorption in electron-positron pair production. Electron/positron transport processes account for angular deflection through multiple Coulomb scattering, collisional energy loss with optional straggling, and the production of secondary particles including K-shell X-rays, knock-on and Auger electrons, bremsstrahlung, and annihilation gamma rays from positron annihilation at rest. Electron transport does not include the effects of external or self-induced electromagnetic fields. Photonuclear physics is available for a limited number of isotopes. Again, MCNPX extends the capabilities discussed above to include physics models for transport of many charged particles as well as arbitrarily sized, heavy-charged particles.

The source definition capabilities of MCNP allow users to specify almost any source. Source variables may be specified discretely or as distributions. Additionally, source variables may be defined to be functions of other source variables providing the user great versatility in defining sources. In addition to fixed-source definitions, the user can specify criticality sources and MCNP can write surface sources of particles crossing surfaces for use in subsequent calculations.

MCNP offers a significant quantity of built-in Monte Carlo variance reduction techniques. Importance splitting/rouletting can be performed if the user specifies splitting parameters for the geometry. Additionally, space-energy-or space-time-dependent splitting can be performed using the weight-window technique, a weight-dependent splitting/rouletting variance reduction technique. MCNP's weight-window generator assists the user in generating the weight-window parameters. Semi-deterministic variance reduction techniques are also available through the DXTRAN technique and point detectors (next-event estimators). Time, energy, and weight cutoffs are also included in MCNP as well as many others.

The tallying capabilities of MCNP are extensive. Some of the basic tallies include surface flux and current tallies, volume flux tallies, point and ring detectors, heating tallies, and pulse-height tallies to simulate radiation-detector responses. The tallies can be performed over multiple individual cells or surfaces as well as the combinations of cells or surfaces and can be binned by energy, time, direction, special multipliers such as cross sections, surface and cell flagging, and user-specified bins. MCNP also has the capability to perform mesh tallies over a user-specified mesh independent of the geometry and which can also be binned by energy.

MCNP offers a significant quantity of platform-independent plotting capability. Geometries can be plotted and ensured accurate through the MCNP geometry plotter. Tally results and cross sections can be viewed through MCNP's tally and cross-section plotting, where both individual tallies and cross sections as well as multiple tallies and cross sections can be viewed and written to postscript files. The results of mesh tallies can also be visualized with the tally plotter.

Other features include, but are not limited to, the ability to model randomly changing geometry, to include diagnostics for determining convergence of criticality calculations (source entropy), proper fission neutron multiplicity, and modeling electron straggling energy loss. Another major feature allows the users to create nuclide tables at new temperatures. The resolved resonance data are Doppler broadened to a higher temperature. The unresolved resonance probability tables and $S(\alpha, \beta)$ thermal treatments can also be interpolated to different temperatures.

D.4.6 Additional Information

Additional information including verification and validation studies for MCNP5 and MCNPX can be found at <http://mcnp-green.lanl.gov/> and <http://mcnp.x.lanl.gov/>, respectively.

D.5 MCSHAPE

MCSHAPE is a general-purpose Monte Carlo code developed to simulate the transport of unpolarized or polarized photons in one- or three-dimensional geometries [Fernandez and Molinari 1993; Fernandez et al. 2004].

D.5.1 Acronyms

Monte Carlo **SHAPE** where SHAPE is an older deterministic code that builds up the albedo X-ray spectra by overlapping multiple scattering contributions (two or three collisions are allowed).

D.5.2 Current Version and Availability

The project MCSHAPE is a work in progress. The last version ready to download is MCSHAPE1D 2.61 (June 2007). The executable code is freely distributed on the web site at <http://shape.ing.unibo.it>.

D.5.3 Principal Authors

The principal authors are Jorge E. Fernandez (project leader) and Viviana Scot. The code is a registered property of Alma Mater Studiorum University of Bologna.

D.5.4 Operating Systems

MCSHAPE runs on 32- and 64-bit Pentium processors. It has been tested under both Windows (several versions) and Linux. A minimum screen resolution of 1024×768 is preferable. The folder with the physical data for the simulation requires about 800 MB, so at least 900 MB of free memory to install the program is recommended. MCSHAPE v.2.61 has been compiled under both Windows XP and Linux using INTEL Fortran version 10.0 and Winteracter library v7.0e. A parallel version is available for Linux clusters.

D.5.5 Applications

MCSHAPE simulates the diffusion of X-rays into matter taking into account the most important photon-atom interactions in the X-ray energy range and follows the evolution of the polarization state of the radiation. MCSHAPE is particularly appropriate for describing the multiple scattering terms to yield the total X-ray spectrum. All aspects that characterize the computed multiple-scattering contributions can be suitably defined: (1) the intensity term, (2) the full polarization state as a function of energy, (3) the number of collisions, and (4) the physical processes.

MCSHAPE3D, the three-dimensional extension of the code, can simulate the propagation of photons from polarized or unpolarized sources in complex heterogeneous media and, therefore, is a valuable tool for the simulation of scanning X-ray fluorescence (XRF) experiments, XRF tomography, and for the interpretation of the experimental results involving photon-matter interactions in three dimensions (3D). The geometry is modeled using a 3D regular grid of cubic voxels. For each voxel, the composition is specified by giving the number of chemical elements, their weight fractions, the atomic characteristics of each element, the total mass attenuation coefficient, and the total density. To complete the model, several types of sources and collimators can be chosen from a library.

MCSHAPE can determine the influence of the detector on the measured spectrum using two stages of analysis [Fernandez and Scot 2009]. In the first one, the transport of photons into the detector is described by a transport model that depends only on the physical properties of the detector material such as composition, density, thickness, etc. In the second stage, the broadening and the energy resolution of the detector, which depends specifically on the detection mechanism, is described by using the independent code RESOLUTION. The influence of the electronics is partially considered as energy discretization, while pulse processing problems are neglected.

The code is self-consistent and is distributed with a library called ELEMENTS (about 800 MB) containing all the physical parameters of the elements with atomic number Z from 1 to 92.

D.5.6 Significant Features

To correctly represent the polarization of photons, MCSHAPE uses the four intensity parameters introduced by Stokes [1852], which contain all the physical information about the polarization state of the photons—i.e., the intensity, the degree of polarization, the orientation and the ellipticity of the ellipse of polarization—at each point in the geometry and in any given direction.

The Boltzmann-Chandrasekhar vector transport equation describes formally the vector flux $\mathbf{f}(\mathbf{r}, \boldsymbol{\Omega}, \lambda)$ of polarized photons in the Stokes system (having components f_I, f_Q, f_U , and f_V) and can be written as

$$\begin{aligned} & \boldsymbol{\Omega} \cdot \nabla \mathbf{f}^{(S)}(\mathbf{r}, \boldsymbol{\Omega}, \lambda) + \mu(\mathbf{r}, \lambda) \mathbf{f}^{(S)}(\mathbf{r}, \boldsymbol{\Omega}, \lambda) \\ &= \int_0^\infty d\lambda' \int_{4\pi} d\boldsymbol{\Omega}' H^{(S)}(\mathbf{r}, \boldsymbol{\Omega}, \lambda, \boldsymbol{\Omega}', \lambda') \mathbf{f}^{(S)}(\mathbf{r}, \boldsymbol{\Omega}', \lambda') + \mathbf{S}^{(S)}(\mathbf{r}, \boldsymbol{\Omega}, \lambda), \end{aligned}$$

where

$$H^{(S)}(\mathbf{r}, \boldsymbol{\Omega}, \lambda, \boldsymbol{\Omega}', \lambda') = L^{(S)}(\pi - \psi) K^{(S)}(\mathbf{r}, \boldsymbol{\Omega}, \lambda, \boldsymbol{\Omega}', \lambda') L^{(S)}(-\psi')$$

is the kernel matrix in the meridian plane of reference; $K^{(S)}(\mathbf{r}, \boldsymbol{\Omega}, \lambda, \boldsymbol{\Omega}', \lambda')$ is the scattering matrix in the scattering plane of reference; $L^{(S)}$ is the four-by-four rotation matrix, which transforms the scattered flux from the scattering plane to the meridian plane of the reference; $\mu(\mathbf{r}, \lambda)$ is the narrow-beam attenuation coefficient, which is independent of the state of polarization of the photons (assuming the matter is isotropic); and $\mathbf{S}^{(S)}(\mathbf{r}, \boldsymbol{\Omega}, \lambda)$ is the source vector flux with components.

The total vector intensity is obtained (component by component) by adding the contributions $f_i^{(k)}(\mathbf{r}, \boldsymbol{\Omega}, \lambda)$ from the different number of collisions, as is done for the scalar model, namely,

$$f_i(\mathbf{r}, \boldsymbol{\Omega}, \lambda) = \sum_{k=0}^{\infty} f_i^{(k)}(\mathbf{r}, \boldsymbol{\Omega}, \lambda).$$

Rayleigh scattering, Compton scattering, and the photoelectric effect, and the prevailing interactions in the X-ray regime (1-1000 keV) are included in the simulation using state-of-the-art models. Because MCSHAPE uses the Stokes representation, it is a unique Monte Carlo code that can follow the complete evolution of the polarization state of the radiation. The current version of MCSHAPE uses a new algorithm for the simulation of Compton scattering that includes Doppler broadening effects [Fernandez and Scot 2007b]. This new algorithm corrects the error in the representation of the Compton peak in the standard approach adopted for generating the Compton profile in other MC codes.

Some of the features of MCSHAPE include: (1) atomic Rayleigh scattering, (2) atomic Compton scattering with a correct Compton profile, (3) the photoelectric

effect with about 1000 edge energies and line widths, (4) infinite or finite target thicknesses, (5) heterogeneous three-dimensional targets, (6) monochromatic or polychromatic sources with arbitrary polarization states, (7) use of reflection or transmission geometries, (8) external detectors with detector responses, and (9) an open database. In future releases of MCSHAPE, electron bremsstrahlung and user-defined geometrical elements are planned.

D.6 PENELOPE

D.6.1 Acronyms and Current Version

PENELOPE is an acronym for **P**enetration and **E**nergy **L**oss of **P**ositrons and **E**lectrons. The current version is PENELOPE2008.1.

D.6.2 Principal Authors

The principal authors are Francesc Salvat, José M. Fernández-Varea, and Josep Sempau, all from the Universitat de Barcelona, Barcelona, Spain.

D.6.3 Operating Systems

The simulation routines are written in FORTRAN and can be run on any computer with a FORTRAN 77 (or higher) compiler. The geometry viewers run under MS Windows and require a Windows emulator to run under other operating systems.

D.6.4 Availability

PENELOPE is free open-source software. The code system is distributed by the OECD Nuclear Energy Agency Data Bank (<http://www.nea.fr/html/dbprog/>) and from the Radiation Safety Information Computational Center.

D.6.5 Applications

PENELOPE performs Monte Carlo simulation of coupled electron-photon transport in material structures consisting of homogeneous bodies limited by quadratic surfaces. A detailed description of the physical models, tracking algorithms, and code structure is given in [Salvat et al. \[2006\]](#), available in PDF format from the OECD Nuclear Energy Agency web site at <http://www.nea.fr/lists/penelope.html>. The code system is entirely self-contained. It consists of the FORTRAN source files, the interaction database, and a set of auxiliary tools (geometry and shower viewers, examples of main programs, etc.). The interaction database covers the energy interval between 50 eV and 1 GeV. Interaction data for energies higher than 1 GeV are estimated by natural

cubic spline extrapolation. Although PENELOPE can transport particles with energies as low as 50 eV, the adopted interaction models (and also the particle trajectory model underlying the Monte Carlo simulation) are reliable only for energies higher than about 1 keV.

PENELOPE simulates the main interactions of photons (Rayleigh scattering, Compton scattering, photoelectric absorption, and electron-positron pair production) and of electrons/positrons (elastic and inelastic collisions, bremsstrahlung emission, and, in the case of positrons, two-photon annihilation). The ionization of K, L, and M electronic shells of atoms caused by Compton scattering, photoelectric absorption, and electron/positron impact and the subsequent emission of X-rays and Auger electrons are also simulated. The energies of the X-rays emitted from primary vacancies are set equal to their experimental values. The information in the database refers mostly to interactions of radiation with single, isolated atoms. Interactions with compound materials are described using Braggs additivity rule, i.e., cross sections for molecules are approximated by the sum of cross sections of the atoms in the molecule. The adopted cross sections are the most accurate available to date, limited only by the required generality of the code.

Random photon histories are generated using the conventional detailed simulation scheme. Electrons and positrons are simulated by means of a mixed (class II) scheme in which hard interactions (i.e., interactions involving angular deflections and energy losses larger than preselected cut-off values) are simulated individually from the corresponding differential cross sections. The collective effect of the soft interactions (with energy loss or angular deflection less than the cut-off values) that occur between each pair of consecutive hard interactions is described using multiple scattering approximations. The effect of soft scattering on the spatial displacements of the transported electron or positron is simulated using the random-hinge method. For each considered material, the actual values of the energy-loss and angular cut-offs are determined by a set of simulation parameters that are set by the user. These parameters also determine the amount of detail (energy and angular resolutions) and the speed of the simulation.

PENELOPE is structured as a family of subroutine packages. The Fortran source files of the basic packages are `penelope.f`—analog simulation of coupled electron-photon transport in unbounded media; `pengeom.f`—tracking of particles through modular quadratic geometries (can operate material structures with up to 10,000 surfaces and 5000 bodies); and `penvared.f`—variance reduction routines (particle splitting, Russian roulette, and interaction forcing). The `penelope.f` and `pengeom.f` subroutines generate particle histories automatically, i.e., they track particles through the material structure, simulate particle interactions and the generation of secondary particles, and manage the secondary stack. To take full advantage of the code capabilities, the user should write a steering main program, which only has to define the initial states of primary particles according to the characteristics of the radiation source, control the evolution of the generated showers, and keep score of the relevant quantities. To spare occasional users from having to write their main program, the distribution

package includes a generic main program, named `penmain`, that simulates electron-photon transport in quadratic geometries, with different sources and a variety of scoring options, which are defined through an input file.

PENELOPE has been applied to a wide variety of practical problems, including electron backscattering, electron probe microanalysis and X-ray generators, response of radiation detectors, dosimetry and characterization of ionization chambers, radiotherapy and simulation of medical electron accelerators and cobalt units, and synchrotron radiation.

D.6.6 Significant Features

The most characteristic feature of PENELOPE is the consistent use of class II schemes for electrons and positrons. When the energy-loss and angular cut-off values are set to zero, the simulation becomes purely detailed, i.e., free from any possible artifact introduced by multiple scattering approximations. Because only the effect of soft interactions needs to be accounted for by the random-hinge method, spatial displacements are accurately described. In many practical cases, the simulation results are very stable under variations of the simulation parameters. The geometry package PENGEO is tailored to minimize the numerical work needed to track particles using either detailed or mixed simulation. The geometry viewers generate images of the geometry by using strictly the same tracking routines as in the simulation. PENELOPE is particularly well adapted to simulations of electron transport at low and intermediate energies. The cross sections for elastic scattering were obtained from elaborate first principles calculations (Dirac partial-wave analysis). Inelastic collisions are described by means of the Sternheimer-Liljequist generalized oscillator strength model (the target atom or molecule is represented as a set of delta oscillators), which yields a realistic description of inelastic collisions in the energy range of interest. In particular, it gives collision-stopping powers consistent with ICRU-recommended values. The ionization of inner shells by electron and positron impact is described as an independent mechanism.

D.7 SCALE

D.7.1 Acronyms

ENDF	Evaluated Nuclear Data File
GeeWiz	Graphically Enhanced Editing WIZard
MAVRIC	MONACO with Automated Variance Reduction using Importance Calculations
MORSE	Multigroup Oak Ridge Stochastic Experiment
SCALE	Standardized Computer Analyses for Licensing Evaluation
SGGP	SCALE Generalized Geometry Package
TORT	Three-dimensional Oak Ridge Discrete Ordinates Neutron/Photon Transport Code

D.7.2 Current Version

SCALE 6.0

D.7.3 Principal Author

The principal author is the Oak Ridge National Laboratory.

D.7.4 Operating Systems

SCALE is developed and tested for operation on LINUX, UNIX, Windows, and Mac Intel operating systems.

D.7.5 Availability

Information about the code package can be obtained from the SCALE web site (<http://www.ornl.gov/sci/scale/>). The SCALE code can be obtained from the Radiation Safety Information Computational Center (<http://rsicc.ornl.gov/>) or the Organisation for Economic Co-operation and Development (OECD) Nuclear Energy Agency (NEA) Data Bank (<http://www.nea.fr/html/databank/welcome.html>) as code package CCC-0732.

D.7.6 Applications

The SCALE system [SCALE 2006] was developed to satisfy a need for a standardized method of analysis for the evaluation of nuclear fuel facility and package designs. At the heart of the SCALE code package are radiation transport methods that utilize both Monte Carlo and discrete-ordinate methods. In its present form, SCALE has the capability to perform criticality, shielding, radiation source term, spent fuel depletion/decay, and reactor physics analyses using well-established codes. The capability to perform such a wide breadth of analyses is a result of the modular design of the SCALE code system. SCALE integrates the calculation capabilities of many computer programs into easy-to-use automated sequences. The programs that perform the calculations are referred to as functional modules. The automated sequences are controlled by other programs referred to as control modules that read the user input, prepare input files for each functional module, call each module in sequence via the SCALE driver program, and automate the passing of data from one module to the next in the calculation sequence. Provisions have also been made to allow the user to execute the functional modules on a stand-alone basis.

The following discussion is focused on the Monte Carlo capabilities within SCALE. The SCALE package has the KENO and MONACO Monte Carlo codes that are used to solve eigenvalue and fixed-source three-dimensional radiation transport analyses, respectively.

KENO consists of two distinct 3D Monte Carlo criticality safety codes in SCALE (KENO-V.a and KENO-VI). These codes continue to be enhanced to provide users with greater capabilities [Bowman et al. 2007a]. In SCALE 6, these codes have

been extended to allow use of either multigroup or continuous-energy cross sections. KENO-V.a, which has been part of SCALE for 25 years, has a simplified geometry system that makes it very easy to use and is much more computationally efficient than other Monte Carlo criticality safety codes. The geometry models consist of units that are constructed by nesting partial or whole spheres, cylinders, and rectangular cuboids oriented along x -, y -, and z -axes. These units can be combined in rectangular arrays or inserted in other units. The restricted orientation of the units allow for very fast particle tracking, but still permit users to construct very complex geometry models. KENO-VI, which was introduced in SCALE about 10 years ago, provides users with a more general geometry system known as the SCALE Generalized Geometry Package (SGGP) for constructing geometry models. KENO-VI has a much larger assortment of bodies (e.g., cone, cuboid, cylinder, dodecahedron, elliptical cylinder, hexprism, plane, etc.). All KENO-VI bodies may intersect and may be arbitrarily rotated and translated. In addition to rectangular arrays, KENO-VI provides dodecahedral (a three-dimensional stack of dodecahedrons) and three hexagonal array types.

MONACO is a three-dimensional Monte Carlo code developed within SCALE for shielding calculations and will be released with SCALE 6. This new code is the result of a modernization effort combining the multigroup neutron and photon physics of the well-known MORSE Monte Carlo shielding code with the flexibility of the SGGP. MONACO uses the same cross-section package as other SCALE modules. Available tallies in MONACO include point detectors, region-based flux tallies, and mesh tallies. Any MONACO tally can be folded with a response function, either user-entered or from a standard list available with each SCALE cross-section library. Mesh tally values and uncertainties can be viewed with a special Java viewer on all computer platforms that run SCALE.

For fixed-source problems an advanced variance reduction scheme is used. The MAVRIC sequence in SCALE combines the results of an adjoint calculation from the 3D deterministic code TORT with the Monte Carlo MONACO. Both an importance map for weight windows and a biased source are automatically generated from the adjoint flux using the CADIS methodology. MAVRIC is completely automated and from a simple user input, it creates the cross sections (forward and adjoint), calculates the first collision source, computes the adjoint fluxes, creates the importance map and biased source, and then executes MONACO. Users can start and stop the calculation at various points so that progress can be monitored and importance maps can be reused for similar problems. For simple problems that do not require the advanced variance reduction, the MAVRIC sequence is an easy way to compute problem-dependent cross sections and execute Monaco with a common user input.

D.7.7 Significant Features

One of the significant and powerful capabilities in SCALE is the automated problem-dependent, cross-section processing capabilities for multigroup radiation transport analyses. SCALE includes both continuous-energy and multigroup ENDF/B-VI cross-section libraries. In terms of visualization capabilities, SCALE has the GeeWiz interface module that serves as a user-control center for executing many SCALE sequences.

With the GeeWiz tool, the user can develop input, perform the calculation, and view the output within a single graphical user interface package. In addition, GeeWiz provides a direct link with the KENO3D visualization package that allows users to interactively view their KENO geometry model as it is constructed. Additionally, SCALE provides the Javapeño interactive plotting program that permits users to plot TSUNAMI sensitivity data, flux profiles, reaction rate data, cross-section data, and covariance data. The KENO code also offers the option of generating its output in HTML format that can be viewed through a web browser, including interactive 2D Java applet plots. In summary, SCALE provides numerous Monte Carlo-based calculation sequences that can be used to perform various criticality safety and shielding analyses. These features enhance SCALE's ability to provide a powerful but easy-to-use set of nuclear fuel cycle safety analysis tools.

D.8 SRIM

D.8.1 Acronyms

SRIM is an acronym for **Stopping and Range of Ions in Matter**. The current version is SRIM-2010.

D.8.2 Principal Authors

The principal authors of the SRIM series of codes are James F. Ziegler and Jochen P. Biersack, although others have contributed. This summary is based on material taken from the SRIM web site (<http://www.srim.org/>).

D.8.3 Operating Systems

SRIM is actually a collection of programs, the main one being the DOS-based TRIM code (for **T**ransport of **I**ons in **M**atter), which originated in 1983. SRIM contains the 1998 version of TRIM and runs on Windows PCs having at least 16 MB of RAM with essentially any Windows operating system. SRIM does not run on Unix or Macintosh systems. However, the DOS version of SRIM may run on a Unix PC that has a DOS emulator.

D.8.4 Availability

The SRIM code can be downloaded from the web site at <http://www.srim.org/>.

D.8.5 Applications

The SRIM code simulates the transport of heavy ions of less than 2 GeV/amu in matter. The models are quantum mechanical and, thus, statistical, in the sense that the ions make macroscopic moves between collisions but then the collision results are averaged. This procedure, common to most charged-particle transport algorithms, leads

to increased efficiency. An ion treated in SRIM is any charged particle as heavy as a proton or greater. Matter is any collection of atoms of a gas, a liquid, or a solid up to atomic number $Z = 92$.

The SRIM code is based on the TRIM code, which is a comprehensive ion-transport module. TRIM treats objects that can be quite complex. The object can be made of up to eight layers, each layer being composed of a compound material. TRIM estimates the final distribution of the ions (in three dimensions). TRIM also is able to estimate kinetic phenomena that are induced by an ion's energy loss. These phenomena include sputtering, target damage, the production of phonons, and ionization. SRIM models the cascades that result from ion impact on target atoms.

The SRIM package can be used to generate tables of stopping powers, ranges of ions in matter, and straggling distributions. SRIM also has been used to study ion implantation, ion sputtering, and ion beam therapy.

The most recent and complete description of SRIM is provided by the textbook by [Ziegler et al. \[2008\]](#). Prior to publication of this book, the main reference for SRIM and TRIM was [Ziegler et al. \[1985\]](#). Both references provide thorough descriptions of the models on which SRIM is based. These references describe the basic physics models of ion transport into matter. The TRIM code was initially based on light ion transport, but SRIM now treats ion transport in matter where both the ion and the atoms in the matter include all elements up to uranium.

D.8.6 Significant Features

The SRIM code allows interruption and resumption. This feature means that intermediate results are not necessarily lost but the simulation can be resumed following some intermediate inquiry on the part of the user. The SRIM code allows ion transport plots to be viewed and saved. Plots of ion tracks can be viewed during calculation, but the primary output of SRIM are files that summarize the detailed ion transport results.

The range algorithms on which SRIM is based are described by [Biersack and Haggmark \[1980\]](#). The general theory of ion stopping is treated in various works by [Ziegler \[1977, 1978, 1980a,b\]](#) and by [Ziegler and Manoyan \[1988\]](#). Typically, SRIM is updated about every five years. The last major changes were in 1995, 1998, 2003, and 2008. A tutorial manual is included in the SRIM package.

D.9 TRIPOLI

D.9.1 Acronym and Current Version

TRIdimensionnel POLIcinétique is a three-dimensional continuous-energy Monte Carlo code. TRIPOLI-4 is the fourth generation of the code. Presently maintained are versions 4 and 5 (T-4 v4 and v5).

D.9.2 Principal Authors

The principal authors are the TRIPOLI project team, CEA-Saclay, and SERMA R&D Unit.

D.9.3 Operating Systems

Linux (32 and 64 bits), AIX, Solaris, and OSF1.

D.9.4 Availability

As of 2010, version 4 is available from the NEA Data Bank (<http://www.nea.fr>), as well as from RSICC (<http://www.rsicc.org>). Version 5 is available for CEA, EDF, and their respective subsidiaries.

D.9.5 Applications

TRIPOLI is mainly used for radiation protection and shielding, core physics, and criticality calculations. It is the Monte Carlo standard for all CEA experimental or naval reactors and nuclear cycle activities as well as for the EDF Nuclear Power Plants. It is also the reference calculation route for the French criticality community, in the CRISTAL criticality computing tool package (CEA, AREVA, EDF, and subsidiaries).

The code is designed for neutron and photon transport in the energy range from 0 to 20 MeV, and it uses continuous-energy cross sections. TRIPOLI-4 produces classic Monte Carlo output such as flux and derived quantities (e.g., reaction rates), current, and k_{eff} for criticality and core physics problems. Time-dependent problems may be addressed as well as perturbation calculations by employing the correlated sampling method.

The code has a flexible parallel mode, enabling it to easily run on heterogeneous networks of processors as well as on massively parallel machines. The code directly reads ENDF and PENDF files produced by the NJOY nuclear data-processing system and is compatible with all nuclear data evaluations in the ENDF/B format such as JEF2, JEFF3, ENDF/B-VI, ENDF/B-VII, JENDL3.3, etc. The user may easily mix materials from different nuclear data evaluations, a feature that also makes TRIPOLI-4 a valuable tool for the nuclear data community.

D.9.6 Significant Features

For radiation protection and shielding, one needs efficient variance reduction tools, among which, the geometry and energy domain attractors are necessary to address deep penetration problems. TRIPOLI, thus, offers users an importance map, automatically estimated by the code either analytically or from a simplified adjoint calculation using the Dijkstra algorithm. This variance reduction feature is available in versions 4 and 5.

For radiation protection applications, Green's functions may be computed to perform source parametric calculations, i.e., different sources can be multiplied by the precomputed Green's functions thereby providing high-quality results in a short time. This functionality is available in versions 4 and 5 (with parallel operation in v5).

A surface restart feature is another approach for parametric problems. With this capability an initial run is made and surface sources are tallied and then used for several subsequent runs, each with possibly different variance reduction options. Reactor

dismantling studies are a good example of such an application; a neutron surface source is first computed at the reactor boundary and is then used to compute tallies at different locations in the concrete outside the reactor.

For core physics applications, entropy diagnostics are provided to assess the source convergence (with both Shannon and Boltzmann entropies).

In version 4 the geometry may be input in either a combinatorial mode or as surface-bounded volumes, and the particle trajectories are tracked with the native TRIPOLI-4 geometry module. Beginning with version 5, TRIPOLI-4 is compatible with the ROOT geometry module developed by the high-energy physics community and maintained by CERN. All ROOT-compatible pre- and postprocessors are therefore available for a TRIPOLI simulation.

Developed separately from TRIPOLI-4 is the SALOME TRIPOLI (<http://www.salome-platform.org>) preprocessor, offering a full three-dimensional graphical environment to make a complete TRIPOLI-4 input deck. This preprocessor is available from the NEA Data Bank and RSICC (2008).

Bibliography

COG

- BUCK, R.M., D.E. CULLEN, D.P. HEINRICHS, E.M. LENT, D.E. NIELSON JR., AND K.E. SALE 2007, *COG—Publicly Available Now to Criticality Safety Practitioners*, UCRL-CONF-224715, 05/28/2007, (ICNC-2007).
- BUCK, R.M. AND E.M. LENT 1993, *COG: A New, High-Resolution Code for Modeling Radiation Transport*, UCRL-52000-93-6.
- BUCK R.M. AND J.M. HALL 1999, *Applications of the COG Multiparticle Monte Carlo Transport Code to Simulated Imaging of Complex Objects*, UCRL-JC-134630 (preprint).
- CULLEN, D.E., L.F. HANSEN, E.M. LENT, AND E.F. PLECHATY 2003, *Thermal Scattering Law Data: Implementation and Testing Using the Monte Carlo Neutron Transport Codes COG, MCNP and Tart*, UCRL-ID-153656, 05/17/2003.
- CULLEN, D.E., R.N. BLOMQUIST, C.J. DEAN, S. GOLUOGLU, Y.-K. LEE, E.L. LENT, R. MACFARLANE, S. MCKINLEY, E.F. PLECHATY, AND J.C. SUBLET 2007, *ENDF/B-VII.0 Data Testing for Three Fast Critical Assemblies*, UCRL-TR-233310, 07/31/2007.
- DESCALLE, M.-A., R.M. BUCK, E.M. LENT, AND K.E. SALE 2004, *COG Validation: SINBAD Benchmark Problems*, UCRL-TR-202654, 03/02/2004.
- HEINRICHS, D.P. AND E.M. LENT 2003, *Photonuclear Benchmarks with a Comparison of COG and MCNPX Results*, UCRL-CONF-200552, 10/15/2003.
- VERBEKE, J.M., A. DOUGAN, L.F. NAKAE, K.E. SALE, AND N.J. SNYDERMAN 2007, *Neutron Correlations in Special Nuclear Materials, Experiments and Simulations*, UCRL-PROC-231582, 06/07/2007.

EGSnrc

- KAWRAKOW, I. 2000a, "Accurate Condensed History Monte Carlo Simulation of Electron Transport. I. EGSnrc, the new EGS4 version," *Med. Phys.*, **27**, 485–498.
- KAWRAKOW, I. 2000b, "Accurate Condensed History Monte Carlo Simulation of Electron Transport. II. Application to Ion Chamber Response Simulations," *Med. Phys.*, **27**, 499–513.

- KAWRAKOW, I. 2005, *EGSnrc C++ Class Library*, Technical Report PIRS-898, National Research Council of Canada, Ottawa, Canada.
- KAWRAKOW, I. AND A.F. BIELAJEW 1998a, "On the Representation of Electron Multiple Elastic-Scattering Distributions for Monte Carlo Calculations," *Nucl. Instrum. Meth. B*, **134**, 325–336.
- KAWRAKOW, I. AND A.F. BIELAJEW 1998b, "On the Condensed History Technique for Electron Transport," *Nucl. Instr. and Meth. B*, **142**, 253–280.
- KAWRAKOW, I. AND W.O. ROGERS 2003, *The EGSnrc Code System: Monte Carlo Simulation of Electron and Photon Transport*, Report PIRS-701 (4th printing), National Research Council of Canada, Ottawa.
- ROGERS, D.W.O., B.A. FADDEGON, G.X. DING, C.-M. MA, J. WEI, AND T.R. MACKIE 1995, "BEAM: A Monte Carlo Code to Simulate Radiotherapy Treatment Units," *Med. Phys.*, **22**, 503–524.
- ROGERS, D.W.O., I. KAWRAKOW, J.P. SEUNTJENS, B.R.B. WALTERS, AND E. MAINEGRAHNG 2003, *NRC User Codes for EGSnrc*, Technical Report PIRS-702 (RevB), National Research Council of Canada, Ottawa.
- ROGERS, D.W.O., B. WALTERS, AND I. KAWRAKOW 2005, *BEAMnrc Users Manual*, NRC Report PIRS 509(a) (revI).
- WALTERS, B.R.B., I. KAWRAKOW, AND D.W.O. ROGERS 2005, *DOSXYZnrc Users Manual*, NRC Report PIRS 794 (rev B).

GEANT

- AGOSTINELLI, S. et al. 2003, "GEANT4—A Simulation Toolkit," *Nucl. Instrum. Meth. B*, **506**, 250–303.
- ALLISON, J. et al. 2006, "GEANT4 Developments and Applications," *IEEE T. Nucl. Sci.*, **53**, 270–278.
- ANNINOS, D., et al. 2006, "Recent Developments and Upgrades to the Geant4 Geometry Modeler," CHEP'06 (Mumbai, India, February 2006).
- APOSTOLAKIS, J., A. BAGULYA, S. ELLES, V.N. IVANCHENKO, O. KADRI, M. MAIRE, AND L. URBAN 2008, "The Performance of the Geant4 Standard EM Package for LHC and Other Applications," *J. Phys. Conf. Series*, **119**, 032004.
- ASO, T., A. KIMURA, S. TANAKA, H. YOSHIDA, N. KANEMATSU, T. SASAKI, AND T. AKAGI 2005, "Verification of the Dose Distributions with Geant4 Simulation for Proton Therapy," *IEEE Trans. Nucl. Sci.*, **52**(4), 896–901.
- BANERJEE, S. 2008a, "Validating Geant4 Versions 7.1 and 8.3 Against 6.1 for BaBar," *J. Phys. Conf. Series*, **119**, 032007.
- BANERJEE, S. 2008b, "Readiness of CMS Simulation Towards LHC Startup," *J. Phys. Conf. Series*, **119**, 032006.
- KASE, Y., N.K. ANEMATSU, T. KANA, AND N. MATSUFUJI 2006, "Biological Dose Calculation with Monte Carlo Physics Simulation for Heavy-ion Radiotherapy," *Phys. Med. Biol.*, **51**, N467–N475.
- UEDA, et al. 2010, "The Status of the Simulation Project for the ATLAS Experiment in View of the LHC Startup," *J. Phys. Conf. Series*, **219**, 032060.
- WRIGHT, D.H., T. KOI, G. FOLGER, V. IVANCHENKO, M. KOSSOC, N. STARKOV, A. HEIKKINEN, AND H.-P. WELLISCH 2006, "Recent Developments and Validations in Geant4 Hadronic Physics," *AIP Conf. Proc.*, **867**, 479–486.
- WRIGHT, D.H., T. KOI, G. FOLGER, V. IVANCHENKO, M. KOSSOC, N. STARKOV, A. HEIKKINEN, AND H.-P. WELLISCH 2007, "Low and High Energy Modeling in Geant4," *AIP Conf. Proc.*, **896**, 11–20.

MCNP

- X-5 MONTE CARLO TEAM 2008, "MCNP – A General Monte Carlo N-Particle Transport Code, Version 5," Los Alamos Report LA-UR-03-1987, Los Alamos Nat. Lab., Los Alamos, NM.
- HENDRICKS, J.S., et al. 2005, "MCNPX Extensions Version 2.5.0," Los Alamos Report LA-UR-05-2675, Los Alamos Nat. Lab., Los Alamos, NM.
- HENDRICKS, J.S., et al. 2008, "MCNPX Extensions Version 2.6.0," Los Alamos Report LA-UR-08-2216, Los Alamos Nat. Lab., Los Alamos, NM.

MCSHAPE

- FERNANDEZ, J.E. AND V.G. MOLINARI 1993, "Diffusion of Polarized Photons in the Frame of the Transport Theory," *Nucl. Instrum. Meth. B*, **73**, 341–351.
- FERNANDEZ, J.E., V.G. MOLINARI, M. BASTIANO, AND V. SCOT, 2004, "Diffusion of Photons with Arbitrary State of Polarization: The Monte Carlo Code MCSHAPE," *Nucl. Instrum. Meth. B*, **213**, 105–107.
- FERNANDEZ, J.E. AND V. SCOT, 2007a, "Deterministic and Monte Carlo Codes for Multiple Scattering Photon Transport with Arbitrary Polarization States," *Spectrochim. Acta B*, **62**, 517–528.
- FERNANDEZ, J.E. AND V. SCOT 2007b, "Unbiased Monte Carlo Simulation of the Compton Profile," *Nucl. Instrum. Meth. B*, **263**, 209–213.
- FERNANDEZ, J.E. AND V. SCOT, 2009, "Simulation of the Detector Response Function with the Code MCSHAPE," *Radiat. Phys. Chem.*, **78**(10), 882–887.
- SCOT, V., J.E. FERNANDEZ, L. VINCZE, AND K. JANSSENS 2007, "3D Extension of the Monte Carlo Code MCSHAPE for Photon-Matter Interactions in Heterogeneous Media," *Nucl. Instrum. Meth. B*, **263**, 204–208.
- STOKES, G.G. 1852, "On the Composition and Resolution of Streams of Polarized Light from Different Sources," *Trans. Cambridge. Phil. Soc.*, **9**, 399–416.

PENELOPE

- ACOSTA, E., X. LLOVET, AND F. SALVAT 2002, "Monte Carlo Simulation of Bremsstrahlung Emission by Electrons," *Appl. Phys. Lett.*, **80**, 3228–3230.
- BARÓ, J., J. SEMPÁU, J.M. FERNÁNDEZ-VAREA, AND F. SALVAT 1995, "PENELOPE: An Algorithm for Monte Carlo Simulation of the Penetration and Energy Loss of Electrons and Positrons in Matter," *Nucl. Instrum. Meth. B*, **100**, 31–46.
- BRUSA, D., G. STUTZ, J.A. RIVEROS, J.M. FERNÁNDEZ-VAREA, AND F. SALVAT 1996, "Fast Sampling Algorithm for the Simulation of Photon Compton Scattering," *Nucl. Instrum. Meth. A*, **379**, 167–175.
- FERNÁNDEZ-VAREA, J.M., R. MAYOL, J. BARÓ, AND F. SALVAT 1993, "On the Theory and Simulation of Multiple Elastic Scattering of Electrons," *Nucl. Instrum. Meth. B*, **73**, 447–473.
- SALVAT, F. AND J.M. FERNÁNDEZ-VAREA 1993, "Semiempirical Cross Sections for the Simulation of the Energy Loss of Electrons and Positrons in Matter," *Nucl. Instrum. Meth. B*, **63**, 255–269.
- SALVAT, F., J.M. FERNÁNDEZ-VAREA, AND J. SEMPÁU 2006, "PENELOPE-2006: A Code System for Monte Carlo Simulation of Electron and Photon Transport," (OECD/Nuclear Energy Agency, Issy-les-Moulineaux, France, 2006).
- SEMPÁU, J., E. ACOSTA, J. BARÓ, J.M. FERNÁNDEZ-VAREA AND F. SALVAT 1997, "An Algorithm for Monte Carlo Simulation of the Coupled Electron-photon Transport," *Nucl. Instrum. Meth. B*, **132**, 377–390.
- SEMPÁU, J., J.M. FERNÁNDEZ-VAREA, E. ACOSTA, AND F. SALVAT 2003, "Experimental Benchmarks of the Monte Carlo Code PENELOPE," *Nucl. Instrum. Meth. B*, **207**, 107–123.

SCALE

- BOWMAN, S.M., M.D. DEHART, M.E. DUNN, S. GOLUOGLU, J.E. HORWEDEL, L.M. PETRIE, Jr., B.T. REARDEN, AND M.L. WILLIAMS 2007, "New Criticality Safety Analysis Capabilities in SCALE 5.1," p. 403–407, *Proc. 8th International Conf. on Nuclear Criticality Safety (ICNC 2007)*, Vol. I, May 28–June 1, 2007, St. Petersburg, Russia.
- BOWMAN, S.M., B.T. REARDEN, AND J.E. HORWEDEL 2007, "GeeWiz Integrated Visualization Interface for SCALE 5.1," p. 12–16, *Proc. 8th Inter. Conf. on Nuclear Criticality Safety (ICNC 2007)*, Vol. II, May 28–June 1, St. Petersburg, Russia.
- DEHART, M.D. AND S.M. BOWMAN 2007, "Improved Radiochemical Assay Analyses Using TRITON Depletion Sequences in SCALE," *Proc. of IAEA Advances in Applications of Burnup Credit to Enhance Spent Fuel Transportation, Storage, Reprocessing and Disposition*, Aug. 29–Sept. 2, 2005, London, IAEA-TECDOC-CD-1547, Session 2, p. 99–108.
- EMMETT, M.B. 2000, MORSE: Present Capabilities and Future Directions, *Appl. Radiat. Isotopes*, **53**(4-5), 863–866.
- GAULD, I.C. AND C.E. SANDERS 2001, "Development and Applications of a Prototypic SCALE Control Module for Automated Burnup Credit Analysis," *Proc. 2001 ANS Embedded Topical Meeting on Practical Implementation of Nuclear Criticality Safety*, Reno, Nevada, Nov. 11–15, 2001.
- GOLUOGLO, S., M.E. DUNN, N.M. GREENE, L.M. PETRIE, and D.F. HOLLENBACK 2007, "Generation and Testing of the Continuous-Energy Cross-Section Library for Use with Continuous-Energy Versions of KENO," p. 364–366, *Proc. 8th Inter. Conf. on Nuclear Criticality Safety (ICNC 2007)*, Vol. I, May 28–June 1, 2007, St. Petersburg, Russia.
- HAGHIGHAT A. AND J.C. WAGNER 2003, "Monte Carlo Variance Reduction with Deterministic Importance Functions," *Prog. Nucl. Energ.*, **42**(1), 25–53.
- REARDEN, B.T. 2004, "Perturbation Theory Eigenvalue Sensitivity Analysis With Monte Carlo Techniques," *Nucl. Sci. Eng.*, **146**(3), 367–382.
- SCALE 2006, *SCALE: A Modular Code System for Performing Standardized Computer Analyses for Licensing Evaluations*, ORNL/TM-2005/39, Version 5.1, Vols. I-III, November 2006. Available from Radiation Safety Information Computational Center at Oak Ridge National Laboratory as CCC-732.
- WILLIAMS M.L., S. GOLOUOGLU, AND L.M. PETRIE 2005, "Recent Enhancements to the SCALE 5 Resonance Self-Shielding Methodology," *T. Am. Nucl. Soc.*, **92**, 751–753.

SRIM

- BIERSACK, J.P. AND L. HAGGMARK 1980, "A Monte Carlo Computer Program for the Transport of Energetic Ions in Amorphous Targets," *Nucl. Instrum. Meth. A*, **174**, 257–269.
- ZIEGLER, J.F. 1977, *Hydrogen Stopping Powers and Ranges in all Elements*, Pergamon Press, New York.
- ZIEGLER, J.F. 1978, *Helium Stopping and Ranges in All Elements*, Pergamon Press, New York.
- ZIEGLER, J.F. 1980a, "The Stopping of Energetic Ions in Solids," *Nucl. Instrum. Meth. A*, **168**, 17–24.
- ZIEGLER, J.F. 1980b, *Handbook of the Range Distributions of Energetic Ions in All Elements*, Pergamon Press, New York.
- ZIEGLER, J.F. 2004, "SRIM-2003," *Nucl. Instrum. Meth. B*, **219-220**, 1027–1036.
- ZIEGLER, J.F., J.P. BIERSACK, AND U. LITTMARK 1985, "The Stopping and Range of Ions in Solids," in *Stopping Powers and Ranges of Ions in Matter*, Pergamon Press, New York.
- ZIEGLER, J.F. AND J.M. MANOYAN 1988, "The Stopping of Ions in Compounds," *Nucl. Instrum. Meth. B*, **35**, 215–228.

ZIEGLER, J.F., J.P. BIRSACK, AND M.D. ZIEGLER 2008. *The Stopping and Range of Ions in Matter*, Lulu Press, Morrisville, NC.

TRIPOLI

- BOTH, J.P., et al. 1994, "A Survey of TRIPOLI-4," *Proc. 8th Intern. Conf. on Radiation Shielding*, Arlington, TX, April, 373–380.
- BOTH J.P. 2000, "Treatment of Cross Section Uncertainties in the Transport Monte Carlo Code TRIPOLI-4," *J. Nucl. Sci. Technol.*, Supplement 1, 420–426.
- BOURGANEL, S. AND O. PETIT 2007, "TRIPOLI-4: Green's functions," *ANS/ENS Inter. Meeting*, Washington, dc, November.
- GOMIT, J.M., et al. 1999, "ICNC'9, The New CRISTAL Criticality Safety Package," *Sixth Inter. Conf. on Nuclear Criticality Safety*, Versailles, France, September, 308–316.
- LEE, Y.K. 2003, "Analysis of the NRC PCA Pressure Vessel Dosimetry Benchmark Using TRIPOLI-4.3 Monte Carlo Code and ENDF/B-VI, JEF-2.2 and IRDF-90 Libraries," *Nuclear Mathematical and Computational Science*, Gatlinburg, TN, April.
- LEE, Y.K. 2008, "Analysis of the VENUS-3 PWR Pressure Vessel Surveillance Benchmark Using TRIPOLI-4.3 Monte Carlo Code," in "Reactor Dosimetry," STP 1490, D.W. Vehar, D.G. Gilliam, and J.M. Adams (eds.), ISBN13: 978-0-8031-3412-6, ASTM, West Conshohocken, PA.
- LEE, Y.K. AND F.X. HUGOT 2007, "TRIPOLI-4: Verification and Validation," *ANS/ENS Inter. Meeting*, Washington, dc, November.
- PÈNELIAU, Y.P. AND J.P. BOTH 1999, "Parallelization of the Monte Carlo Code TRIPOLI-4," *Mathematical and Computation, Reactor Physics and Environmental Analysis in Nuclear Applications*, Madrid, Spain, September, 412.
- PETIT, O. AND C. DIOP 2007, "TRIPOLI-4: The Automatic Variance Reduction Scheme," *ANS/ENS Inter. Meeting*, Washington, dc, November.
- TRAMA, J.C. AND C. JOUANNE 2007, "TRIPOLI-4: Nuclear Data Management," *ANS/ENS Inter. Meeting*, Washington, dc, November.
- TRAMA, J.C. AND F.X. HUGOT 2007, "TRIPOLI-4: Parallelism Capability," *ANS/ENS Inter. Meeting*, Washington, dc, November.

E Minimal Standard Pseudorandom Number Generator

Numerous pseudorandom number generators are in current use around the world, and there are many reasons given for preferring one over another. Chapter 3 of this text discusses various ways of generating pseudorandom numbers, by which is meant numbers uniformly distributed on the unit interval $(0, 1)$. Many of the examples and problems at the ends of the chapters in this text require a method of generating pseudorandom numbers. Different pseudorandom number generators will yield slightly different results, and the same pseudorandom number generator will yield a string of pseudorandom numbers that depends on the “seed” that is used to start the sequence. Thus, the answer one person may obtain to an example or a problem that requires pseudorandom numbers may differ from the answer another reader obtains to the same problem. This should not be viewed with alarm, as almost all Monte Carlo answers are approximate and small differences among solutions using different pseudorandom number sequences are inevitable. Nevertheless, there may be some merit to providing a basic pseudorandom number generator that readers can use if they wish.

As discussed in Section 3.5.3, [Park and Miller \[1988\]](#) proposed a *minimal standard* pseudorandom number generator to produce pseudorandom numbers uniformly over the interval $(0, 1)$. It is this generator, with an initial seed of 73,907, that is used for the many examples throughout this book. In this appendix, the generator is coded in four computer languages to allow you, the reader, to duplicate the example results. These routines will work correctly on any machine of which the maximum integer is $2^{31} - 1$ or larger.

E.1 FORTRAN77

In FORTRAN77, variables beginning with the letters *i–n* are, by default, integers and all others are real. Thus, the type declarations in the routine below are not necessary, but are included for good programming practice.

```
      FUNCTION random(iseed)
C -----
C Park & Miller minimal standard for a RN generator
C -----
      INTEGER iseed,ia,im,iq,ir,k
      REAL random,am
      PARAMETER (ia=16807, im=2147483647, am=1./im,
&               iq=127773, ir=2836)
      k = iseed/iq
```

```

iseed = ia*(iseed-k*iq)-ir*k
IF(iseed.LT.0) iseed = iseed+im
random = am*iseed
RETURN
END

```

E.2 FORTRAN90

Although the above FORTRAN77 routine also works in FORTRAN90, here it is using the new constructs allowed by FORTRAN90.

```

REAL FUNCTION random(seed)
! -----
! Park & Miller minimal standard for a RN generator
! -----
IMPLICIT NONE
INTEGER :: temp,seed
INTEGER, PARAMETER :: a=16807, m=2147483647,&
& q=127773, r=2836
REAL,PARAMETER :: minv=1./m
temp = seed/q
seed = a*(seed-temp*q)-r*temp
if(seed < 0) seed = seed + m
random = minv*seed
return
END FUNCTION random

```

E.3 Pascal

Although Pascal is not used much anymore, it is with this language that [Park and Miller \[1988\]](#) first implemented their algorithm.

```

function random : real;
(* integer version of the minimal standard *)
const
  a = 16807;
  m = 2147483647;
  q = 127773;
  r = 2836;
var
  lo, hi, test : integer
begin
  hi := seed div q;
  lo := seed mod q;
  test := a * lo - r * hi;
  if test > 0 then
    seed := test
  else
    seed := test + m;
  random := seed / m
end;

```

E.4 C and C++

The following routine works in both C and C++. The initial value of `seed` is set outside this routine (and never altered subsequently except by this routine). A random number is then generated with `random(&seed)`.

```
float random(int *seed)
// -----
// Park & Miller minimal standard for a RN generator
// -----
{
    int temp;
    const int A=16807, M=2147483647, Q=127773, R=2836;
    const float minv=1./M;

    temp = *seed/Q;
    *seed = A*(*seed-temp*Q)-R*temp;
    if(*seed < 0)
    {
        *seed = *seed + M;
    }

    return minv * (float)*seed;
}
```

E.5 Programming Considerations

The above routines have the initial seed x_0 set outside the routine and subsequent values of x_i are stored in this variable by the routine. The value of `seed` must only be changed from its initial value by the routine `random`. The returned value `random` is $\rho_i = x_i/m$.

Although these function routines work well, they can be modified to use special features of the various languages so that no function argument is needed. For example, in FORTRAN77 a labeled COMMON block could be used to store the value of `iseed`. In FORTRAN90, this function would be better written as a module so that the seed need not be passed as an argument. In C++, the function could be written as a new class.

Bibliography

PARK, S.K. AND K.W. MILLER 1988, "Random Number Generators: Good Ones Are Hard to Find," *Commun. ACM*, **31**, 1192–1201.

Index

A

- Absorbing media, 277–278
- Absorption states, 209
- Accuracy of tally, 90
 - random number generator, 90
 - user errors, 90
- Adjoint
 - flux density, 261, 264
 - importance, 264
 - linear equations, 207–209
 - operator, 261
 - transport equation, (*see* Transport equation)
- Antithetic variates, 117–120

B

- Bayes
 - Bayesian analysis, 153
 - theorem, 33, 154–159
 - decision theory, 159
 - making inferences examples, 162
 - making inferences, 159–166
 - vs classical analysis, example, 156
- Bernoulli distribution, 325
- Bernoulli, 3, 337
- Beta distribution, 159, 314–316
- Biased estimators, 126
- Bienaymé-Chebyshev inequality, 338
- Binomial distribution, 326–329
- Bivariate normal distribution, 150, 333
- Boltzmann equation, (*see* Transport equation)
- Buffon's needle problem
 - analytic result, 9, 19
 - Bayesian analysis, example, 155–157
 - history, 5
 - inverse MC, example, 194
 - MC analog simulation, 41
 - MC quadrature, 10
 - using biasing and weights, 97
 - using Markov chains, 150
 - with antithetic sampling, 120
 - with correlated sampling, 115

- with importance sampling, 103
 - with weighted MC quadrature, 128
- Buffon, Comte de, 3
 - Burn-in, 143

C

- Cauchy distribution, 320, 340
- CDF (*see* Probability distributions)
- Center-of-mass system, 284
- Central limit theorem, 37–39, 344
- Central moments, 342
- Chandrasekhar's H -function, 294
- Chi-squared distribution, 321
- Combinatorial geometry, 271
 - surfaces used, 271
- Combining variance reduction methods, 124
- Composition sampling method, 76
- Composition-rejection sampling method, 79
- Compton scattering, 283
- Computer precision, 275–277
- Conditional PDF, (*see* Probability distributions)
- Confidence interval, 39
- Congruential generators, 48–60, 65
- Convergence
 - almost surely, 339
 - in probability, 338
 - strongly, 339
 - weak, 338
- Correlated sampling, 113–120
 - antithetic variates, 116–120
 - with one known expected value, 114–117
- Correlation coefficient, 334
- Covariance, population, 32
- Cranberg formula, 96
- Cross section
 - differential scattering, 283
 - differential, 247
 - for mixtures, 248
 - interaction coefficient, 242–244
 - macroscopic, 242–244

Cross section (*continued*)

- mass interaction coefficient, 248
- microscopic, 247
- neutron scattering, 284
- scattering coefficient, 245–247

Cumulative distribution, definition, 23

Current vector, 241

D

Decomposition sampling method, 78

Degrees of freedom, 28, 321, 323

Differential equations

- algebraic equation approximation, 217
- Helmholtz equation, 225–229
 - 2D, 225–229
 - 3D, 229
- analytic solution, 228
- constant source, 226–228
- Green's function, 223
- MC solution example, 228
- MC solution, 225, 230
- variable source, 223

Laplace's equation

- 2D, 218–220
- 3D, 220
- analytic solution, 217
- MC solution, 219

MC solution, 217

Poisson's equation, 222–225

- constant source, 222
- Green's function, 223
- MC example, 223
- MC solution example, 223
- variable source, 223

walk on circles method, 220, 226

- MC example, 228

walk on spheres method, 221, 229

- MC example, 223

Diffusion equation, (*see* Helmholtz equation)

Dimensionality reduction, 122

Dirac delta function, 30

Direction cosines, scattered particle, 279–283

Direction vector, 236

Distributions, (*see* Probability distributions)**E**

Eigenvalue problems, 230

Emission rate density, 255

Error function, 317

Escape probability

- purely absorbing medium, 277
- with scattering, 293

Estimating underlying PDF, 85, 185–194

Exponential attenuation, 244

Exponential distribution, 24, 71, 310–312

- sampling from, 71, 311

F

Fermi, 6

Figure of merit, 83

Fission neutrons, 96

Fixed source MC problem, 273

Fluence, 240

- at a point, 290–292

Flux density, 239, 251

- adjoint, 264

- interpretation, 239

Fredholm integral equations, (*see* Integral equations)**G**

Gamma distribution, 165, 307, 312–314

Gaussian distribution, 25, 317

Geometric distribution, 329

Gibb's sampler, 152

H

Hastings ratio, 138

Helmholtz equation, (*see* Differential equations)

Hermitian operator, 217, 230

I

Importance sampling, 101–106

- example, 104

- Monte Carlo quadrature, 105

- vs stratified sampling, 112

Incomplete beta function, 314

Incomplete gamma function, 312, 318

Inner product, 204, 212, 215, 261

Integral equations, 211–216

- deterministic solutions, 211

- Fredholm, 174, 211

- MC solution, 212–214, 302

- example, 215–216

Integral transport equation, (*see* Transport equation)

- formal solution, 296

Integrals, (*see also* Quadrature)

- as expected values, 9
- evaluation by MC, 10
- MC examples, 39–41
- multidimensional, 39, 41, 87
- with quasi-random numbers, 129

Integration volume partitioning, 121

Interaction coefficient, (*see* Cross section)

Inverse CDF sampling, 70–72

Inverse MC, 171–199

- by iteration, 177
- by simulation, 196–198
- existence and uniqueness, 172
- symbolic, 178–194
- system design optimization, 176
- uncertainties, 179–181

Inverse problem, formulation, 173–176

- integral formulation, 174
- practical formulation, 171
- sensitivity analysis, 181

Isotropic scattering, 32

Isotropic sources, 273

J

Joint distribution, 31, 33, 333–335

- sampling from, 80

K

Kelvin, 6

L

L'Ecuyer random number generator, 62

Laboratory coordinate system, 285

Laplace's equation, (*see* Differential equation)

Law of large numbers, 35, 337–340

- weak vs strong laws, 340

Leakage Estimator, 292

Leclerc, (*see* Buffon)

Legendre polynomial expansion, 284

Likelihood distribution, 156, 161

Linear algebraic equations, 203–211

- adjoint, Markov chains, 204
- finite Markov chain solution, 209
- iterative solution, 203
- Markov chain solution, 204–206
- example, 207

Linear Differential equations, (*see* equations)

Linear integral equations, (*see* Integral equations)

Logistic distribution, 309

Lognormal distribution, 319

M

Marginal distribution, 31, 155

Markov chains, 133–153

- Bayes theorem, with data, 161–166
- burn-in, 143
- continuous, 215
- discrete vs continuous, 135
- ergodic, 136
- example for an integral, 142
- finite, 209
- history, 136
- linear algebraic equations, 203–211

MCMC, (*see* Markov chains)

Mean free path, 245

Mean

- population, 25, 29
- sample, 27
- sum of random variables, 34

Metropolis ratio, 138

Metropolis-Hasting algorithm, 133, 137–143

- example, Buffon's needle problem, 150
- Gibb's sampler, 152
- multidimensional sampling, 148–150
- proposal distribution, 144–148
- sample distribution, 140

Metropolis, 7, 133, 137

Moment generating function, 341

- for normal distribution, 343
- properties, 342
- uniqueness, 344

Monte Carlo transport codes, (*see* Transport codes)

Monte Carlo, (*see also* Inverse MC)

- accuracy vs precision, 89–93
- analog vs nonanalog, 41, 97
- as quadrature, 8–12, 39, (*see also* Quadrature)
 - antithetic variates, 117–119
 - dimensionality reduction, 121
 - example, 10, 86
 - importance sampling, 105
 - systematic sampling, 110
 - transformation example, 100
 - volume partitioning, 121
- as simulation, 12–15, 39
- biased estimators, 125

Monte Carlo (*continued*)

- direct vs inverse, 171–173
- early history, 2–8
- integral transport equation, 296–300
 - as simulation, 299
- linear algebraic equations, 42, 203–211
- multidimensional quadrature, 39, 40, 87
- overview, 1
- particle transport, (*see also* Particle transport)
 - simulation, 269–304
- solution of integral equations, 296
- variance reduction
 - antithetic variates, 117–120
 - basic methods, 98
 - biased estimators, 125
 - combining methods, 124
 - correlated sampling, 113–119
 - dimensionality reduction, 121
 - exponential transformation, 303
 - implicit absorption, 303
 - importance sampling, 102–106, 288–289
 - interaction forcing, 303
 - Russian roulette, 122–124, 302
 - splitting, 122–124, 302
 - stratified sampling, 111
 - systematic sampling, 107–111
 - transformations, 100
 - truncation methods, 302
 - weighted MC integration, 127
- Multinomial distribution, 334
- Multiplicative congruential generator, (*see* Pseudorandom numbers)
- Multivariate normal distribution, 333

N

- Negative binomial distribution, 329
- Neumann expansion, 213, 296
- Neumann series, 214
- Neutron scattering, 284
- Next event estimator, 290
- Normal distribution, 24, 37, 309, 317

O

- One-speed particle transport examples, 293
- Operators, 261
- Optical thickness, 257
- Ordered random variates, for quadrature, 126

P

- Pareto distribution, 309, 323
- Partial current, 241
- Particle density, 238
- Particle reflection from half-space, 294
- Particle transport, (*see also* Transport codes; Transport equation)
 - deterministic vs stochastic, 269
 - integral equation, 255–260, 296–300
 - integro-differential form, 255
- Monte Carlo, 270–302
 - absorbing medium, 277
 - analog vs nonanalog, 286
 - choosing interaction type, 277
 - Compton scattering, 283
 - computer precision, 275
 - examples, 277, 293
 - geometry specification, 270–273
 - one-speed, 293–295
 - path-length estimation, 274–277
 - scattering, 279–283
 - scoring, 286–293
 - simulation, 299
 - sources, 273
 - time dependence, 285
 - variance reduction, 300–304
 - weights, 286
- sampling scattered directions, 279–283
- Path-length estimator, 288
- Path-length estimation, 274–277
 - cell travel distance, 274
 - concave vs convex cells, 275
 - effect of computer precision, 275
- PDF (*see* Probability distributions)
- PDF, linear in parameters, 191
 - separable, 188
- Phase space, 258
- Photon scattering, 283
- Poisson distribution, 325, 330–333
- Poisson's equation, (*see* Differential equations)
- Posterior distribution, 156
- Precision of tally, 90–93
- Primitive roots, 57
- Prior distribution, 155
- Probability concepts, 153
 - frequentists vs Bayesians, 153
- Probability density function, definition, 22

Probability distributions

- Bernoulli, 325
- beta, 159, 315, 314–316
 - moments, 314
 - sampling from, 315
- binomial, 326–329
 - radioactive decay, 328
 - sampling from, 328
- bivariate normal, 334
- Cauchy, 309, 319
 - moments, 320
 - sampling from, 320
- CDF definition, 23
- chi-squared, 321
- conditional, 32, 154
- continuous, 307–324
- covariance, population, 32
- discrete, 29, 324–333
- exponential, 24, 309–312
 - moments, 310
 - sampling from, 71, 311
- for scattering, general, 279–283
 - neutron, 284
 - photon, 283–284
- gamma, 165, 307, 312–314
 - moments, 313
 - sampling from, 313
- Gaussian, 24, 317
 - moments, 317
 - sampling from, 317
- geometric, 329
 - sampling from, 329
- joint, 31, 33, 80, 333–335
- likelihood, 156, 162
- logistic, 309
- lognormal, 319
 - moments, 319
 - sampling from, 319
- marginal, 155
- mean, population, 26, 28, 32
 - sample, 27
- multinomial, 334
- multiple variable, 33
- multivariate normal, 333
 - sampling from, 333
- negative binomial, 329
 - sampling from, 330
- normal approximation, 322
 - normal, 24, 37, 309, 317
 - moments, 317
 - sampling from, 318
 - Pareto, 323
 - sampling from, 324
 - PDF definition, 22
 - Poisson, 325, 330–333
 - approximation to binomial, 331
 - sampling from, 332
 - posterior, 156
 - prior, 156
 - proposal, 138
 - Rayleigh, 316
 - standard deviation
 - population, 26
 - sample, 29
 - standardized, 308
 - Student's t , 322
 - moments, 323
 - translated exponential, 311
 - sampling from, 312
 - uniform, 23, 308
 - sampling from, 71, 309
 - variance
 - population, 26, 28, 32
 - sample, 28
 - Weibull, 309, 316
 - moments, 316
 - sampling from, 316
- Proposal distribution, 144–148
- Pseudorandom numbers, 47
 - carry-with-add generators, 64
 - combining generators, 62
 - bit mixing, 62
 - L'Ecuyer, 62
 - Wickmann-Hill, 62
 - congruential generators, 48–63
 - combining generators, 62
 - desired properties, 51
 - optimum prime moduli, 59
 - shuffling output, 60
 - skipping ahead, 61
 - structure, 49
 - tests for, 52
 - effects on MC accuracy, 89
 - generation of, 47–65
 - inverse congruential generators, 65
 - lagged Fibonacci generators, 64

Pseudorandom numbers (*continued*)

- minimal standard, 57
 - coding, 57, 373–375
 - deficiencies, 57
- multiple recursive generators, 63
- nonlinear congruential generators, 65
- practical generators, 54–60
 - modulus = 2^α , 54
 - prime modulus, 56
- vs random numbers, 47

Q

Quadrature, (*see also* Monte Carlo)

- antithetic variates, 117–120
- by Markov chains, 142
- dimensionality reduction, 121
- examples, 10, 41, 100
- importance sampling, 101
- multidimensional, 87
- ordered variates, 127
- systematic sampling, 110
- volume partitioning, 121
- weighted Monte Carlo, 127

Quasirandom numbers, 48, 129

Quota sampling, (*see* Stratified sampling)

R

Radiation

- average travel distance, 245
- current vector, 241
- description, 236–241
- fluence, 240
- flux density, 241
- interactions, 242–251
- mean free path, 245
- partial current, 241
- particle density, 238
- reaction rate density, 249
- uncollided, attenuation, 244

Radioactive decay, 27

Random numbers, (*see* pseudorandom numbers)

Random walks, (*see also* Markov chains)

- continuous, 215, 218
- discrete, 200–205
- walking on circles, 220, 226, 228
- walking on spheres, 221, 223

Ratio of uniforms sampling method, 80

- Rayleigh, 6
 - distribution, 316
- Reaction density estimator, 289
- Reaction rate density, 249
- Rectangular distribution, (*see* Uniform distribution)
- Rejection sampling method, 74–76
- Relative error, 83
- Rotation matrices, 280–282
- Round-off errors, 92
- Russian roulette, 122–124

S

Sample statistics

- mean, 27
- standard deviation, 34
- variance, 32

Sampling random variables, 69–82, (*see also* Probability distributions)

- composition method, 76
- composition-rejection method, 79
- decomposition method, 78, 93
- from a nearly linear PDF, 79
- from joint distribution, 80
- general discussion, 81
- Gibb's sampler, 152
- inverse CDF method, 308
 - continuous variables, 70
 - discrete variables, 72
- isotropic scattering, 32
- linear transformation, 307
- Metropolis-Hasting algorithm, 137–142
- multidimensional, 148–152
- ratio of uniforms method, 80
- rejection method, 74–76
- table lookup, 73
- uniform over disk, 76

Scattering coefficient, (*see* Cross section)

Schwartz's inequality, 106, 116

Scoring, 82–88

- accumulators, 82
- accuracy and precision, 89–93
- estimating PDF, 85
- figure of merit, 83
- fluence at a point, 290–292
- fluence in a volume, 288–290
- fluence over a surface, 287

- Leakage estimator, 292
 - example, 293
- multidimensional integrals, 87
- next event estimator, 290–292
- particle transport, 286–293
- path-length estimator, 288
- random number generator errors, 89
- reaction density estimator, 289
- relative error, 83
- round-off error, 91–93
- statistical tests, 83–85
- successes-over-trials, 85
- surface current, 290
- systematic error, 89, 269
- use of weights, 86, 286, 302
- user errors, 90
- variance of the variance, 84
- Simulation, (*see* Monte Carlo)
- Sobol's quasirandom numbers, 129
- Solid angle, 236, 259
- Spectral radius, 204
- Splitting, 122–124, 302
- Standard deviation
 - population, 27
 - sample, 29
- Statistical analysis
 - Bayesian, example, 158
 - frequentists, example, 156
- Statistical error, 269
- Stratified sampling, 111–113
 - vs importance sampling, 113
- Strong law of large numbers, 339
- Student's *t* distribution, 322
- Successes-over-trials, 85
- Sums of random variables, 34
- Surface current estimator, 290
- Symbolic MC, 178–196
 - domain unknown, 194–196
 - PDF known, 181–185
 - PDF unknown, 185–194
- Systematic error, 89, 269
- Systematic sampling, 107–111
 - as importance sampling, 110
- T**
- Table lookup sampling, 73
- Tally, (*see* Scoring)
- Tests for randomness, 53
- Transition function, 135
 - limiting value, 135
 - reversibility, 139
- Transition matrix, 134
 - limiting value, 134
- Transition probability, 134, 139
- Transport codes
 - COG, 347
 - EGSnrc, 349–351
 - GEANT4, 351–353
 - MCNP, 356–359
 - MCSHAPE, 353–356
 - PENELOPE, 359–361
 - SCALE, 361–364
 - SRIM, 364
 - TRIPOLI, 365–367
- Transport equation
 - adjoint
 - derivation, 262–263
 - utility, 263
 - balance relation, 251
 - derivation, 251–253
 - in phase space, 258
 - integral form, 255–260
 - alternate forms, 260
 - Monte Carlo, 296–300
 - integro-differential form, 255
 - Monte Carlo, variance reduction, 300–304
 - solution by MC, 300–304
- Transport in absorbing media, 277
- Transport theory, (*see* Particle transport)
- U**
- Ulam, 7
- Uniform distribution, 23, 308–310
 - sampling from, 71
- Unknown PDF, approximate by histogram, 185–188
- V**
- Variance of variance, 84
- Variance reduction, 97–129
 - biased estimators, 125
 - combining methods, 124
 - correlated sampling, 113–120
 - dimensionality reduction, 121
 - exponential transformation, 303
 - general methods, 98
 - implicit absorption, 303

Variance reduction (*continued*)

- importance sampling, 101–106, 300–302
- integration volume partitioning, 121
- interaction forcing, 303
- nonanalog methods, 300–304
- problem transformation, 100
- Russian roulette, 122–124, 302
- splitting, 122–124, 302
- stratified sampling, 111–113
- systematic sampling, 107–111
- truncation methods, 302
- weighted MC integration, 126–129
- weighted uniform sampling, 126
- with antithetic variates, 117–120
- with quasi-random numbers, 129

Variance

- population, 25, 29

sample, 28

sum of random variables, 34

von Neumann, 7

W

Weak law of large numbers, 337–339

Weibull distribution, 309, 316

Weighted MC integration, 127

example, 128

Weighted uniform sampling, 126

Weights, importance sampling, 102

Russian roulette, 123

scoring, 87, 286

splitting, 123

Wickman-Hill pseudorandom number

generator, 62