杨耀臣 著

蒙特卡罗方法与人口仿真学

中国科学技术大学出版社

蒙特卡罗方法与人口仿真学

杨耀臣　著

中国科学技术大学出版社
1999・合肥

图书在版编目(CIP)数据

蒙特卡罗方法与人口仿真学/杨耀臣著. —合肥:中国科学技术大学出版社,1999.7

ISBN 7-312-01118-7

Ⅰ. 蒙…

Ⅱ. 杨…

Ⅲ. ①蒙特卡罗法 ②人口—仿真—研究

Ⅳ. C92

中国版本图书馆 CIP 数据核字(1999)第 16527 号

中国科学技术大学出版社出版发行

(安徽省合肥市金寨路 96 号,230026)

肥西新华印刷厂印刷

全国新华书店经销

开本:850×1168/32　印张:7　字数:182 千

1999 年 7 月第 1 版　1999 年 7 月第 1 次印刷

印数:1—1000 册

定价:10.00 元

内容简介

本书首先介绍了蒙特卡罗方法的基本知识、一般原则及其在粒子输运、解微分方程和统计力学中的应用，然后介绍了用蒙特卡罗方法进行人口仿真学研究的实现过程及其有关人口战略的研究案例.

本书适合科研和教学单位、工商企业及政府机关中具有大学以上文化程度并略通概率论的有关人员学习参考.

前言

蒙特卡罗方法具有其他任何计算方法所无法比拟的高精确度，然而它需要花费高昂的计算机CPU费用，因此在本世纪六七十年代，蒙特卡罗方法的这一致命缺点严重限制了其运用的深度和广度.今天，随着计算机技术的迅猛发展，一台普通的586微机就可以轻而易举地获得2～3GB硬盘存储量、32～64MB内存以及每秒百万次运行速度的高指标，这便为蒙特卡罗方法——一种利用重复的统计实验来求解物理、数学问题的方法焕发新的生机创造了良好的物质条件.

过去，蒙特卡罗方法被尊称为计算方法中的“贵族”，在核工业研究设计中，只是在最关键的设计阶段才搬出这位“贵族老爷”充当检查官.因此在那时候，蒙特卡罗方法仅被局限在自然科学中的粒子物理领域里应用.今天，由于计算机技术日新月异的变化，情况大大改观了，蒙特卡罗方法已远远超出了单纯的粒子物理领域，而广泛渗透到自然科学中诸如机械、电子、力学、火灾科学乃至商业预测等领域，因此，希望了解、学习和使用蒙特卡罗方法的人也迅速增加.作者出版本书的目的之一便是深入浅出地介绍一点蒙特卡罗方法的经典知识，使希望了解这一方法的读者获得入门知识.

适用于蒙特卡罗方法的问题一定是满足统计规律的事物，而许多发生在人文与社会科学领域中的问题也都具有这种特性.如当今世界面临的四大社会问题：灾难性的世界人口增长、人类资源的无度开发与浪费、日趋严重的环境污染和破坏以及呈指数增长的信息爆炸，都可以直接或间接地分解出一系列符合特定统计规律的事物，求助于蒙特卡罗方法予以解决.

在安徽省原省委书记卢荣景同志的关心与支持下，作者首次将蒙特卡罗方法应用于人口学，完成了《用蒙特卡罗方法研究人口战略》课题的研究，并获得安徽省高等学校人文、社会科学研究优秀成果二等奖．为了推进自然科学与社会科学的结合，特别是为了促进蒙特卡罗方法在我国人口学的高层次研究中发挥更好的作用，作者编写了本书，并在书中将作者研制的“人口战略的蒙特卡罗研究软件包(PSMC)”进行了详细的剖析，作为引玉之砖奉献给读者．

作为用蒙特卡罗方法研究人口战略的案例，本书的最后一章介绍了6篇在国内学术刊物或国际学术会议上发表过的论文或向政府决策机构提出的政策建议．

本书适合科研和教学单位、工商企业及政府机关中具有大学以上文化程度并略通概率论的读者阅读．

杨耀臣

1998年8月于中国科学技术大学

目　次

第1章 绪 论

1.1 蒙特卡罗方法简介

蒙特卡罗方法是一种利用重复的统计实验来求解物理和数学问题的方法，这类问题可以直接或间接地用一个随机过程来描述，蒙特卡罗方法就是设法去模拟这个过程.这种模拟过程有时候也叫做仿真过程，通过模拟或仿真过程，找出我们希望求出的某些结果或观察我们感兴趣的某种现象.

一个事件在一个随机过程中的出现是无法确知的，但是这个事件可能出现的机会，或者用学术一点的词汇，叫做出现的“概率”，我们是知道的.换句话说，若我们能用一个随机过程去描述一个问题，我们必须预先知道此事件在这个过程中出现的概率的分布形式.例如，在一个投掷骰子的游戏中，我们无法确知在某次投掷时哪个面会朝上，但是知道每个面有六分之一的机会出现.能找出事件在一个随机过程中发生的概率分布，是用蒙特卡罗方法求解一个问题的基本要求.

要模拟或仿真一个随机过程，我们必须要有基本的工具.好比玩掷骰子游戏时，骰子便是工具；玩轮盘赌游戏时，轮盘是工具；玩掷硬币游戏时，则硬币是工具.在用蒙特卡罗方法求解一个问题时，这个基本工具就是一组数字，称为随机数.这组数字描绘着问题中某种概率的分布形式，但是在确定的某一次特殊的实验中，我们都无法预测到底哪一个数字会出现.有了这个随机数工具，即有了描绘问题的概率分布，我们便可以从头到尾模拟问题的整个过程.

研究一个问题必然有其最终的目的，比如，在西方允许博彩业存在，赌客去赌的最终目的是赢钱，而买彩票的目的是中奖. 模拟一种随机的过程，往往是希望观察某种现象，而不像在摸彩时以一次的结果定乾坤. 而用蒙特卡罗方法模拟事件的过程，要经过多次重复实验，以求得平均的数据. 就好像西方国家的赌客，在赌博过程中，往往是在赌了许多次后离开赌场时，他们才能真正地论断其输赢. 所以，蒙特卡罗方法又称为重复性的统计实验方法.

既然蒙特卡罗方法是模拟一个随机的过程，以求得期望的解答，因此蒙特卡罗方法显然可以用来求解机遇性的问题或无规则运动的问题. 但是在许多情况下，蒙特卡罗方法也可以用来求解肯定性的问题. 对于这类问题，我们总是可以找到一个与之相对应的随机过程，从而将一个肯定性的问题转化为概率性的问题.

现在让我们举几个概率性的问题，来说明蒙特卡罗方法在这些方面应用的好处.

比如说，我们想知道中国人口增长和分布的趋势，根据现有的婚姻、生育、生存、死亡等统计资料，我们找出相对应的概率分布，称其为分布函数. 对一个个人口样本，用一组随机数来模拟他（她）的发展，这个人口个体的集合就相当于一组人口样本. 分析这组样本，便可以找出某些问题的解答. 这种通过计算机来模拟的蒙特卡罗方法比取实际的人口样本进行调查的方法要方便得多. 首先，我们不需要等好几十年来观察实际的发生与发展过程；其次，我们可以改变某一分布概率中的一些参数（相当于某些人口政策），用以研究其对人口分布和增长的影响，这种做法在实际观察中是绝对做不到的. 本书的第 7 章和第 8 章分别介绍了用蒙特卡罗方法进行人口仿真的实现过程及其用于人口战略上的几个案例.

中子在介质中的运动是随机性的，因此我们可以用蒙特卡罗方法来模拟它们的运动. 这样，我们就不必建一个真正的反应堆来讨论和研究装置的物理现象，省却了建堆带来的许多问题. 当然，中子的运动也可以用一个肯定的方程式（如玻尔兹曼方程）来描

述，在许多情况下，用解析或数值的方法可以求出中子的行为. 但是，这类方法往往因受到几何的限制而不得不作许多简化，从而导致不能完全反映事物的本来面目. 用蒙特卡罗方法来求解粒子输运问题是应用蒙特卡罗方法最透彻、最直观的例子之一. 本书第 4 章将详细讨论这个问题.

蒙特卡罗方法也经常用来求解一些在数学上很难写出恰当的方程式的随机过程的问题，例如电话交换系统的设计. 在某一段时间内对电话的使用和使用时间的长短是由电话用户中的个体决定的，交换系统的设计便必须以使用次数和使用时间长短的分布为依据. 在本世纪 40 年代前后，数学的发展尚难以解决这个问题. 因此在 1949 年，英国的邮政总局建立了一个“类比计算机”来模拟电话事务的不规则运动，这是一个具有历史意义的蒙特卡罗方法的试验例子.

对于一些肯定性的问题，从理论上我们可以找出解答的“性质”，但是在实际求解时并不一定求得出来. 但是，如果我们能在基础结构上找出相对应的随机过程，那么这个问题的解便可以由蒙特卡罗方法求得. 最明显的例子是解扩散方程，这类问题往往是难以精确求解的. 但是，这个问题和“醉汉无规则行走”问题是相通的，“醉汉”问题是一个随机性的问题，是可以用蒙特卡罗方法求解的，所以扩散问题也可以用蒙特卡罗方法求解. 蒙特卡罗方法曾经被用来解联立的线性方程式，求方程式的特征值，求逆矩阵等，所有这些问题都是通过寻找一种对应的随机过程而求得解的.

根据上面的叙述，看起来好像蒙特卡罗方法能求解一切问题，不论是概率性的还是肯定性的. 但是，我们必须了解蒙特卡罗方法的“有效性”. 如果一个问题的解可以用其他方法较为迅速地求得的话，那么蒙特卡罗方法的相对有效性便降低了. 有些时候，其他方法虽不能求得精确的值，但能求到近似值，如果这个近似值足够我们使用，而在资金、时间上又经济得多，蒙特卡罗方法的相对有效性也就降低了许多. 譬如在核电站设计中，关于反应堆的临界计

算就是先通过少群中子参数预算，然后利用多群参数复算，最后用蒙特卡罗方法验算.

我们还必须注意到一点，那就是蒙特卡罗方法是一种统计方法，它的结果不是确定到一点，其基本原因当然是我们取用的样本只是总体中有限的一部分.然而，虽然蒙特卡罗方法得到的结果不是确定的，但如果我们能减小误差范围，那么结果的可靠性便增加了.一般而言，误差的减小和所取样本总数增加的平方根成反比，要误差减小到$\frac{1}{10}$，必须增加所取样本总数 100 倍，这样做往往是不经济的.然而，蒙特卡罗方法是一种数学方法，在许多情况下，我们可以采用多种抽样技巧，使抽样得到相应的改善，如此便使我们既能得到预期的结果，又能使误差减小许多.因此，为了提高蒙特卡罗方法的使用效率，我们必须对这种方法的数学基础与本质有较深刻的了解和认识.

在第 2 章里，我们将讨论随机数的意义和产生方法，特别给读者详细介绍作者在实际工作中使用十分有效的在计算机中建立随机数发生器的方法.在第 3 章里，我们将用一次积分的例子说明蒙特卡罗方法的一般原则.我们选用一次积分而不选用一些明显而又可以模拟的随机过程，是因为积分式在蒙特卡罗方法中扮演着一个非常重要的角色.因此，许多蒙特卡罗方法在用积分式说明时在意义上比较清楚，在数学逻辑上也比较严密.在第 4 章里，我们讨论一般粒子输运的过程，读者可以从中了解如何用蒙特卡罗方法去直接或间接地模拟，如何求解问题的预期结果等.第 5 章讨论蒙特卡罗方法在求解偏微分方程中的应用，我们用热扩散方程来说明如何了解一个数学问题背后所带有的概率性质.在第 6 章中，简略讨论蒙特卡罗方法在统计力学上的应用.第 7 章是本书的重点内容，它为高层次人口学研究工作者提供了用蒙特卡罗方法进行人口仿真学研究的具体步骤.据作者所知，如此全面和完整地把蒙特卡罗方法移植到人口学研究中，并得到了完全符合逻辑的结

果，在国内外尚属首次. 作者把自己研制的大型计算机软件“人口战略的蒙特卡罗研究软件包(PSMC)”完全打开，作为引玉之砖，供有志于用这种方法做高层次人口研究的学者参考. 第 8 章的内容是作者用上述软件进行人口仿真学研究的成果，这些文字都曾在《中国人口科学》等学术刊物或国际学术会议上发表过，可能从某个角度反映了作者在我国人口管理与决策上的一管之见.

1.2 概率论和统计学中的一些基础知识

1. 概率分布

假设 $x\in(a,b)$，$p(x)$是 x 在(a,b)上的概率分布，则

(1) $p(x)\geqslant 0,\forall\ x\in(a,b)$；

(2) $\int_a^b p(x)\mathrm{d}x = 1$ 或 $\sum_i p(x_i) = 1$.

2. 积累概率分布

假设 $x,X\in(a,b)$，$p(x)$为 x 在(a,b)上的概率分布，则

$$P(X) = \int_a^X p(x)\mathrm{d}x$$

为 $p(x)$的积累概率分布函数. 特别地，有

$$P(a) = 0,\quad P(b) = 1.$$

而且，当 $X_1\geqslant X_2$ 时有

$$P(X_1)\geqslant P(X_2).$$

3. 期望值或概率平均值

假设 $x\in(a,b)$，$p(x)$为 x 在(a,b)上的概率分布，则

$$\langle f\rangle = \int_a^b f(x)p(x)\mathrm{d}x$$

或

$$\langle f\rangle = \sum_i f(x_i)p(x_i)$$

为 $f(x)$ 的期望值.

4. $f(x)$ 的方差

假设 $x \in (a,b)$，$p(x)$ 为 x 在 (a,b) 上的概率分布，则 $f(x)$ 的方差 σ^2 为

$$\sigma^2 = \mathrm{Var}\langle f \rangle = \int_a^b [f(x) - \langle f \rangle]^2 p(x) \mathrm{d}x$$
$$= \langle f^2 \rangle - \langle f \rangle^2$$

或

$$\sigma^2 = \sum_i [f(x_i) - \langle f \rangle]^2 p(x_i).$$

5. $f(x)$，$g(y)$ 的合方差

假设 $x \in (a,b)$，$y \in (c,d)$，且 $p(x)$ 为 x 在 (a,b) 上的概率分布，$p(y)$ 为 y 在 (c,d) 上的概率分布，$p(x,y)$ 为 x，y 同时发生的概率，则

$$\int_a^b \int_c^d p(x,y) \mathrm{d}y \mathrm{d}x = 1,$$

$$\int_a^b p(x,y) \mathrm{d}x = p(y),$$

$$\int_c^d p(x,y) \mathrm{d}y = p(x).$$

若 x，y 为互相独立的事件，则

$$p(x,y) = p(x) \cdot p(y).$$

$f(x)$，$g(y)$ 的合方差定义为

$$\mathrm{Cov}(f,g) = \int_a^b \int_c^d [f(x) - \langle f \rangle][g(y) - \langle g \rangle] p(x,y) \mathrm{d}x \mathrm{d}y$$

或

$$\mathrm{Cov}(f,g) = \sum_i \sum_j [f(x_i) - \langle f \rangle][g(y_j) - \langle g \rangle] p(x_i,y_j).$$

相关系数 ρ 定义为

$$\rho = \frac{\mathrm{Cov}(f,g)}{\mathrm{Var}(f) \cdot \mathrm{Var}(g)}.$$