

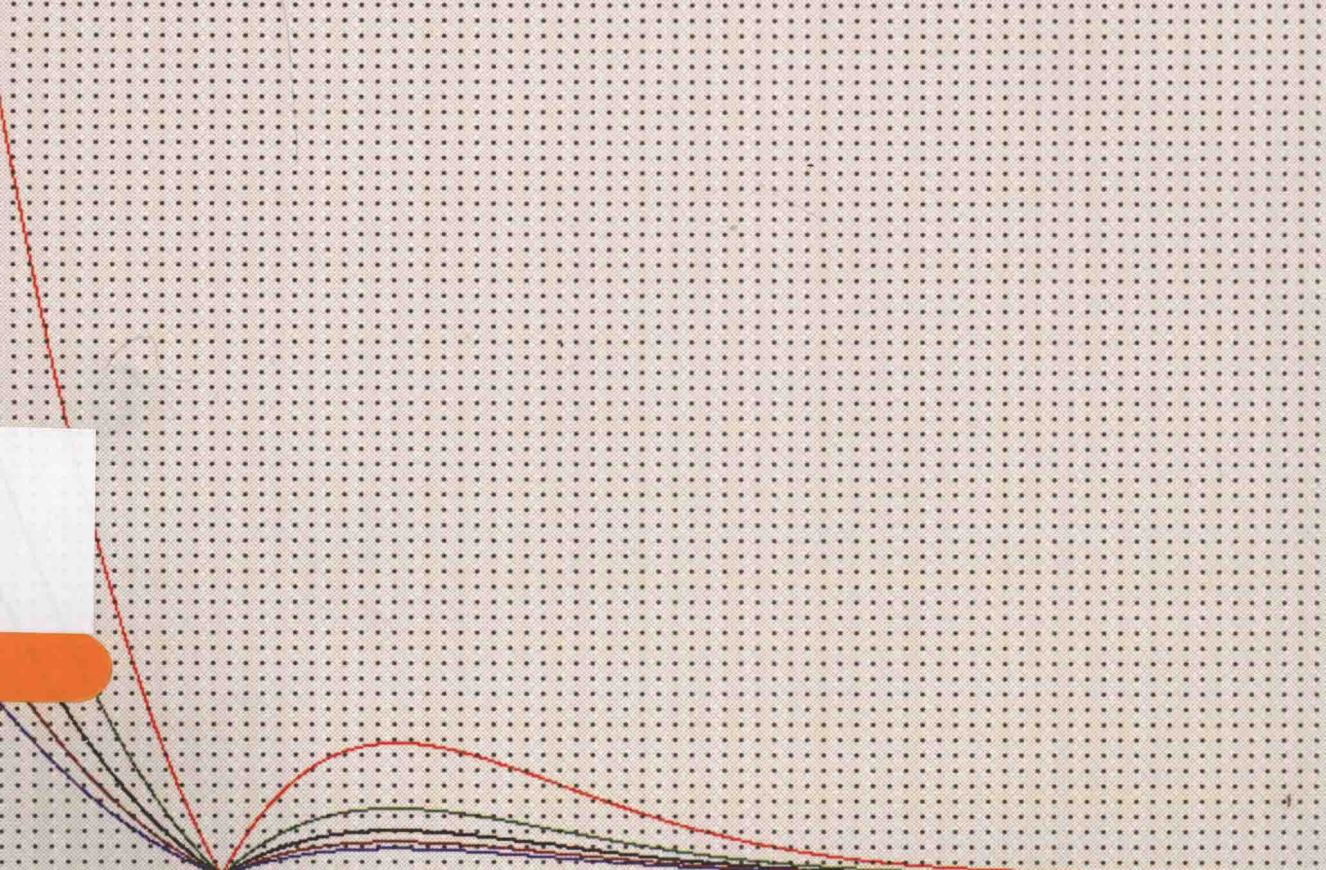
国家社会科学基金项目成果

# 蒙特卡罗模拟与概率统计

/ 基于SAS研究 /

Monte Carlo Simulation for Probability  
and Mathematical Statistics  
based on SAS

江海峰 著



中国科学技术大学出版社

· 国家社会科学基金项目成果 ·

# 蒙特卡罗模拟与概率统计

/ 基于SAS研究 /

Monte Carlo Simulation for Probability  
and Mathematical Statistics  
based on SAS

江海峰 著



中国科学技术大学出版社

## 内 容 简 介

本书借助 SAS 软件强大的统计分析功能,利用蒙特卡罗模拟思想,以概率论与数理统计内容为研究对象,从实验的角度再现相关定理的证明。本书共十二章,分为三篇:第一篇简要介绍蒙特卡罗模拟与 SAS 软件基础,共两章;第二篇着重介绍蒙特卡罗模拟在概率论与分布中的应用,共四章;第三篇主要介绍蒙特卡罗模拟在数理统计中的应用,共六章。

本书可为蒙特卡罗模拟思想的学习、概率论与数理统计的课堂与实验教学、SAS 软件的了解与深入研究提供参考,也有助于提升使用蒙特卡罗模拟技术和 SAS 软件解决实际问题的能力。

### 图书在版编目(CIP)数据

蒙特卡罗模拟与概率统计:基于 SAS 研究/江海峰著. —合肥:中国科学技术大学出版社, 2015. 4

ISBN 978-7-312-03488-6

I. 蒙… II. 江… III. 蒙特卡罗法—应用—统计分析—应用软件 IV. C819

中国版本图书馆 CIP 数据核字(2014)第 248640 号

出版 中国科学技术大学出版社  
安徽省合肥市金寨路 96 号,230026  
<http://press.ustc.edu.cn>  
印刷 合肥学苑印务有限公司  
发行 中国科学技术大学出版社  
经销 全国新华书店  
开本 787 mm×1092 mm 1/16  
印张 30.75  
字数 768 千  
版次 2015 年 4 月第 1 版  
印次 2015 年 4 月第 1 次印刷  
定价 56.00 元

# 前 言

蒙特卡罗模拟是一种基于“随机数”的计算方法,通过设定随机过程(数据生成系统),反复生成时间序列,并计算参数估计量和统计量,进而研究其分布特征的方法。这一方法源于美国第一次世界大战研制原子弹的“曼哈顿计划”。该计划的主持人之一数学家冯·诺伊曼用著名的赌城——摩纳哥的蒙特卡罗(Monte Carlo)来命名这种方法。由于该方法主要被用来研究随机现象,而概率论与数理统计正是研究随机现象及其背后隐藏规律的科学,因此从这个角度来说,蒙特卡罗模拟在概率论与数理统计中有着广泛的应用前景。早在17世纪,人们就知道用事件发生的“频率”来估计事件的“概率”;19世纪人们用投针试验的方法来估计圆周率 $\pi$ 。然而这种人工实验无疑受到实际条件的限制,直到20世纪40年代电子计算机的发明,特别是20世纪末以来高速电子计算机的出现,使得用数学方法在计算机上大量、快速地模拟这样的试验成为可能。

目前,概率论与数理统计作为一门数学基础课程,在高校中普遍开设,所使用的教材基本上遵循着理论介绍、例题讲解和习题练习这一传统模式,虽然不少教材中带有实验教学内容,但是那也不过是使用一些软件(例如Excel和Matlab)的计算功能来完成数据的运算,或者说在一些已有公式的指导下发挥软件的计算功能而已,而对如何用实验方法来演示这些公式、定理等则鲜有涉及,特别是高等教育教学改革工程实施以来,这门课程的教学出现了轻理论重应用的趋势,一些定理和公式的证明在不少教材中已经被淡化,这给学生的学习带来了一定的困难,作者在教学中也遇到了这样的问题,并一直思考以找到破解这个困境的方法。解决这个问题必须具备三个条件:首先要对蒙特卡罗模拟方法有一定程度的了解,其次是必须深刻理解概率论与数理统计的知识框架及其内在的体系,最后必须熟练掌握一种软件并能够将所要解决的问题转换为软件语言来贯彻蒙特卡罗模拟思想。幸好作者一直从事SAS软件的教学研究,从而为解决这个问题奠定了坚实的基础,并结合对概率论与数理统计、SAS软件教学总结以及蒙特卡罗模拟的理



解,撰写了这本书,其目的就是借助 SAS 软件用实验手段再现定理公式的证明以及习题答案的求解,同时对未知研究领域做探索性研究,给概率论与数理统计教学提供一个真正意义上的实验平台。

本书在撰写过程中坚持以下原则:

(1) 在研究内容上,本书从概率论开始到随机变量的分布、抽样分布,最后到参数估计、假设检验和回归分析等,紧扣概率论与数理统计的知识框架。

(2) 在篇章安排上,首先介绍蒙特卡罗模拟的基础知识,然后介绍蒙特卡罗模拟的 SAS 软件基础,最后再介绍如何利用蒙特卡罗模拟来解决概率论与数理统计中的问题,这样的安排使那些没有相关基础知识的读者也能使用本书进行应用研究。

(3) 在程序的编写与解释上,本书的程序按照所要解决的问题过程进行编写,每个小部分都分割开,这使得读者不但能了解所要解决问题的步骤,而且也知道使用什么对应的程序来完成。在程序的解释上,除对每个程序做整体阐述以外,还在程序中每个关键语句后面都备有注释,这样的布局使读者更容易读懂程序。

(4) 在数据的处理上,由于蒙特卡罗模拟会产生大量的模拟数据,因此不可能保存所有的数据;另外,本书程序在处理随机数生成时,一般将初始种子值设置为 SAS 系统时间,因此不同运行时刻会得到不同的实验结果,这样处理可以使读者得到自己的模拟结论,并与本书得到的结论进行相互验证,从而更好地理解蒙特卡罗模拟思想的精髓。

本书是在实践教学基础上产生的,从作者在实际教学中的应用效果来看还是令人满意的,因此本书首要的应用是配合概率论与数理统计的教学,特别是为该课程的实验教学提供一个平台;其次,本书使用了大量 SAS 软件程序,对那些想学习 SAS 软件的相关人员来说,不失为一本好参考资料,因为本书是介绍如何编写程序来解决问题的,是问题导向型,而不是枯燥的语言语法介绍;另外,对那些想了解蒙特卡罗模拟知识的读者来说,本书也具有一定的参考价值。

本书共分三篇,第一篇为基础篇,包括第一章和第二章,主要介绍蒙特卡罗模拟和 SAS 软件基础,读者尤其要掌握 SAS 软件基础;第二篇为蒙特卡罗模拟在概率论与分布中的应用,包括第三章、第四章、第五章和第六章,主要介绍蒙特卡罗模拟求解随机事件的概率与分布之间的关系、随机变量的期望与方差、切比雪夫不等式、大数定律与中心极限定理等内容的研究;第三篇为蒙特卡罗模拟在数理

统计中的应用,包括第七章、第八章、第九章、第十章、第十一章和第十二章,主要介绍蒙特卡罗模拟在抽样分布、区间估计、假设检验、回归分析、方差分析以及 Bootstrap 中的应用。

本书部分内容为安徽工业大学青年教师基金“蒙特卡罗模拟在概率统计中的应用研究(QS200808)”的成果,部分内容为国家社会科学基金“基于 Bootstrap 方法下单位根检验研究(13BJY011)”的阶段性成果。在编写过程中,得到了我的博士生导师和同门以及安徽工业大学原经济学院有关领导和商务实践中心老师们的大力支持,作者在此一并表示感谢!

由于本人水平有限,书中难免有错误和不妥之处,某些程序可能还可以进一步简化等,恳请同行学者和专家以及广大读者不吝赐教,批评指正。

**编 者**

2014年10月

# 目 录

前言 .....	( i )
----------	-------

## 第一篇 蒙特卡罗模拟与 SAS 软件基础

第一章 随机数的产生与检验 .....	( 3 )
第一节 蒙特卡罗模拟方法发展简介 .....	( 3 )
第二节 均匀分布随机数的产生与检验 .....	( 4 )
一、均匀分布随机数的产生与检验 .....	( 4 )
二、均匀分布随机数相关检验 .....	( 6 )
第三节 两种常用产生随机数的方法 .....	( 17 )
一、逆转方法产生特定分布随机数 .....	( 17 )
二、拒绝-接受法 .....	( 30 )
第四节 几种特殊分布随机数的产生 .....	( 33 )
一、具有特定偏度和峰度的非正态分布随机数的产生 .....	( 33 )
二、具有特定相关程度的多元正态分布随机数的产生 .....	( 40 )
三、具有特定相关程度和偏斜程度的多元非正态分布随机数的产生 .....	( 45 )
第二章 蒙特卡罗模拟的 SAS 基础 .....	( 52 )
第一节 SAS 中常用的随机数发生器 .....	( 52 )
一、ranbin 与二项分布随机数发生器 .....	( 52 )
二、rancau 与柯西分布随机数发生器 .....	( 53 )
三、ranexp 与指数分布随机数发生器 .....	( 53 )
四、rangam 与伽玛分布随机数发生器 .....	( 53 )
五、rannor 与标准正态分布随机数发生器 .....	( 53 )
六、ranpoi 与泊松分布随机数发生器 .....	( 54 )
七、rantri 与三角分布随机数发生器 .....	( 54 )
八、ranuni 与均匀分布随机数发生器 .....	( 54 )
九、rand 与特定分布的随机数发生器 .....	( 55 )
第二节 SAS 数据库与数据集 .....	( 57 )
一、SAS 数据库 .....	( 57 )
二、SAS 数据集 .....	( 58 )



第三节 SAS 编程的语法基础 .....	(60)
一、SAS 语法基础 .....	(61)
二、data 步中的一些常用语句 .....	(63)
第四节 几个常用 SAS 过程介绍 .....	(68)
一、print 与显示过程 .....	(68)
二、sort 与排序过程 .....	(69)
三、means 与均值过程 .....	(70)
四、univariate 与单变量过程 .....	(72)
五、freq 与频数过程 .....	(73)
六、reg 与回归过程 .....	(73)
七、append 与追加记录过程 .....	(74)
八、gplot 与折线(散点)图过程 .....	(74)
九、printto 与输出控制过程 .....	(77)
十、ttest 与双样本检验过程 .....	(78)
十一、glm 与方差分析过程 .....	(78)
第五节 SAS 宏与 IML 模块 .....	(79)
一、SAS 宏 .....	(79)
二、IML 基本知识介绍 .....	(82)
<b>第二篇 蒙特卡罗模拟在概率论与分布中的应用</b>	
第三章 随机事件发生的概率与 MCS 研究 .....	(89)
第一节 经典实验与 MCS 求解 .....	(89)
一、MCS 分析的理论基础 .....	(89)
二、蒲丰的针及其理论求解 .....	(90)
三、MCS 模拟与估计结果 .....	(91)
四、抛硬币实验与 MCS 求解 .....	(93)
第二节 随机取数模型与 MCS 求解 .....	(94)
第三节 其他类型概率求解的 MCS 分析 .....	(107)
第四章 随机变量分布特点与 MCS 研究 .....	(116)
第一节 分布与其参数关系的 MCS 分析 .....	(116)
一、二项分布的形态与参数的关系 .....	(116)
二、泊松分布的形态与参数的关系 .....	(129)
三、卡方分布形状与自由度的关系 .....	(132)
四、F 分布形状与自由度的关系 .....	(136)
第二节 分布之间替代关系的分析 .....	(138)
一、二项分布、超几何分布与泊松分布 .....	(138)



二、标准正态分布与 $t$ 分布 .....	(140)
三、卡方分布与 $t$ 分布 .....	(144)
四、卡方分布与 $F$ 分布 .....	(148)
五、密度函数与分布函数替代的差异 .....	(152)
<b>第五章 随机变量的数字特征与 MCS 研究 .....</b>	<b>(153)</b>
<b>第一节 常见分布的期望与方差介绍 .....</b>	<b>(153)</b>
一、期望与方差的定义 .....	(153)
二、常见分布的期望与方差 .....	(153)
<b>第二节 常见分布期望与方差的 MCS 分析 .....</b>	<b>(156)</b>
<b>第三节 切比雪夫不等式与 MCS 分析 .....</b>	<b>(161)</b>
<b>第四节 大数定律与 MCS 分析 .....</b>	<b>(165)</b>
一、切比雪夫大数定律与 MCS .....	(165)
二、贝努里大数定律与 MCS .....	(169)
三、辛钦大数定律与 MCS .....	(170)
四、大数定律的一个应用 .....	(174)
<b>第六章 中心极限定理与 MCS 研究 .....</b>	<b>(176)</b>
<b>第一节 拉普拉斯中心极限定理的 MCS 分析 .....</b>	<b>(176)</b>
一、拉普拉斯中心极限定理 .....	(176)
二、拉普拉斯中心极限定理的 MCS 证明 .....	(176)
<b>第二节 林德伯格-勒维中心极限定理的 MCS 分析 .....</b>	<b>(185)</b>
一、林德伯格-勒维中心极限定理 .....	(185)
二、林德伯格-勒维中心极限定理的 MCS 证明 .....	(185)
<b>第三节 李雅普诺夫中心极限定理的 MCS 分析 .....</b>	<b>(196)</b>
一、李雅普诺夫中心极限定理 .....	(196)
二、李雅普诺夫中心极限定理的 MCS 证明 .....	(196)
<b>第三篇 蒙特卡罗模拟在数理统计中的应用</b>	
<b>第七章 抽样分布检验与 MCS 研究 .....</b>	<b>(205)</b>
<b>第一节 三大分布的 MCS 检验 .....</b>	<b>(205)</b>
一、正态分布的线性组合及其 MCS 检验 .....	(205)
二、卡方分布的形成与 MCS 检验 .....	(209)
三、 $t$ 分布的形成与 MCS 检验 .....	(212)
四、 $F$ 分布的形成与 MCS 检验 .....	(214)
<b>第二节 单正态总体下的抽样分布与 MCS 检验 .....</b>	<b>(217)</b>
一、样本方差的卡方分布与 MCS 检验 .....	(217)
二、样本均值 $t$ 分布与 MCS 检验 .....	(220)



三、样本均值与样本方差独立的 MCS 检验 .....	(222)
第三节 双正态总体下的抽样分布与 MCS 检验 .....	(225)
一、双样本 $t$ 分布与 MCS 证明 .....	(225)
二、双样本 $F$ 分布与 MCS 证明 .....	(228)
三、双样本均值与样本方差之间独立性与 MCS 检验 .....	(232)
四、关于检验通过率的一个说明 .....	(236)
第八章 区间估计与 MCS 研究 .....	(237)
第一节 区间估计基本知识回顾 .....	(237)
一、单正态总体期望 $\mu$ 的区间估计 .....	(237)
二、单正态总体方差 $\sigma^2$ 的区间估计 .....	(238)
三、双总体期望差 $\mu_1 - \mu_2$ 的区间估计 .....	(238)
四、双总体方差比 $\sigma_1^2 / \sigma_2^2$ 的区间估计 .....	(239)
第二节 单正态总体区间估计的 MCS 分析 .....	(240)
一、关于期望 $\mu$ 区间估计的 MCS 研究 .....	(240)
二、关于方差区间估计的 MCS 研究 .....	(243)
第三节 双正态总体区间估计的 MCS 分析 .....	(247)
一、关于期望差 $\mu_1 - \mu_2$ 的 MCS 分析 .....	(247)
二、关于期望差 $\sigma_1^2 / \sigma_2^2$ 的 MCS 分析 .....	(252)
第四节 正态总体单侧区间估计的 MCS 分析 .....	(259)
一、双侧区间估计公式与单侧区间估计的关系 .....	(259)
二、单总体期望的单侧区间估计与 MCS 分析 .....	(259)
三、双总体期望差的单侧区间估计与 MCS 分析 .....	(262)
第五节 非正态总体区间估计的 MCS 分析 .....	(266)
一、三种分布参数区间估计的推导 .....	(266)
二、三种分布参数区间估计的 MCS 分析 .....	(267)
第九章 假设检验与 MCS 研究 .....	(270)
第一节 单正态总体的假设检验与 MCS 分析 .....	(270)
一、MCS 分析的依据 .....	(270)
二、单正态总体期望 $\mu$ 的假设检验 .....	(271)
三、单正态总体方差 $\sigma^2$ 的假设检验 .....	(279)
第二节 双正态总体的假设检验与 MCS 分析 .....	(285)
一、双正态总体两个期望相等的假设检验 .....	(285)
二、双正态总体两个方差相等的假设检验 .....	(295)
第三节 单正态总体假设检验中两类错误及其 MCS 研究 .....	(308)
一、假设检验中的两类错误及其概率 .....	(308)
二、MCS 分析的依据 .....	(308)



三、单总体期望的检验研究 .....	(308)
四、单总体方差的检验研究 .....	(313)
第四节 双正态总体假设检验中两类错误及其 MCS 研究 .....	(317)
一、双总体期望相等的检验研究 .....	(317)
二、双总体方差相等的检验研究 .....	(321)
第五节 假设检验中两类错误与影响因素分析 .....	(329)
一、犯第二类错误概率与样本容量关系的研究 .....	(329)
二、犯两类错误概率之间关系的研究 .....	(329)
第六节 双总体 $t$ 检验的稳健性与 MCS 分析 .....	(330)
一、方差不相等与 MCS 检验 .....	(330)
二、非正态分布与 MCS 检验 .....	(337)
三、分布非独立性与 MCS 检验 .....	(344)
四、多种条件违背与 MCS 检验 .....	(348)
第十章 线性回归模型与 MCS 研究 .....	(353)
第一节 相关系数检验与 MCS 研究 .....	(353)
一、相关系数及检验概述 .....	(353)
二、零相关系数 $t$ 检验的 MCS 分析 .....	(355)
三、零相关系数临界值的 MCS 分析 .....	(358)
四、其他样本容量非正态分布零相关系数临界值的 MCS 分析 .....	(362)
五、非零相关系数正态检验的 MCS 分析 .....	(364)
六、非零相关系数检验临界值存在性的 MCS 分析 .....	(369)
第二节 经典假设下参数估计量性质与 MCS 研究 .....	(379)
一、参数估计量无偏性和一致性的 MCS 分析 .....	(379)
二、参数估计量正态性的 MCS 分析 .....	(381)
第三节 经典假设下的假设检验与 MCS 研究 .....	(383)
第四节 非经典假设下的假设检验效率与 MCS 研究 .....	(386)
一、扰动项非正态分布与 MCS 分析 .....	(386)
二、扰动项一阶自相关与 MCS 分析 .....	(389)
三、扰动项异方差与 MCS 分析 .....	(392)
四、扰动项为非零均值与 MCS 分析 .....	(394)
五、随机解释变量与 MCS 分析 .....	(397)
第五节 MCS 在非经典回归分析中的应用 .....	(399)
一、DF 检验临界值与 MCS 分析 .....	(399)
二、线性趋势时间序列检验与 MCS 分析 .....	(405)
三、二次曲线趋势时间序列检验与 MCS 分析 .....	(413)



第十一章 方差分析与 MCS 研究 .....	(420)
第一节 方差分析原理与 MCS 分析 .....	(420)
一、方差分析原理 .....	(420)
二、方差分析的 MCS 检验分析 .....	(421)
第二节 违背条件下的方差分析与 MCS .....	(423)
一、异方差性与 MCS 分析 .....	(423)
二、非正态性与 MCS 分析 .....	(425)
三、非独立性与 MCS 分析 .....	(428)
第三节 违背条件组合下的方差分析与 MCS .....	(432)
一、异方差性和非正态性总体与 MCS 分析 .....	(432)
二、异方差性和非独立总体与 MCS 分析 .....	(433)
三、非独立和非正态性总体与 MCS 分析 .....	(434)
四、三种违背条件与 MCS 分析 .....	(438)
第四节 多种影响因素方差分析与 MCS .....	(438)
第十二章 Bootstrap 与 MCS 研究 .....	(446)
第一节 相关系数与 Bootstrap 研究 .....	(446)
一、相关系数检验与 Bootstrap 研究 .....	(446)
二、相关系数区间估计与 Bootstrap 研究 .....	(451)
第二节 经典假设检验与 Bootstrap 研究 .....	(459)
一、正态分布与 Bootstrap 研究 .....	(460)
二、同方差与 Bootstrap 研究 .....	(462)
三、序列不相关与 Bootstrap 研究 .....	(465)
四、结构稳定性与 Bootstrap 研究 .....	(467)
第三节 单位根检验与 Bootstrap 研究 .....	(470)
一、单位根检验量的特点 .....	(470)
二、DF 检验的 Bootstrap 实现 .....	(471)
参考文献 .....	(478)



# 第一篇 蒙特卡罗模拟与 SAS 软件基础



# 第一章 随机数的产生与检验

作为一种重要的辅助分析方法,蒙特卡罗模拟有着悠久的历史,其基础是产生特定分布的随机数,其中 0 至 1 上的均匀分布随机数是产生其他众多分布随机数的基础,而利用均匀分布随机数产生其他分布随机数的常用方法有两种,即逆转法(Inverse Transform Method)和拒绝-接受法(Rejection-Acceptance Method)。有时候我们需要产生具有一定特征分布的随机数,例如具有一定偏斜程度和相关关系的随机数。因此,本章第一节介绍蒙特卡罗模拟的发展历史,第二节介绍均匀分布随机数产生的理论方法与相关检验,第三节介绍如何利用两种产生随机数的方法生成几种常见分布随机数及其检验,最后,在第四节我们介绍如何产生具有一定特征的随机数。本章已经使用了不少的 SAS 程序,读者可以首先参考第二章的介绍,然后再阅读本章后三节的内容。

## 第一节 蒙特卡罗模拟方法发展简介

蒙特卡罗模拟(Monte Carlo Simulation)方法,或称计算机随机模拟方法,是一种基于“随机数”的计算方法,通过设定随机过程(数据生成系统),反复生成数据序列,并计算参数估计量和统计量,进而研究其分布特征的方法。这一方法源于美国第一次世界大战中研制原子弹的“曼哈顿计划”。该计划的主持人之一、数学家冯·诺伊曼用驰名世界的赌城——摩纳哥的蒙特卡罗(Monte Carlo)来命名这种方法,为它蒙上了一层神秘的色彩。作为地名,蒙特卡罗在欧洲的摩纳哥,以著名赌城而得名。若再晚些时候,蒙特卡罗模拟也许就称作拉斯维加斯(Las Vegas,位于美国的内华达(Nevada)州,著名赌城)模拟方法了。

人类很早以前就有了发现和利用蒙特卡罗模拟方法的基本思想。早在 17 世纪,人们就知道用事件发生的“频率”来估计事件的“概率”。19 世纪,人们利用蒲丰投针试验的方法来估计圆周率  $\pi$ 。随着 20 世纪 40 年代电子计算机的发明,特别是 20 世纪末以来高速电子计算机的出现,使得用数学方法在计算机上大量、快速地模拟这样的试验成为可能。

蒙特卡罗模拟方法的基本原理及思想如下:当所要求解的问题是某种事件出现的概率,或者是某个随机变量的期望值时,它们可以通过某种“试验”的方法,得到这种事件出现的频率,或者这个随机变数的平均值,并用它们作为问题的解。蒙特卡罗模拟方法通过抓住事物运动的几何数量和几何特征,利用数学方法来加以模拟,即进行一种数字模拟实验,它是以一个概率模型为基础,按照这个模型所描绘的过程,通过模拟实验得到结果,作为问题的近似解。因



此,可以把蒙特卡罗模拟解题归结为 3 个主要步骤:构造或描述概率过程、实现从已知概率分布抽样、建立各种估计量。

构造或描述概率过程:对于本身就具有随机性质的问题,如粒子输运问题,主要是正确描述和模拟这个概率过程;对于本来不是随机性质的确定性问题,如计算定积分,就必须事先构造一个人为的概率过程,它的某些参数正好是所要求问题的解,即要将不具有随机性质的问题转化为随机性质的问题。

实现从已知概率分布抽样:构造了概率模型以后,由于各种概率模型都可以看作是由各种各样的概率分布构成的,因此产生已知概率分布的随机变量(或随机向量),就成为实现蒙特卡罗方法模拟实验的基本手段,这也是蒙特卡罗模拟方法被称为随机抽样的原因。最简单、最基本、最重要的一个概率分布是 $(0,1)$ 上的均匀分布(或称矩形分布)。随机数就是具有这种均匀分布的随机变量,随机数序列就是具有这种分布总体的一个简单子样,也就是一个具有这种分布的相互独立的随机变数序列。产生随机数的问题,就是从这个分布进行抽样的问题。在计算机上,可以用物理方法产生随机数,但价格昂贵,不能重复,使用不便。另一种方法是用数学递推公式产生。这样产生的序列,与真正的随机数序列不同,所以称为伪随机数(Pseudo Random Number),或伪随机数序列。不过,经过多种统计检验表明,它与真正的随机数或随机数序列具有相近的性质,因此可把它作为真正的随机数来使用。由已知分布进行随机抽样有多种方法,与从 $(0,1)$ 上均匀分布抽样不同,这些方法都是借助于随机序列来实现的,也就是说,都是以产生随机数为前提的。由此可见,随机数是我们实现蒙特卡罗模拟的基本工具。

建立各种估计量:一般说来,构造了概率模型并从中抽样后,即实现模拟实验后,我们就要确定一个随机变量,作为所要求的问题的解,我们称它为无偏估计。建立各种估计量,相当于对模拟实验的结果进行考察和登记,从中得到问题的解。当然,还可以引入其他类型的估计,如极大似然估计、渐进有偏估计等。但是,在蒙特卡罗计算中,使用最多的是无偏估计。

## 第二节 均匀分布随机数的产生与检验

随机数在蒙特卡罗模拟中起着核心作用,因为几乎所有模拟首先的工作就是产生一定数目特定分布的随机数,因而如何产生这些随机数、随机数的随机性检验和是否满足特定分布的检验是执行蒙特卡罗模拟的前提工作,本节首先介绍一种最基本的随机数——均匀分布随机数的产生与相关检验。

### 一、均匀分布随机数的产生与检验

这里所提到的均匀分布实际特指 $U(0,1)$ ,由于对它进行适当的变换就可以得到任意其他的均匀分布,因此 $U(0,1)$ 是最基本的均匀分布,以后所提的均匀分布,如果没有特殊说明,就是指这种分布。所谓服从均匀分布随机数,实际上是指从 $U(0,1)$ 中随机抽取系列随机变量 $u_1, u_2, \dots$ 。它必须满足两个要求:首先每个随机变量 $u_i$ 均匀地落在 $[0,1]$ 上,其次是每个随机





变量  $u_i$  独立于其他随机变量。

然而,利用随机数发生器产生满足这两个条件的随机数是十分困难的,首先,随机数发生器产生的随机数一般既不会出现 0,也不会出现 1;其次,由随机数发生器产生的随机数前后具有一定的内在联系,因而也不可能是完全随机的。正因为如此,通常称由计算机产生的随机数为伪随机数,真正的随机数或称纯随机数只出现在理论研究中,以后没有特别说明,所提到的随机数就是指伪随机数。

产生均匀分布随机数的方法较多,其中使用比较广泛的是线性同余随机数生成方法 (Linear Congruential Generator)。该方法由两个公式组成,分别为

$$\begin{cases} x_{i+1} = ax_i \bmod m \\ u_{i+1} = x_{i+1}/m \end{cases} \quad (1.1)$$

其中  $a$  为乘子,函数  $\bmod$  为求余函数, $m$  是一个足够大的正整数,这也是这种方法得名的原因。使用这种方法首先必须给定一个初始值  $x_0$ ,也形象地称为种子值 (Seed),其取值范围必须是  $1 \sim m$  的整数,通常由用户自行设定,但只要初始种子值相同,产生的随机数序列就完全一样。为了更好地说明随机数是如何利用上述公式产生的,现令  $a=6, m=11, x_0=1$ ,则产生随机数的过程如下:

$$\begin{aligned} x_1 &= 6 \times 1 \bmod 11 = 6, & u_1 &= 6/11 \\ x_2 &= 6 \times 6 \bmod 11 = 3, & u_2 &= 3/11 \\ x_3 &= 6 \times 3 \bmod 11 = 7, & u_3 &= 7/11 \\ x_4 &= 6 \times 7 \bmod 11 = 9, & u_4 &= 9/11 \\ x_5 &= 6 \times 9 \bmod 11 = 10, & u_5 &= 10/11 \\ x_6 &= 6 \times 10 \bmod 11 = 5, & u_6 &= 5/11 \\ x_7 &= 6 \times 5 \bmod 11 = 8, & u_7 &= 8/11 \\ x_8 &= 6 \times 8 \bmod 11 = 4, & u_8 &= 4/11 \\ x_9 &= 6 \times 4 \bmod 11 = 2, & u_9 &= 2/11 \\ x_{10} &= 6 \times 2 \bmod 11 = 1, & u_{10} &= 1/11 \end{aligned}$$

从这一组参数取值中我们不难发现,当随机数产生到一定数目时,就会出现重复,这是这种方法难以避免的,再继续向下运算,余数从 1 至 10 都出现了,这种情况称为全周期。如果更改上述的参数设置,就可能不会出现全周期情况,例如取  $a=3$ ,其他参数不变,整个过程为

$$\begin{aligned} x_1 &= 3 \times 1 \bmod 11 = 3, & u_1 &= 3/11 \\ x_2 &= 3 \times 3 \bmod 11 = 9, & u_2 &= 9/11 \\ x_3 &= 3 \times 9 \bmod 11 = 5, & u_3 &= 5/11 \\ x_4 &= 3 \times 5 \bmod 11 = 4, & u_4 &= 4/11 \\ x_5 &= 3 \times 4 \bmod 11 = 1, & u_5 &= 1/11 \end{aligned}$$

显然,这里只有 5 个值,如果取  $a=3, x_0=2$ ,则余数为 2、6、7、10、8,实际上也只有 5 个数。而通常的蒙特卡罗模拟往往需要成千上万个随机数,因此必须选择一个足够大的  $m$ ,同时也必须选择合适的乘子  $a$ ,表 1.1 给出了几种取值组合情况,可以确保研究需要。