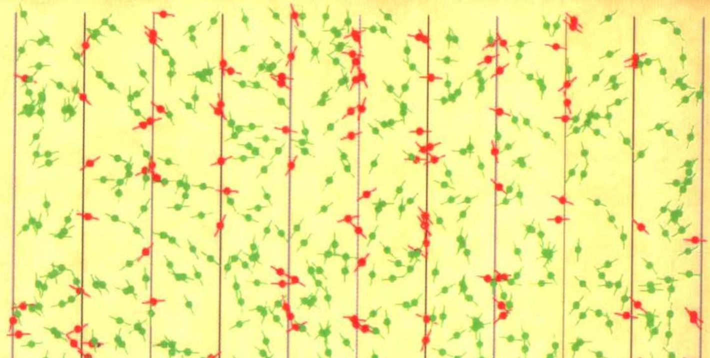
TURING 图灵数学·统计学丛书 11

ELSEVIER 爱思唯尔



Simulation

统计模拟

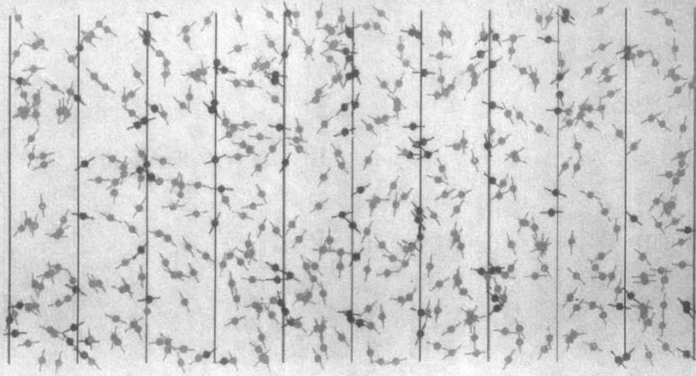
（第4版）

[美] Sheldon M. Ross 著

王兆军 陈广雷 邹长亮 译

人民邮电出版社
POSTS & TELECOM PRESS

TURING 图灵数学·统计学丛书 11



Simulation
统计模拟

（第4版）

Sheldon M. Ross 著
陈广雷 邹长亮 译

人民邮电出版社
北 京

图书在版编目（CIP）数据

统计模拟：第4版 /（美）罗斯，（Ross, S. M）著；王兆军，陈广雷，邹长亮译.
—北京：人民邮电出版社，2007.7
（图灵数学·统计学丛书）

ISBN 978-7-115-16085-0

Ⅰ. 统… Ⅱ. ①罗…②王…③陈…④邹… Ⅲ. 统计－模拟实验－高等学校－教材
Ⅳ. C8-33

中国版本图书馆 CIP 数据核字（2007）第 052005 号

内 容 提 要

统计模拟是数理统计中非常有用的工具之一，它是利用计算机产生某概率模型的随机数，再通过这些随机数来模拟真实模型. 本书首先介绍了产生某些分布随机数的一些方法，之后又较详细地介绍了统计模拟中常用的一些方法，如离散事件模拟方法、方差缩减技术、模拟数据的统计分析方法、统计验证方法、MCMC 方法等；并通过某些实例，对这些方法的应用进行了较详细的说明. 本书最后还提供了不同难度的习题.

本书可作为高等院校数学、统计学、科学计算、保险学和精算学等专业的本科教材，也可作为工程技术人员和精算师等应用工作者的参考用书.

图灵数学·统计学丛书

统计模拟（第4版）

◆ 著　［美］Sheldon M. Ross
译　王兆军　陈广雷　邹长亮
责任编辑　明永玲　王　利

◆ 人民邮电出版社出版发行　北京市崇文区夕照寺街 14 号
邮编　100061　电子函件　315@ptpress.com.cn
网址　http://www.ptpress.com.cn
北京铭成印刷有限公司印刷
新华书店总店北京发行所经销

◆ 开本：700×1000　1/16
印张：16.75
字数：344 千字　2007 年 7 月第 1 版
印数：1–4 000 册　2007 年 7 月北京第 1 次印刷

著作权合同登记号　图字：01-2007-1485 号

ISBN 978-7-115-16085-0/O1

定价：39.00 元

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

版权声明

Simulation, 4th Edition by Sheldon M. Ross, ISBN: 0-12-598063-9.

Authorized Simplified Chinese translation edition published by the Proprietor.
ISBN: 978-981-259-847-9.

Elsevier (Singapore) Pte Ltd.
3 Killiney Road,
#08-01 Winsland House I
Singapore 239519
Tel: (65)6349-0200
Fax: (65)6733-1817

First Published 2007
2007 年初版

译者简介

王兆军 南开大学教授、博士生导师. 现任南开大学数学科学学院统计学系系主任、中国概率统计学会理事、中国现场统计研究会理事、天津数学会秘书长. 1987年毕业于南开大学，获理学学士学位；1990年毕业于华东师范大学，获理学硕士学位；1995年于南开大学获理学博士学位.

陈广雷 中国人民武装警察部队学院副教授. 1992年毕业于东北师范大学，获理学学士学位；2000年毕业于北京师范大学，获理学硕士学位；现为南开大学在读博士生.

邹长亮 2003年毕业于南开大学，获理学学士学位；2006年毕业于南开大学，获理学硕士学位；2006年9月起在南开大学攻读博士学位.

译 者 序

统计模拟方法不仅是统计专业的一门重要课程，而且也是越来越多的理工学院、商学院、经济学院、医学院、农学院部分专业中的选修课程. 虽然国内目前关于统计计算的教材已有若干本，但却没有一本书专门系统介绍统计模拟，尤其是最近几年发展起来且得到广泛应用的新方法，如 Bootstrap(自助法)、MCMC(马氏链蒙特卡洛法) 和模拟退火算法等. 然而，这本由 Sheldon M. Ross 编写的《统计模拟》教材，恰好是这方面的补充与完善.

本书首先介绍了随机数的产生方法，之后又给出了几种产生连续与离散随机变量的方法，并通过某些例子简述了某些模拟方法及模拟数据的统计分析方法. 然后，提出了若干方差缩减技术，如对偶变量法、控制变量法、分层抽样法、重要抽样法、公共变量法等，并介绍了几种模拟检验方法，如 χ^2 拟合优度检验、科尔莫戈罗夫－斯米尔诺夫检验等. 在介绍 MCMC 方法的一章，又讲述了 Hasting-Metropolis 算法、吉布斯抽样、模拟退火和重要抽样方法等. 最后又对随机模拟的其他一些专题进行了讨论.

本书在讲述方法的同时，还注重这些方法在金融、优化等方面的应用，且给出了较丰富的参考文献. 虽然本书内容很丰富，但因其所需的概率统计知识相对较少，所以很适合作为本科生自学或教材用书，而且其中某些内容也可供统计专业的、研究生参考学习.

我们很高兴能有机会将该书推荐给国内的读者，也非常感谢人民邮电出版社图灵公司的编辑在本书出版过程中给予我们的大力支持和帮助. 另外，也要感谢几位没有署名的审稿者.

本书的翻译工作是由 3 名作者合作完成的，其中第 1 章至第 6 章由王兆军翻译，第 7 章、第 8 章由陈广雷翻译，第 9 章至第 11 章由邹长亮翻译，全书由王兆军统校.

由于译者无论是英文、中文水平还是统计专业知识都很有限，译文中难免会有不妥之处，欢迎广大读者批评指正.

译　者

2006 年 12 月于南开园

前　　言

概述

在用随机模型去描述实际现象时，人们既要考虑所选模型的逼真性，又要考虑所选模型的可分析性. 即是说，一个无法进行数学分析的模型，即使它能很好地反映实际问题，也是毫无用处. 于是，人们转而把注意力放在考虑一个能近似反映实际问题的模型. 然而，随着高性能且价格相对便宜的计算机的出现，产生了一种用计算机尽可能真实地模拟随机现象，之后再利用模拟结果去分析它的新方法.

本书将展示如何利用随机模拟结果对模型进行分析的方法. 特别地，本书将给出如何利用计算机产生随机数（或称伪随机数）的方法，讨论如何利用这些随机数来产生任一分布的随机变量的值，阐述如何利用离散随机事件的概念去模拟一个依时间而变化的随机模型的行为，以及给出通过模拟一个随机系统的行为而得到某感兴趣参数的估计. 在模拟过程中，我们将考虑何时停止模拟及如何给定所考虑的估计的置信度等问题. 本书将给出几种能改进常用模拟估计的模拟方法. 另外，我们还将探讨如何利用随机模拟去判断所用随机模型是否与实际数据相符的问题.

本版的新内容

本版的变化主要体现在以下几个方面.

- 大部分章都增加了新习题.
- 增加关于生成贝努利随机变量序列的新结果 (例 4e).
- 关于选取最优指数分布的筛选法生成伽玛随机变量的新结果 (5.2 节).
- 新增例 8p 给出了求取当所有癌细胞被杀死后仍存活的健康细胞数目的分布. 重写了关于分层抽样的 8.4 节，包括有关事后分层及每层最优模拟次数的材料.
- 新增 8.5 节, 此节包括把分层抽样应用于到达为泊松过程的系统分析 (8.5.1 节)、单调函数的多重积分的计算 (8.5.2 节) 和复合随机向量 (8.5.3 节) 三部分.
- 新增了有效利用方差缩减技术计算随机排列和随机子集函数的 8.9 节.

各章内容

第 1 章 是绪言，它给出了一个有意思的典型案例. **第 2 章**是概率论简介. 本章自成体系并在假设读者不熟悉概率论的情况下给出简单介绍. **第 3 章**涉及随机数及如何由计算机产生伪随机数. **第 4 章**和**第 5 章**考虑如何利用随机数来产生离散随机变量和连续随机变量.

第 6 章 介绍离散随机事件方法以分析随时间而变化的随机系统, 并给出了多个例子, 如涉及单个或多服务员的排队系统、保险风险模型、仓储问题、机器维修模型及股票期权的执行模型等.

第 7 章 的主要内容是统计, 我们假设本书的一般读者事先没有学过统计. 本章先从非常基本的统计概念开始讲起, 最后以自助法 (Bootstrap) 结束 (自助法在随机模拟中非常有用).

第 8 章 涉及方差缩减技术这一重要内容. 我们将尝试通过寻找一个具有相同均值和较小方差的估计以改进常用的模拟估计. 本章首先介绍对偶变量 (Antithetic variables) 方法. 本章附录的证明显示, 当把此方法用来估计某单调函数的期望时, 它总能减小方差, 并只占用较少的计算机存储空间. 其次, 我们引入控制变量并阐述它在减小方差上的作用, 且用在排队系统、可靠性系统、列表重排序问题和 21 点游戏 (Blackjack) 中的例子说明如何有效地利用控制变量. 我们也将提出如何利用回归软件包承担控制变量的计算量. 之后考虑利用条件期望以减小方差, 并把它应用于估计 π 和分析有限容量的排队系统中. 结合控制变量法, 条件期望法也被用来估计更新过程中某时刻的平均事件数. 作为减小方差的另一工具, 分层抽样被应用于处理带有变化到达率的排队问题和积分求取的例子中. 在估计某视频纸牌游戏平均收益的例子中, 我们解释和说明了条件期望和分层抽样这两种减小方差方法间的关系. 分层抽样在到达为泊松过程的排队系统、多重积分的计算和复合随机向量中的应用. 然后, 我们考虑重要抽样法. 重要抽样方法是一种非常有效的估计小概率的方差缩减方法. 为说明这一点, 我们引入倾斜分布的概念, 说明如何利用它和重要抽样法来估计一个卷积的小尾部概率. 我们还把重要抽样方法应用于排队论、随机游动、随机排列和基于小概率事件的条件期望的计算上. 本章最后的一个方差缩减方法涉及公共随机数流的应用. 8.7 节介绍用多种方差缩减技术相结合为奇异股票期权 (exotic stock option) 定价.

第 9 章 涉及某些统计验证方法. 当有实际数据可用时, 这些统计方法可以用来验证随机模型的有效性. 本章将介绍如 χ^2 检验和科尔莫戈罗夫 – 斯米尔诺夫等拟合优度检验. 本章的其他小节讨论两样本问题、多样本问题和检验一个过程是否为泊松过程的统计假设检验.

第 10 章 涉及马氏链蒙特卡洛 (MCMC) 方法. 这些方法极大地拓展了近几年随机模拟的应用. 对于一个随机向量 $\boldsymbol{X}$, 用随机模拟方法估计 $\theta = E[h(\boldsymbol{X})]$ 的标准步骤, 是先模拟产生多个与 $\boldsymbol{X}$ 独立同分布的随机变量, 之后用基于这些随机变量的 $h(\boldsymbol{X})$ 的均值作为 θ 的估计. 这就是通常所说的原始估计, 它可以用第 8 章中的方差缩减技术进行改进. 然而, 要使用这种方法, 就必须知道 $\boldsymbol{X}$ 的分布且能模拟此分布. 但正如本章的许多例子, 虽然我们知道 $\boldsymbol{X}$ 的分布, 但我们却不能直接模拟

此随机向量 $\boldsymbol{X}$; 又如本章的其他例子, 虽然其分布可以写成若干因子的乘积, 但并不完全已知. 对于这两种情况, 上述通常估计 θ 的方法均不可行. 但是近几年, 一个基于生成一个马氏链的新方法在随机模拟中得到了广泛的应用, 此马氏链的极限分布同 $\boldsymbol{X}$ 的分布, 且用函数 h 在此马氏链的各状态点的平均值估计 θ. 本章将研究 MCMC 方法. 我们首先在 10.2 节引入马氏链并给出它的某些性质. 10.3 节介绍 Hastings-Metropolis 算法, 它是用来生成一个极限分布为给定因子乘积的马氏链的常用方法. 一种产生来自一"组合"集之随机元素的方法也在 10.3 节给出. 10.4 节介绍的吉布斯抽样是最常用的 Hastings-Metropolis 算法. 本节讨论的例子有: 在某区域内生成任两点距离均大于给定值的随机点, 乘积排队网络的分析, 用来预测某些棒球选手本垒打次数的分层贝叶斯统计模型的分析, 以及基于所有试验结果都至少出现一次的条件多项分布的随机模拟等. 10.5 节介绍用于确定性优化问题的模拟退火方法及一个旅行商问题的例子. 本章最后一节介绍重要重抽样算法, 它是第 4 章和第 5 章的筛选 (Acceptance-Rejection) 抽样法的推广, 本节也考虑此算法在贝叶斯统计中的应用.

第 11 章 涉及随机模拟的其他一些专题. 我们将在 11.1 节中熟悉别名方法, 如不考虑计算时间, 它是产生离散随机数的非常有效的方法. 11.2 节讨论二维泊松过程的模拟问题. 我们在 11.3 节给出一个关于相关伯努利随机变量和的协方差的等式, 并指出如何利用此等式估计某些小概率, 且估计方差非常小. 本节还考虑应用此等式去估计系统可靠性 (此估计的效率在估计小系统可靠性问题时高于现有的其他的估计), 去估计指定模式在给定时间发生的概率等问题. 11.4 节考虑马氏链首达时的均值和分布的模拟估计, 以及二维正态变量的尾部概率的计算. 在上述内容的基础上, 11.5 节给出了服从指定马氏链平稳分布的随机变量的模拟方法.

致谢

我们十分感激 Yontha Ath(加州大学长滩分校)、David Butler (俄勒冈州立大学)、Matt Carlton (加利福尼亚州立理工大学)、James Daniel (得克萨斯大学奥斯汀分校)、William Frye (鲍尔州立大学)、Mark Glickman (波士顿大学)、Chuanshu Ji (北卡大学)、Yonghee Kim-Park (加州大学长滩分校)、Donald E. Miller (圣玛丽学院)、Krzysztof Ostaszewski (伊利诺伊州立大学)、Bernardo Pagnocelli, Erol Peköz (波士顿大学)、Yuval Peres (加州大学伯克利分校) 和 Esther Portnoy (伊利诺伊大学厄巴纳 – 尚佩恩分校) 等, 他们提出了十分有益的建议. 我们还要谢谢那些希望匿名的本书的审稿者.

目　录

第1章 绪　　言

考虑如下案例. 一个药剂师想建立一个自己填写处方的小药店. 他计划早九点开门, 并希望下午五点前平均每天 (周一至周五期间) 有 32 个处方需要他填写. 经验显示, 他填写一张处方所需的时间是一个均值和标准差分别为 10 分钟和 4 分钟的随机变量. 他假设下午五点后不再收新处方; 但若需要, 他将留在药店把一天中没有完成的处方写好. 对于此案例, 这个药剂师感兴趣的问题可能是:

1. 平均来看, 他晚上几点可以离开药店?

2. 晚上五点半时他仍在药店工作的天数的比例是多少?

3. 他填写一张处方所需的平均时间是多少? (要考虑到填写新处方时需等待他写完前面所有的处方)

4. 30 分钟之内能完成的处方比例是多少?

5. 如果他把早九点至晚五点期间接受所有处方的策略做一改变, 只有当手中需要填写的处方少于 5 个时才接受新处方, 则平均每天会流失多少处方?

6. 上述策略的改变将如何影响问题 1 ~ 4 的答案?

为了用数学方法分析此案例并回答上述问题, 我们需要先建立一个概率模型. 为此, 我们有必要对上述案例做某些合理的确切的假设. 比如, 为了用一概率模型 1
描述平均每天 32 位顾客的到来时间, 我们必须对此做些假设. 一个可能的假设是一天中单位时间内 (比如每小时) 到来的顾客人数在概率意义下是一个常数, 尽管另一个更接近实际的假设是顾客到来的人数与时间有关. 然后, 我们必须指定一个概率分布 (均值为 10, 标准差为 4) 来刻画填写一个处方的所需时间, 以及做一些假设以保证一个处方所需的服务时间服从这个分布或作为其他随机变量的函数 (如等待服务处方的数目或时间). 这就是说, 我们必须对顾客到来时间和所需服务时间做某些概率假设, 也必须判断用来描述一天的概率分布是依赖于这一天是星期几, 还是个与时间无关的常数. 当这些或其他的假设被指定后, 我们就可以构造此案例的概率模型了.

一旦概率模型确定后, 就可以从理论上定量地给出上述问题的答案. 然而, 在实际中我们很难定量地给出这些问题的答案, 于是, 我们一般通过随机模拟来给出答案. 随机模拟就是通过计算机编程产生随机数以模拟此模型在很多天内的可能结果, 并且利用统计理论估计上述问题的答案. 换句话说, 通过计算机编程产生用以描述顾客到来时间和服务所需时间的具有特定概率分布的随机变量的值, 并由此

得到多天的关于上述问题的随机变量的值，之后利用统计方法估计它们. 例如，如果 1000 天的随机模拟结果显示药剂师在五点半仍在工作的有 122 天，则我们估计问题 2 的答案为 0.122.

为了能完成上述分析，人们必须懂得某些概率论知识，如概率分布、随机变量的独立性等，因此第 2 章为概率论简介. 由于随机数是随机模拟研究的基础，故第 3 章将讨论随机数及如何由计算机产生随机数. 第 4、5 章给出产生服从任一分布的随机数的方法（第 4、5 章分别考虑离散与连续分布的随机数）. 学完第 5 章之后，读者就应有能力对实际问题建立概率模型，并知道如何产生此模型相应随机变量的随机数. 由于有些系统是依时间连续变化的，故第 6 章将考虑如何把上述产生随机数的方法应用于一个实际系统. 在第 6 章中，我们将提出“离散随机事件”的概念，并基于它而得到模拟一系统的系统化方法. 离散随机事件模拟方法即是通过
2 计算机编程（可用读者熟悉的任一种计算机语言）多次模拟此系统而使问题得以解决. 验证此程序可行与否的某些步骤也将在第 6 章给出. 在利用模拟输出结果解答所关心的模型的问题时必须用到某些统计理论，这些将在第 7 章介绍. 第 7 章将首先介绍一些最简单实用的统计概念，之后再引入最近发展起来的在随机模拟中非常有用的“自助法 (Bootstrap)”. 统计研究表明，由随机模拟得到的估计值的方差是衡量随机模拟效率的一个重要指标. 估计值的方差越小，就表明为得到某一给定精度所需的模拟次数越少. 第 8 章将给出一些改进模拟估计方差的方法. 由于方差减小可以提高模拟效率，故减小方差是随机模拟研究中非常重要的课题. 当有实际数据可用时，如何利用随机模拟的结果去验证概率模型与实际情况的接近程度的方法将在第 9 章给出. 第 10 章介绍马氏链蒙特卡洛方法 (MCMC). 近些年，MCMC 方法极大地拓广了随机模拟的研究领域. 第 11 章研究其他一些主题.

习　题

1. 下面数据是一个单服务员系统中前 13 个顾客的到达时间和所需的服务时间. 当服务员空闲时，顾客到达后立即得到服务；否则，就要排队等候. 当服务员完成一位顾客的服务后，则立即服务下一个等候时间最长的顾客.

到达时间:	12	31	63	95	99	154	198	221	304	346	411	455	537
服务时间:	40	32	55	48	18	50	47	18	28	54	40	72	12

(a) 求这 13 位顾客的离开时间.

(b) 如果有两个服务员且每一服务员可以服务任一顾客，求这 13 位顾客的离开时间.

(c) 如果假设服务员完成一位顾客的服务后，对等候时间最短的顾客进行服务，求这 13 位顾客的离开时间.

2. 考虑一个先来先服务的服务站. 分别以 A_n, S_n 和 D_n 记顾客 n 的到达时间、所需服务时间和离开时间. 假设此服务站只有一名服务员且开始时站内无顾客. 3

(a) 当 $D_0 = 0$ 时, 请验证

$$D_n - S_n = \max\{A_n, D_{n-1}\}, \quad \forall\, n > 0$$

(b) 当有两名服务员时, 请给出相应的递推公式.

(c) 当有 k 名服务员时, 请给出相应的递推公式.

(d) 请编写求取离开时间的计算机程序 (离开时间是到达时间和所需服务时间的函数), 并验证练习 1 中 (a) 和 (b) 的答案. 4

第2章　概率基础

2.1　样本空间和事件

对于一个事先不知确切结果的试验，其所有可能试验结果构成的集合称为样本空间，记为 S. 例如，考虑一个编号为 1 至 7 的七匹马的赛马试验，其样本空间为

$$S = \{(1,2,3,4,5,6,7)\text{的所有排列}\}$$

其中 $(3,4,1,7,6,5,2)$ 表示 3 号马跑第一，4 号马跑第二，等等.

样本空间的任一子集 A 称为一个事件，即事件是一个包含试验的某些可能结果的集合. 如果一个试验结果包含在 A 中，则我们说事件 A 发生或出现. 例如，在上例中，如果

$$A = \{S\text{中第一个数为5的试验结果}\}$$

则 A 是 5 号马跑第一的事件.

对于任何两个事件 A 和 B，我们定义一个称为二者并的新事件 $A \cup B$，它包含事件 A 或 B 中的所有试验结果. 同样，我们定义一个称为二者交的新事件 AB，它仅包含既在事件 A 又在事件 B 中的所有试验结果. 这就是说，如事件 A 或 B 发
5 生，则事件 $A \cup B$ 发生；如事件 A 和 B 同时发生，则事件 AB 发生. 我们同样可以定义多个事件的并与交. 即，事件 $A_1, \cdots, A_n$ 的并 $\cup_{i=1}^{n} A_i$，被定义为包含在任一 A_i 中的所有试验结果的事件；事件 $A_1, \cdots, A_n$ 的交 $A_1A_2\cdots A_n$，被定义为包含在所有 A_i 中的所有试验结果的事件.

对于任一事件 A，我们定义其补事件为在样本空间 S 中但不在 A 中的所有试验结果的集合，记为 A^c. 事件 A^c 发生当且仅当事件 A 不发生. 因为试验结果肯定在样本空间 S 中，故 S^c 不包含任何试验结果且它不会发生. 我们称 S^c 为空集且记为 $\emptyset$. 如果 $AB = \emptyset$，则 A 与 B 不会同时发生（其原因为二者没有相同试验结果），我们称之为互斥.

2.2　概率公理

假设对于一个样本空间为 S 的试验中的任一事件 A，存在一个常数 $P(A)$，满足下面三个公理：

公理 1 $0 \leqslant P(A) \leqslant 1$.

公理 2 $P(S) = 1$.

公理 3 对任一列互斥事件$A_1, A_2, \cdots$,有

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i),\ n = 1, 2, \cdots, \infty$$

则称$P(A)$为事件 A 的概率.

公理 1 指出事件 A 中任一试验结果出现的概率介于 0 与 1 之间；公理 2 指出样本空间中的试验结果以概率 1 出现；公理 3 指出任意多个互斥事件，至少发生其中之一的概率等于各事件概率之和.

由此三个公理，概率的许多结论可以得到证明. 比如，由于 A 和 A^{C} 总是互斥的，且 $A \cup A^{\mathrm{C}} = S$，于是由公理 2 和公理 3 有

$$1 = P(S) = P(A \cup A^{\mathrm{C}}) = P(A) + P(A^{\mathrm{C}})$$

或等价地有

$$P(A^{\mathrm{C}}) = 1 - P(A)$$

也可以说，一个事件不发生的概率等于 1 减去此事件发生的概率. 6

2.3 条件概率和独立性

考虑掷一枚硬币两次的试验，因为每次试验结果不是正面就是反面，故此试验的样本空间可以写成

$$S = \{(\mathrm{H,H}), (\mathrm{H,T}), (\mathrm{T,H}), (\mathrm{T,T})\}$$

其中 (H,T) 表示第一次掷出现正面，第二次掷出反面. 现假设四种结果的出现是等可能的，则每一结果出现的概率为 1/4. 如果已知第一次试验的结果为正面，则在此信息下，两次结果均为正面的概率是多少？为求此概率，我们推理如下：当给定第一次结果为正面时，试验结果最多有两种情况：(H,H) 或 (H,T). 由于这两种结果在原两次投掷试验中发生的概率相同，故在此试验中我们可以假设二者发生的概率相同，即当给定第一次结果为正面时，上述两种结果 (H,H) 和 (H,T) 发生的条件概率各为 1/2, 另两个结果的条件概率为 0. 因此，所求的概率即为 1/2.

如果以 A 和 B 分别表示两次投掷均为正面的事件和第一次投掷为正面的事件，则上面得到的概率即称为 B 发生下 A 的条件概率，记为

$$P(A|B)$$

对于所有随机试验和任两事件 A 和 B, 我们均可以用上述同样方法得到 $P(A|B)$ 的一般表达式. 也就是说, 事件 B 发生后, 为使事件 A 发生, 则出现的试验结果必须同时属于 A 和 B, 即在 AB 中. 因为 B 已经发生, 故 B 即为新的样本空间. 因此, 事件 $A|B$ 发生的概率等于事件 AB 相对于事件 B 的概率, 即

$$P(A|B)=\frac{P(AB)}{P(B)}$$

为确定事件 A 发生的概率, 人们经常通过另一事件 B 作如下考虑: 求在事件 B 发生条件下 A 的条件概率及在事件 B 不发生条件下 A 的条件概率, 之后注意到

7

$$A=AB\cup AB^{\mathrm{C}}$$

又因为 AB 与 AB^{C} 互斥, 故有

$$P(A)=P(AB)+P(AB^{\mathrm{C}})=P(A|B)P(B)+P(A|B^{\mathrm{C}})P(B^{\mathrm{C}})$$

当我们利用上一公式时, 我们就说我们是利用**事件 B 发生与否的条件**来计算事件 A 的概率 $P(A)$ 的.

例 2a 一保险公司将投保人分成易出险与不易出险两类. 其数据表明, 易出险客户在一年内提出一次索赔的概率为 0.25, 而不易出险客户在一年内提出一次索赔的概率为 0.1. 如果一个新的投保人是易出险客户的概率为 0.4, 则他在一年内索赔一次的概率为多少?

解 以 C 表示一年内索赔一次的事件, B 表示此投保人为易出险客户的事件. 则

$$P(C)=P(C|B)P(B)+P(C|B^{\mathrm{C}})P(B^{\mathrm{C}})=(0.25)(0.4)+(0.1)(0.6)=0.16.$$

□

假设事件 B_i, $i=1,\cdots,n$ 中恰有一个事件发生, 即假设事件 $B_1,B_2,\cdots,B_n$ 互斥且它们的并为样本空间 S. 于是, 我们可以利用 B_i 发生的条件来计算事件 A 的概率, 其原因为

$$A=AS=A\left(\bigcup_{i=1}^{n}B_i\right)=\bigcup_{i=1}^{n}AB_i$$

由此可得

$$P(A)=\sum_{i=1}^{n}P(AB_i)=\sum_{i=1}^{n}P(A|B_i)P(B_i)$$