

**Markov Processes
and Applications:
Algorithms, Networks,
Genome and Finance**

**马尔可夫过程
及其应用：**

算法、网络、基因与金融

Étienne Pardoux 著
许明宇 译

非外借

高等教育出版社

**Markov Processes
and Applications:**
Algorithms, Networks,
Genome and Finance

**马尔可夫过程
及其应用：**

算法、网络、基因与金融

Étienne Pardoux 著
许明宇 译

图书在版编目(CIP)数据

马尔可夫过程及其应用:算法,网络,基因与金融 /
(法)艾蒂安·巴赫杜(Etienne Pardoux)著;许明宇
译. -- 北京:高等教育出版社,2019.3

ISBN 978-7-04-051380-6

I. ①马… II. ①艾… ②许… III. ①马尔柯夫过程
IV. ①O211.62

中国版本图书馆CIP数据核字(2019)第036776号

策划编辑 李华英
责任校对 王雨

责任编辑 李华英
责任印制 毛斯璐

封面设计 李树龙

版式设计 马敬茹

出版发行 高等教育出版社
社址 北京市西城区德外大街4号
邮政编码 100120
印刷 高教社(天津)印务有限公司
开本 787mm×1092mm 1/16
印张 19.75
字数 360千字
购书热线 010-58581118
咨询电话 400-810-0598

网 址 <http://www.hep.edu.cn>
<http://www.hep.com.cn>
网上订购 <http://www.hepmall.com.cn>
<http://www.hepmall.com>
<http://www.hepmall.cn>
版 次 2019年3月第1版
印 次 2019年3月第1次印刷
定 价 79.00元

本书如有缺页、倒页、脱页等质量问题,请到所购图书销售部门联系调换
版权所有 侵权必究
物料号 51380-00

前言

这本书的基础理论部分包括第一章中介绍的蒙特卡罗 (Monte Carlo) 方法, 第二章中取值于有限集或可数集的离散时间的马尔可夫 (Markov) 链 (马氏链) 理论, 以及第六章和第七章的泊松 (Poisson) 过程和带跳的马尔可夫过程理论。后两章中讨论的过程为时间连续的取值于有限集或可数集的过程。

基于这些理论基础, 书中给出了其在多个方面的许多应用。第三章中讨论了随机算法, 主要研究马氏链蒙特卡罗方法, 这个方法是 20 世纪 50 年代在统计力学研究中引入的。利用像变换, 这个方法成为贝叶斯统计中的重要方法, 特别是当数据量大且复杂的时候。另一个重要的方法是随机优化算法, 即模拟退火法。

另一个重要的应用是在分子生物学中的应用。书中讨论了两个独立的应用。第四章研究了 DNA 过程 (即基因链) 的解释与排列。这里首要的工具是隐马氏链, 而这个方法也可以用于信号处理和语音辨识。另一个应用是生态学中的应用, 特别是多个现存物种之间的关系, 而这个关系可以用“物种发展树图”来描述和研究。物种发展树图的模型构造有许多方法, 这里讨论了概率进化模型。第七章研究了带跳的马尔可夫过程及其在这一方面的应用。

第五章介绍控制问题和滤波问题, 包括著名的 Kalman-Bucy 滤波问题, 它有很多应用, 比如经常用在卫星导航中。

第八章的主题是排队论和网络 (例如互联网, 电话线网等)。

最后在第九章给出了金融数学的一个介绍。其中包括离散时间模型和连续时间模型。特别还包括伊藤 (Itô) 随机分析和扩散过程, 这类过程是连续时间取值于欧几里得空间 \mathbb{R}^d 的马尔可夫过程。需要提醒注意的一点是, 这是唯一的有

一些未给出证明的结果的一章, 因为如果我们把所有证明都写出来, 这章就太长了, 而这些证明在别的书中很容易找到。

每一章都配有一定数目的习题。某些习题是针对书中讲的算法编写程序, 即如何利用程序, 例如 Matlab, 实现数值模拟。一半以上的习题的解答可以参考第十章的习题解答。其余的习题解答可以在网站 www.dunod.com 上找到。作者建议学生自己来做习题, 而不用参考习题解答。这是学生理解本书的最好方法。大部分的习题可以看做“教材”内容的一部分, 部分习题则是对于书中的理论的多种应用。

这本书最初是基于法国工程师学校的二三年级多门课程教材 (首先是马赛力学高等专科学校, 然后是马赛中心大学) 及普罗旺斯大学的硕士教材整合而成的。它包含了关于马尔可夫过程及其应用的多门课程的内容。本课程安排在硕士一年级的第二学期, 在硕士二年级安排为选修。同样本课程对选择“概率统计”专业的学生也很有用。

本书 (特别是第九章) 要求学生了解概率论和测度论的基础理论, 包括条件期望理论, 学生应该在硕士一年级时学到这些内容。事实上, 本书中的绝大部分内容的理解仅用到离散分布和密度函数, 以及常见的极限定理 (例如大数定律与中心极限定理), 而这些内容对于数学专业本科毕业生和工程师学校一年级的学生应该是不陌生的。

在此我感谢 Geneviève Foissac, Julien Berestycki, Fabienne Castell, Yuri Golubev, Arnaud Guillin, Stéphanie Léocard, Laurent Miclo, 还有 Rémi Rhodes

Étienne Pardoux
法国马赛, 2007 年 3 月

目录

第一章 蒙特卡罗方法与模拟	1
1.1 方法的介绍	2
1.2 收敛性定理	3
1.3 随机变量的模拟	5
1.4 减小方差技巧	9
1.5 习题	13
第二章 马氏链	17
2.1 定义与基本性质	17
2.2 一些例子	21
2.2.1 在 $E = \mathbb{Z}^d$ 上的随机游走	21
2.2.2 Bienaymé–Galton–Watson 过程	22
2.2.3 离散事件的等待序列	22
2.3 强马尔可夫性	22
2.4 常返态和暂留态	25
2.5 不可约常返的情形	27
2.6 非周期情形	33
2.7 可逆马氏链	39
2.8 均衡态的收敛速度	40
2.8.1 可翻转的有限状态情形	41

2.8.2 一般情形	43
2.9 马氏链的统计结果	43
2.10 习题	45
第三章 随机算法	57
3.1 马氏链蒙特卡罗方法	57
3.1.1 一个应用	59
3.1.2 Ising 模型	61
3.1.3 图像的贝叶斯分析	63
3.1.4 加热的马氏链	64
3.2 不变概率的模拟	65
3.2.1 完美模拟	65
3.2.2 与历史耦合	68
3.3 收敛到不变测度的速度	71
3.4 模拟退火法	73
3.5 习题	75
第四章 马氏链与基因组	78
4.1 如何解读 DNA?	78
4.1.1 CpG 岛	79
4.1.2 在原核基因组中探测基因	79
4.2 独立同分布序列模型	80
4.3 马尔可夫模型	81
4.3.1 在 CpG 岛中的应用	81
4.3.2 在原核生物的整组基因中寻找基因	82
4.3.3 马氏链 M_k 的统计	83
4.3.4 分段马氏链	83
4.3.5 局部齐次马氏链	84
4.4 隐马氏链	85
4.4.1 似然估计的计算	86
4.4.2 Viterbi 算法	87
4.4.3 参数估计	88
4.5 隐半马尔可夫模型	93
4.5.1 隐马尔可夫模型的局限	93
4.5.2 什么是半马氏链?	94
4.5.3 隐半马尔可夫模型	95

4.5.4	半马氏链的 Viterbi 算法	96
4.5.5	在原核生物基因组中寻找基因	96
4.6	对齐两个序列	98
4.6.1	Needleman-Wunsch 算法	99
4.6.2	隐马氏链比对算法	100
4.6.3	对齐序列的后验概率分布	104
4.6.4	一个配对的后验概率	106
4.7	一个多重比对算法	107
4.8	习题	109
第五章	马氏链的控制和滤波	110
5.1	确定性最优控制	110
5.2	马氏链的控制	112
5.3	线性二次最优控制	113
5.4	马氏链的滤波	114
5.5	Kalman-Bucy 滤波	116
5.5.1	目标	116
5.5.2	滤波问题的解	117
5.6	部分观测的二次线性控制	121
5.7	习题	123
第六章	泊松过程	124
6.1	点过程和记数过程	124
6.2	泊松过程	125
6.3	马尔可夫性	128
6.4	渐近行为	131
6.5	习题	133
第七章	带跳马尔可夫过程	137
7.1	一般性结果	137
7.2	无穷小生成元	141
7.3	强马尔可夫性	144
7.4	嵌入马氏链	146
7.5	状态分类: 暂留态和常返态	149
7.6	不可约的常返的情形	151

7.7	可逆性	155
7.8	进化和生态学的马尔可夫模型	156
7.8.1	进化模型	158
7.8.2	生态学中的似然估计方法	163
7.8.3	生态学的贝叶斯进展	166
7.9	在离散偏微分方程中的应用	169
7.10	模拟退火法 (第 3.4 节的后续)	170
7.11	习题	176
第八章	排队与网络	181
8.1	排队模型 $M/M/1$	181
8.2	排队模型 $M/M/1/K$	184
8.3	排队模型 $M/M/s$	185
8.4	排队模型 $M/M/s/s$	187
8.5	机器修理	187
8.6	排队模型序列	188
8.7	排队模型 $M/G/\infty$	188
8.8	排队模型 $M/G/1$	189
8.8.1	嵌入马氏链	189
8.8.2	正常返情形	190
8.9	开放 Jackson 网络	193
8.10	闭 Jackson 网络	197
8.11	电话网络	199
8.12	Kelly 网络	202
8.12.1	单个排队队列	202
8.12.2	多类网络	205
8.13	习题	206
第九章	金融数学引论	208
9.1	基本概念	208
9.1.1	期权	209
9.1.2	套利	209
9.1.3	市场的有效性和完备性	210

9.2	离散模型的欧式期权定价	211
9.2.1	模型	211
9.2.2	可允许策略	211
9.2.3	鞅	213
9.2.4	有效完备的市场	214
9.2.5	看涨 - 看跌期权的定价	216
9.2.6	Black-Scholes 公式 (1)	217
9.3	Black-Scholes 模型与公式	219
9.3.1	随机分析引论	220
9.3.2	随机微分方程	226
9.3.3	Feynman-Kac 公式	228
9.3.4	Black-Scholes 偏微分方程	229
9.3.5	Black-Scholes 公式 (2)	231
9.3.6	Black-Scholes 模型的推广	231
9.3.7	Black-Scholes 公式 (3)	232
9.3.8	Girsanov 定理	235
9.3.9	马尔可夫性与偏微分方程	236
9.3.10	写在几个标的资产上的未定权益	238
9.3.11	完备性和有效性	240
9.3.12	有效计算的注释	241
9.3.13	历史与隐含波动率	241
9.4	美式期权的离散模型	242
9.4.1	Snell 包络	243
9.4.2	Doob 分解	245
9.4.3	Snell 包络和马氏链	246
9.4.4	回到美式期权	247
9.4.5	美式和欧式期权	247
9.4.6	美式期权和马尔可夫模型	248
9.5	在 Black-Scholes 模型下的美式期权	249
9.6	利率与债券	250
9.6.1	利率期货	250
9.6.2	利率和债券的期货	251
9.6.3	债券期权	253
9.6.4	一个利率模型	254
9.7	习题	256

第十章 部分习题解答	259
10.1 第一章习题解答	259
10.2 第二章习题解答	264
10.3 第三章习题解答	277
10.4 第四章习题解答	279
10.5 第五章习题解答	280
10.6 第六章习题解答	281
10.7 第七章习题解答	284
10.8 第八章习题解答	290
10.9 第九章习题解答	292
参考文献	295
索引	298

第一章 蒙特卡罗方法与模拟

引言

为了给出蒙特卡罗 (Monte Carlo) 方法的基本想法, 我们首先考虑一个积分的数值计算问题. 为了计算如下积分:

$$\int_{[0,1]} f(x)dx$$

基于 $\sum_{i=1}^n w_i f(x_i)$ 这种类型的公式, 我们可以给出许多不同的逼近算法, 这里 w_i 是正数, 其和为 1, 且 x_i 是区间 $[0, 1]$ 中的点. 例如, 当 $w_i = 1/n$ 时, 点 $x_i = i/n$ 即为区间的均分点, 我们得到经典的计算方法 (梯形算法). 当然还有许多不同算法, 比如我们熟知的高斯算法和辛普森 (Simpson) 算法. 蒙特卡罗方法也是类似的, 它的计算方法是: 定义 $w_i = 1/n$, 然后我们“随机”选择 x_i (但不是完全随意的选取, 所选取的点需要服从 $[0, 1]$ 上的均匀分布). 下面我们将会看到, 算法收敛性由分布的大数定律得到, 而 $1/\sqrt{n}$ 阶收敛速度由中心极限定理给出. 显然, 这个收敛速度是比较慢的, 在 1 维情形下与别的积分算法比较显得并没有优势. 但是, 其他的方法在维数增加的时候都变得不适用了. 事实上, 结果的精确度可写为离散化后的相邻两点的距离的函数. 如果我们用 n 个点来离散化区域 $[0, 1]^d$, 相邻两点的距离为 $n^{-1/d}$ 阶的, 于是为了在区域 $[0, 1]^d$ 中利用 1 阶方法得到接近于 $1/n$ 阶的精确率, 我们需要将区域离散为大约 n^d 块. 另一方面蒙特卡罗方法对于维数的变化并不敏感.

在历史上, 这个方法最初是由布丰公爵 (Buffon) 在 1777 年研究 π 的逼近

算法时提出的. 不过蒙特卡罗方法的真正产生是与电子计算机的出现相联系的. 第一篇描述这个方法的文章可以追溯到 20 世纪 40 到 50 年代. 之后关于这一方法的研究也越来越多, 这很大程度上是由于这一算法实现简单, 即可编程性很强, 以及利用现代计算机可以在很短的时间内生成大规模的随机数.

1.1 方法的介绍

为了应用蒙特卡罗方法, 我们首先将要计算的数值写为某个随机变量的期望的形式. 通常这一步比较简单, 例如计算定积分, 但是我们也可以考虑更复杂的情形, 比如我们也可以利用它求解椭圆或抛物型偏微分方程 (参见 7.9 节和 9.3 节).

下一步, 我们就要计算写成 $\mathbb{E}(X)$ 形式的数值, 这里 X 是一个随机变量. 为了计算 $\mathbb{E}(X)$, 需要了解如何模拟服从 X 的分布的独立随机变量 X_1, \dots, X_N . 剩下的就是用如下的公式来逼近 $\mathbb{E}(X)$:

$$\mathbb{E}(X) \approx \frac{1}{N}(X_1 + \dots + X_N).$$

我们给出一个利用蒙特卡罗方法计算的例子, 即计算一个定积分. 具体来说, 我们的计算分为两步: 将问题写为随机变量的期望的形式和如何对随机变量进行模拟.

假设我们要计算如下形式的定积分:

$$I = \int_{[0,1]^d} f(u_1, \dots, u_d) du_1 \cdots du_d.$$

假设 $X = f(U_1, \dots, U_d)$, 这里随机变量 U_1, \dots, U_d 是独立同分布的, 且每一个都服从 $\mathcal{U}(0, 1)$, 我们有:

$$\mathbb{E}(X) = \mathbb{E}(f(U_1, \dots, U_d)) = \int_{[0,1]^d} f(u_1, \dots, u_d) du_1 \cdots du_d.$$

这样我们完成了第一步 (将积分写成期望的形式).

现在来看第二步随机模拟, 首先产生一系列服从 $[0, 1]$ 上的均匀分布的独立随机变量 $(U_i, i \geq 1)$, 并定义 $X_1 = f(U_1, \dots, U_d)$, $X_2 = f(U_{d+1}, \dots, U_{2d})$ 等. 于是序列 $(X_i, i \geq 1)$ 为一列独立同分布随机变量, 都与 X 服从相同的分布. 下面我们就可以应用蒙特卡罗方法.

需要强调的一点是, 这个方法很容易用程序来实现. 并且这个方法不依赖于 f 的光滑性, 只要满足可测性就可以.

我们通常要计算更一般的积分, 即 \mathbb{R}^n 中的积分, 它有如下形式:

$$I = \int_{\mathbb{R}^n} g(x)f(x)dx = \int_{\mathbb{R}^n} g(x_1, \dots, x_n)f(x_1, \dots, x_n)dx_1 \cdots dx_n,$$

这里 $f(x)$ 为非负函数且 $\int f(x)dx = 1$. 如果 X 为一个取值于 \mathbb{R}^n 的服从分布 $f(x)dx$ 的随机变量, 那么 I 就等于 $\mathbb{E}(g(X))$, 下面的问题是如何模拟服从所要求分布的随机变量. 关于这个问题, 对于一些常用的分布我们会在下面的 1.3 节中给出回答.

但是我们要首先回答下面两个问题:

- 这个算法如何收敛? 为什么?
- 算法的精度如何估计?

1.2 收敛性定理

应用概率论中的两个重要的定律可以给出这两个问题的答案. 强大数定律给出了算法的收敛性, 而中心极限定理可以使我们得到具体的收敛速率.

强大数定律和蒙特卡罗方法

定理 1.1 假设 $(X_n, n \geq 1)$ 为一列与 X 服从相同分布的独立的随机变量. 如果 $\mathbb{E}(|X|) < +\infty$, 那么, 对于几乎所有的 ω (这意味着存在一个集合 $N \subset \Omega$, 满足 $\mathbb{P}(N) = 0$, 并且此时 $\omega \notin N$):

$$\mathbb{E}(X) = \lim_{n \rightarrow +\infty} \frac{1}{n}(X_1 + \cdots + X_n)(\omega).$$

这个定理引出了蒙特卡罗方法的极限定理: 不过我们只能将它应用于可积的随机变量 (这并不是一个很强的条件).

中心极限定理和蒙特卡罗方法 为了对所研究的问题有一个思路, 需要对误差进行估值:

$$\varepsilon_n = \mathbb{E}(X) - \frac{1}{n}(X_1 + \cdots + X_n).$$

中心极限定理给出了误差 ε_n 的渐近估计行为, 它仍是一个随机变量. 事实上 ε_n 的分布趋向于一个标准高斯分布.

定理 1.2 假设 $(X_n, n \geq 1)$ 为一列独立的随机变量并且与随机变量 X 服从相同的分布. 我们假设 $\mathbb{E}(X^2) < +\infty$. 记 σ^2 为 X 的方差:

$$\sigma^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \mathbb{E}((X - \mathbb{E}(X))^2),$$

那么:

$$\frac{\sqrt{n}}{\sigma}\varepsilon_n \text{ 依分布收敛于标准正态随机变量 } Z.$$

换句话说, 对于 $a < b$, 有

$$\lim_{n \rightarrow +\infty} \mathbb{P}\left(\frac{\sigma}{\sqrt{n}}a \leq \varepsilon_n \leq \frac{\sigma}{\sqrt{n}}b\right) = \int_a^b e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}}.$$

在实际计算中, 如果 n 不是太小 (这也是应用蒙特卡罗方法所必需的), 上面的概率可以用它的极限代替, 于是我们可以将 ε_n 近似看做零均值方差 σ^2/n 的高斯随机变量.

注 1.1 (1) 这是一个很强的正面结果, 因为它给出算法的收敛速度实际上可以由模拟的次数所控制. 并且它可以给出可靠的误差区间, 也就是说我们的估计有一个可控制的误差范围, 而不需要更多的计算, 这也正是这个方法有效的地方.

(2) 不过中心极限定理并无法给出误差的界限, 因为高斯随机变量的取值空间为整个 \mathbb{R} . 当我们讨论蒙特卡罗方法的估计误差时, 或者给出 ε_n 的标准差, 即 σ/\sqrt{n} , 或者给出结果的 95% 的置信区间. 这说明结果有 95% 的可能性出现在给定的区间内 (但仍有 5% 的可能性不在这个区间内, 并不能完全确定!). 当然, 数值 95% 可以用其他更接近 1 的数值来代替.

考虑到随机变量 X 的方差在误差估计中的重要性. 我们可以考虑选择 X 的不同的分布函数使其保持期望 $\mathbb{E}(X)$ 不变 (这是我们最感兴趣的), 同时我们希望将随机变量 X 替换为另一个期望相同但是方差更小的随机变量. 这一技巧被称为方差减小方法 (参见 1.4 节).

需要注意到误差收敛到 0 的速度并不是很快. 但是在某些情况下, 这个方法是我们能够运用的最好的方法 (积分问题或维数超过 4 的抛物型偏微分方程等). 同时需要指出的是, 当应用这一方法来计算积分时, 收敛速度不依赖于 f 的光滑性.

下面我们给出应用中心极限定理估计蒙特卡罗方法收敛速度的两个例子. 我们可以看到由这一结果如何给出蒙特卡罗方法的实用局限.

好的例子 我们考虑计算 $p = \mathbb{P}(f(U) \leq \lambda)$, 这里 U 为服从任意分布的随机变量. 假设 $X = \mathbf{1}_{\{f(U) \leq \lambda\}}$. 于是 $\mathbb{E}(X) = p$, 且 $\sigma^2 = \text{Var}(X) = p(1-p)$. 那么, 对于服从 X 的分布的 n 次独立采样样本 X_1, \dots, X_n , 我们有:

$$p_n = \frac{X_1 + \dots + X_n}{n} \approx p + \frac{\sigma}{\sqrt{n}}Z.$$

注意到 $p(1-p) \leq 1/4$, 如果我们想把值 p 的误差的标准差 $\frac{\sigma}{\sqrt{n}}$ 控制在 0.01 的范

围内, 那么 n 的取值将不少于 2500. 如果我们取 $n = 2500$, 利用中心极限定理, 我们得到对于 p 的值的 95% 的置信区间就是 $[p_n - 1.96 \times 0.01, p_n + 1.96 \times 0.01]$. 如果 p 的真实值的估计为 0.50 阶左右, 这给出一个可以接受的误差范围.

另一方面, 当需要估计的值 p 很小时, 前面的采样数 n 可能就不足以估计模拟的误差的阶数 (因为需要误差的阶的大小低于所要估计的数值), 我们必须 (直观上显然) 取采样数目多于 $1/p$.

难的例子 假设我们要计算 $\mathbb{E}(\exp(\beta Z))$, 这里 Z 为服从标准正态分布的随机变量. 我们容易验证:

$$E = \mathbb{E}(e^{\beta Z}) = e^{\frac{\beta^2}{2}}.$$

如果对于这个问题应用蒙特卡罗方法, 我们假设 $X = e^{\beta Z}$. X 的方差是 $\sigma^2 = e^{2\beta^2} - e^{\beta^2}$. 考虑服从 X 的分布的 n 次采样 X_1, \dots, X_n , 我们有:

$$E_n = \frac{X_1 + \dots + X_n}{n} \approx E + \frac{\sigma}{\sqrt{n}}Z.$$

相对估计误差的标准差是 $\frac{\sigma}{E\sqrt{n}} = \sqrt{\frac{e^{\beta^2} - 1}{n}}$. 如果我们想要保证误差小于一个给定的 $\varepsilon > 0$, 那么需要选定 $n \approx \varepsilon^{-2}(e^{\beta^2} - 1)$. 如果 $\varepsilon = 1$ 且 $\beta = 5$, 那么有 $n = 7 \times 10^{10}$, 这是很多次计算 (确切地说太大了!). 例如, 当 $\beta = 5$ 时, 计算次数达到 27×10^4 的阶次. 如果我们取样本 10^5 次, 我们就接近了估计值 85×10^4 , 计算结果的 95% 置信区间为 $[-467647, 2176181]$. 这太糟糕了. 我们唯一可以确定的是这个方法不适用.

这个例子说明当要用蒙特卡罗方法估计的随机变量的方差很大时, 实际计算的极限满足下面的规则:

规则: 当应用蒙特卡罗方法估计随机变量的期望时, 我们要系统地估计所用的模拟方法给出的方差.

这个例子给出了蒙特卡罗方法的实用方面的局限性, 当随机变量的方差很大时, 蒙特卡罗方法并不适用. 我们必须重新调整模拟, 以得到想要的随机变量的估计. 这里还要指出的是方差减小方法可以保证蒙特卡罗计算的有效性. 我们将在 1.4 节中讨论这个问题.

1.3 随机变量的模拟

对于服从 $[0, 1]$ 上均匀分布的随机变量的模拟 所有计算机程序中都具有一个伪随机数生成器. 它可以生成一系列确定的数, 带有一些周期性, 但是它们的统计性质带有一些独立的 $[0, 1]$ 上均匀分布随机变量的特征. 构造一个好的“随机

数生成器”的困难在于如何构造一个递归函数使得在合理的时间里可以由它生成一个数列,使其在统计上看做一系列独立的 $[0, 1]$ 上均匀分布的随机变量,并且其周期越大越好. 这类随机数生成器的研究主要基于动力系统. 我们可以在 [33] 和 [45] 中找到伪随机数发生器的许多经典算法. 最近, Matsumoto 和 Nishimura [36] 提出的一个伪随机数生成器的周期为 $2^{19937} - 1!$

事实上,所有的随机数生成器实际上是在 $\left[1, \frac{1}{M}, \frac{2}{M}, \dots, \frac{M-1}{M}, 1\right]$ 上的均匀分布中取一个采样值,这里 M 为一个很大的数.

下面我们考虑如何基于服从均匀分布的随机变量的模拟来得到服从其他分布的随机变量模拟.

模拟一个 Bernoulli 分布的随机变量: 设 $0 < p < 1$, 如果 U 为服从 $U(0, 1)$ 的随机变量,那么 $X = 1_{U \leq p}$ 为服从参数为 p 的 Bernoulli 分布的随机变量.

模拟一个二叉树分布的随机变量: 如果 U_1, \dots, U_n 为服从 $U(0, 1)$ 的独立随机变量,那么 $X = 1_{U_1 \leq p} + 1_{U_2 \leq p} + \dots + 1_{U_n \leq p}$ 为服从 $B(n, p)$ 的随机变量 (参数为 n 和 p).

模拟一个几何分布的随机变量: $X = \inf\{k \geq 1; U_k \leq p\}$ 就是一个服从参数为 p 的几何分布的随机变量. 基于下面的引理,我们有一个更有效的算法,这将在习题 5.1 中讨论.

分布函数的逆函数 我们回顾一个经典的结果:

引理 1.1 假设 X 为一个随机变量, F 为其分布函数 (记 $F(x) = \mathbb{P}(X \leq x)$). 这里,对于 $0 \leq t \leq 1$, 定义

$$F^{-1}(t) = \inf\{x; F(x) \geq t\}.$$

如果 U 服从分布 $U[0, 1]$, 那么 $F^{-1}(U)$ 与 X 有相同的分布.

证明 我们利用定义 $F(x) = \mathbb{P}(X \leq x)$, 立刻得到

$$\mathbb{P}(F^{-1}(U) \leq x) = \mathbb{P}(U \leq F(x)) = F(x).$$

这个方法适用于知道 F 及 F 的逆函数的显式形式的情况. 特别的,我们来看指数分布的情形.

指数分布的模拟 回忆一下一个随机变量 X 服从参数为 λ 的指数分布, 当且仅当, 对于所有的 $t \in \mathbb{R}_+$ 有

$$\mathbb{P}(X > t) = \exp(-\lambda t).$$

假设 F 为 X 的分布函数, $F(t) = 1 - e^{-\lambda t}$, 且

$$F^{-1}(x) = -\frac{\log(1-x)}{\lambda}.$$