Monte Carlo 方法与计算机模拟应用研究

魏丽英[1]，陈晓鹏[2]
(1.桂林电子科技大学 数学与计算科学学院，广西 桂林 541004；
2.山西省潞安集团 地质处，山西 长治 046204)

摘 要：介绍了 Monte Carlo 方法，提出其在模拟 Buffer 问题时存在的一个问题，并给出改进的方法；提出了用 Monte Carlo 方法产生任意分布随机变量的原理及方法，并对 Beta 分布和标准正态分布随机变量进行了计算机模拟和效果检验。

关键词：Monte Carlo 方法；Buffer 问题；随机变量模拟

中图分类号：TP39 **文献标识码**：A **文章编号**：1674-7720(2011)12-0071-03

Research of Monte Carlo method and applications of computer simulation

Wei Liying[1]，Chen Xiaopeng[2]
(1.School of Mathematics & Computational Science，Guilin University of Electronic Technology，Guilin 541004，China；
2.Geological Department，Lu An Group，Changzhi 046204，China)

Abstract：This paper gives an introduction to the Monte Carlo method，and a question be pointed for simulation of Buffer question using Monte Carlo Method，also presents the way to improve；This paper also gives the theory and method of generation arbitrary distributed random variables using Monte Carlo method，the random variables of Beta distribution and standard normal distribution are carried out computer simulation，and test results.

Key words：Monte Carlo method；Buffer question；random variable simulation

模拟是指把某一现实的或抽象的系统的状态和特征，用另一个称为模型的系统来代替或模仿。电子计算机的出现及计算机科学技术的迅猛发展，给许多学科带来了巨大的影响。计算机不但使问题的求解变得更加方便、快捷和精确，而且使得解决实际问题的领域更加广泛。计算机适合于解决规模大、难以解析化以及不确定的数学模型，为系统模拟提供了极有力的工具。计算机模拟就是用计算机程序在计算机上模仿各种实际系统的运行过程，并通过计算了解系统随时间变化的行为或特性。

Monte Carlo 方法，即随机模拟法，是用计算机模拟随机现象，通过仿真试验，得到实验数据，再进行分析推断，得到某些现象的规律或某些问题的求解的方法，它是计算机模拟的基础，它对解决随机性问题具有很强的能力。Monte Carlo 方法的基本思想是：首先建立一个概率模型，使所求问题的解正好是该模型的参数或其他有关的特征量；然后通过模拟统计，即多次随机抽样实验，统计出某事件发生的频率。只要实验次数足够多，该频率便近似于事件发生的概率，这实际上就是概率的统计定义[1]。

Monte Carlo 方法属于试验数学的一个分支，它研究的问题大致可分为两种类型，一种是问题本身是随机的；另一种是本身属于确定性问题，但可以建立它的解与特定随机变量或随机过程的数字特征或分布函数之间的联系，因而也可用随机模拟方法解决。本文旨在给出对 Monte Carlo 方法应用的一点思考，只涉及前一种问题。

1 Monte Carlo 方法与 Buffer 问题

Monte Carlo 方法源于 1777 年法国科学家 Buffer 提出的一种计算圆周率 π 的方法——随机投针法，即著名的 Buffer 投针问题，其步骤为[1]：

(1)在纸上画出间距为 d 的平行直线；

(2)取长度为 $t(t<d)$ 针，随机向画有平行直线的纸上投掷 n 次，其中针与直线相交的次数记为 m；

(3)由统计实验估计出针与直线相交的概率 $P\approx\frac{m}{n}$；

(4)计算出 π 的近似值。

分析可知，针与直线相交的充要条件为 $x\leqslant\frac{t}{2}\sin\theta$，

如图 1 所示，其中 x 为针的中点与最近一条直线的距离。

可令 $x \sim U(0, d/2)$，$\theta \sim U(0, \pi)$，建立直角坐标系 (θ, x)，如图 2 所示，由几何概率可知：

$$P=\frac{\int_0^{\pi}\frac{t}{2}\sin\theta d\theta}{\frac{d}{2}\pi}=\frac{2t}{d\pi}$$，由此可知 π 的估计式为：

$$\pi \approx \frac{2tn}{dm}$$。

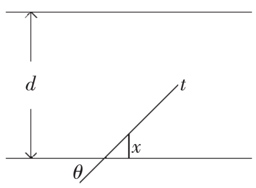

图 1 Buffer 问题简图 1

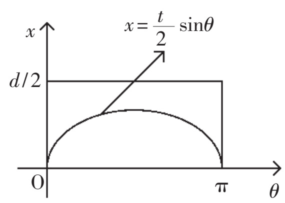

图 2 Buffer 问题简图 2

这里存在一个问题：算法的目的是计算 π 的值，而由此产生的模拟步骤中显然已利用了 π 的值(产生随机变量 θ)，这是矛盾的。笔者认为可直接将 $\sin\theta$ 当作一个随机变量，考虑用舍选抽样法来直接产生 $\sin\theta$，其中 $\theta \sim (0, \pi)$，其原理如下：

如图 3 所示，$\theta \sim U(0, \pi)$，在半圆周 $D=\{(x, y)|x^2+y^2=1, y>0\}$上的每一个随机点与 θ 一一对应，故随机变量的抽样可由随机变量 $Y=(x, y) \in D$ 的抽样代替，而随机变量 $\sin\theta$ 的抽样可由 Y 中分量 y 的抽样代替，这是因为此时 $\sin\theta=\frac{y}{\sqrt{x^2+y^2}}=y$。

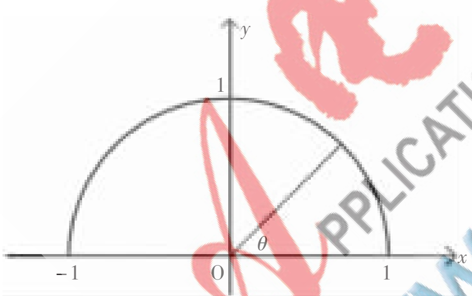

图 3 随机变量 θ 的分布

改进后模拟步骤如下：

(1)给定初值 t, d, n；

(2)产生相互独立的均匀随机数列 x_1, x_2，其中 $x_1 \sim U(-1, 1)$，$x_2 \sim U(0, 1)$，$i=i+1$；

(3)检验条件 $x_1^2+x_2^2=1$ 是否成立(因误差原因具体模拟时可将条件变为 $1-\varepsilon \leqslant x_1^2+x_2^2 \leqslant 1+\varepsilon$，其中 ε 为很小的正数)，若成立，则令 $\sin\theta=x_2$，$k=k+1$，否则回到步骤(2)；

(4)产生均匀随机数 $y \sim U(0, d/2)$，检验条件 $y \leqslant \frac{1}{2}tx_2$ 是否成立，若成立，则 $m=m+1$，否则回到步骤(2)；

(5)检验条件 $i<n$ 是否成立，若成立，则回到步骤(2)，否则计算 π 的估计值 $\pi=\frac{2tk}{dm}$。

用 Matlab 编程实现，取 $\varepsilon=0.01$ 时的统计结果如表 1 所示。

表 1 计算机模拟统计结果

t	d	n	π
1	2	50 000	3.188 5
1	2	100 000	3.144 7
2	3	50 000	3.148 8
2	3	100 000	3.165 9
2	4	50 000	3.169 6
2	4	100 000	3.163 9

从表 1 可看出，运算精度与参数选取有很大关系，一个启示是实验模拟时要精心设置参数，并多次实验提高精度。

2 Monte Carlo 方法模拟随机变量

Monte Carlo 方法可用来产生具有一定分布特征的随机数。

引理[2]：设 ξ、η 是两个独立的随机变量，ξ 在(a, b)中均匀分布，而 η 在(0, 1)中均匀分布；又设 $f(x)$是集中在(a, b)中的密度函数(即满足 $\int_a^b f(x)dx=1$ 的密度函数)；任取正常数 m，使 $mf(x)<1$ 对一切 $x \in (a, b)$成立，则有：

$$P(\xi \leqslant d|mf(\xi) \geqslant \eta)=\int_a^d f(x)dx, (a<d<b)$$

这就是说，在事件 $mf(\xi) \geqslant \eta$ 下，ξ 的条件密度函数为 $f(x)$。

引理启示构造密度函数为 $f(x)$的随机变量的方法：先取两列随机变量 $u_i \sim U(a, b)$，$v_i \sim U(0, 1)$，如果 $mf(u_i) \geqslant v_i$，则令 $x_j=u_i$ 作为第 j 个具有密度函数 $f(x)$的随机变量，否则舍弃(u_i, v_i)，再考察下一对，如此反复。

引理中假定了(a, b)是在有限区间，实际应用中若不是有限区间，总可选取有限区间(a, b)，使 $\int_a^b f(x)dx \geqslant 1-\varepsilon$，其中 ε 是充分小的正数，然后在(a, b)中运用上述方法(当然要在误差许可的范围内，否则不行)。现举例实验如下。

2.1 Beta 分布

Beta 分布[3]的概率密度函数为

$$f(x)=\frac{1}{B(\alpha, \beta)}x^{\alpha-1}(1-x)^{\beta-1}, (0 \leqslant x \leqslant 1, \alpha, \beta>0)$$

式中 $B(\alpha, \beta)=\int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt$

由微积分知识可知 $f(x)$在[0, 1]上有最大值点 $x_0=\frac{\alpha-1}{\alpha+\beta-2}$，故可取 $m=\frac{1}{f(x_0)}$，则有 $mf(x)<1$(除 x_0 点外)成立，于是产生具有 Beta 分布的随机变量的算法如下：

(1)确定 α、β，计算 $m=\frac{1}{f(\frac{\alpha-1}{\alpha+\beta-2})}$；

(2)产生随机变量 $u_i \sim U(0, 1)$，$v_i \sim U(0, 1)$；

(3)检验 $mf(u_i) \geqslant v_i$ 是否成立，若成立，则令 $x_j=u_i$，$j=j+$

1，否则转到步骤(2)，直到 $i>n$(n 为预先给定的实验次数)。

用 Matlab 编程实现，取 $\alpha=2,\beta=3,n=100$，结果如表 2 所示。

经检验其样本均值为 0.400 0，样本方差为 0.043 0，而此 Beta 分布的期望是 0.4，方差是 0.04，可见其模拟程度很好。

2.2 标准正态分布

标准正态分布的概率密度函数为：

$$f(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2},x\in \mathbf{R}$$

$f(x)$ 有最大值 $f(0)=\frac{1}{\sqrt{2\pi}}$，故可取 $m=\frac{1}{f(0)}=\sqrt{2\pi}$；标准正态分布中 $x\in \mathbf{R}$，是无限区间，而事实是：若 $\xi\sim N(0,1)$，则 $P(-3\leqslant\xi\leqslant 3)=\int_{-3}^{3}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx=0.997\ 4$，这称为正态分布的“$3\sigma$”原理。故可令 $(a,b)=(-3,3)$，于是产生标准正态分布随机变量的算法如下：

(1)产生随机变量 $u_i\sim U(-3,3)$，$v_i\sim U(0,1)$；

(2)检验 $mf(u_i)\geqslant v_i$ 是否成立，若成立则令 $x_j=u_i$，$j=j+1$，否则转到步骤(1)，直到 $i>n$(n 为预先给定的实验次数)。

用 Matlab 编程实现，取 $n=100$，结果如表 3 所示。

经检验其样本均值为 0.001 6，样本方差为 0.914 5，可见其模拟程度较高。由标准正态随机变量，进行变换 $Y=\mu+\sigma X$，可知 $Y\sim N(\mu,\sigma^2)$，可得一般正态分布随机变量。

Monte Carlo 方法是通过设计适当的随机试验而完成某种计算任务的方法，它对解决随机性问题具有很强的能力。随着电子计算机的高速发展，Monte Carlo 方法在自然科学及社会科学领域的应用越来越广泛，许多用确定性方法难以解决的随机性问题都可以用它方便地解决，特别是在随机模拟方面更显示了其独特的魅力。Buffer 问题反映了蒙特卡罗方法的思想，通过设计适当的随机试验来完成某种计算任务。但在利用计算机模拟随机试验时出现矛盾是由于 Buffer 问题本身不涉及常数 π 值，在利用计算机模拟试验中没有回避这一问题而造成的，因此应该设法避免[4]。

表 2 Beta 分布随机变量

0.687 9	0.777 5	0.361 1	0.333 2	0.021 3	0.143 6	0.728 9	0.404 2
0.282 0	0.142 4	0.274 8	0.430 5	0.107 3	0.773 4	0.206 9	0.541 0
0.409 1	0.449 0	0.572 4	0.288 7	0.121 2	0.423 2	0.693 5	0.491 0
0.307 3	0.093 5	0.433 1	0.617 3	0.348 2	0.304 6	0.241 6	0.381 2
0.599 3	0.641 8	0.423 4	0.489 9	0.536 9	0.130 1	0.154 5	0.392 2
0.592 1	0.294 7	0.402 1	0.152 7	0.123 4	0.498 2	0.532 5	0.753 1
0.047 3	0.740 2	0.608 1	0.445 5	0.250 9			

表 3 标准正态随机变量

1.794 0	1.872 0	−0.760 6	0.142 6	−0.117 2	1.076 5	−1.211 9	0.435 5
0.112 6	−0.983 9	−1.716 8	−1.098 8	0.778 0	−1.066 1	1.127 2	−1.218 9
−0.159 7	−0.325 3	−0.993 1	−0.563 6	0.764 4	−0.737 1	0.378 5	0.640 1
0.740 4	0.912 5	−0.105 5	1.037 7	−0.006 9	0.614 8	−1.128 9	0.941 1
0.825 1	−0.370 2	−1.175 5	1.522 0	−0.702 9	−1.346 8	0.137 8	

参考文献

[1] 重庆大学数学系.数学实验[M].北京：科学出版社，2000.

[2] 王梓坤.概率论基础及其应用[M].北京：北京师范大学出版社，2007.

[3] 杨肇夏.计算机模拟及其应用[M].北京：中国铁道出版社，1999.

[4] 左国新，陈应保.蒲丰问题与蒙特卡罗方法[J].高等函授学报(自然科学版)，2005，18(2)：3-4.

(收稿日期：2010-12-31)

作者简介：

魏丽英，女，1981 年生，硕士研究生，主要研究方向：应用数学。