蒙特卡罗方法和事例产生器

平荣刚

中国科学院高能物理研究所
(pingrg@ihep.ac.cn)

北京谱仪十一培训学校，2023年9月29-10月6日，内蒙古大学

目 录

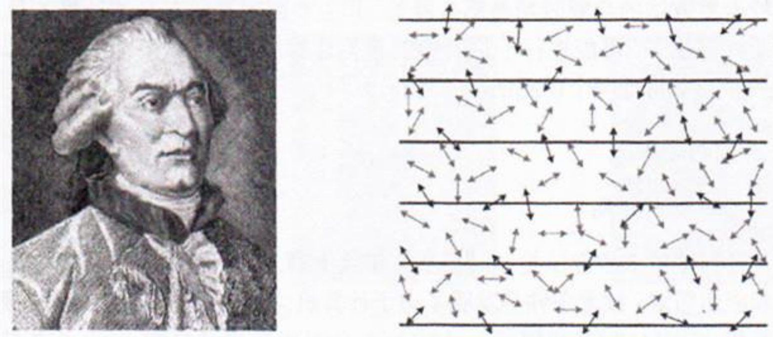
1.1 蒙特卡罗方法简史

蒙特卡罗方法又称随机抽样技巧或统计试验方法。半个多世纪以来，由于科学技术的发展和电子计算机的发明，这种方法作为一种独立的方法被提出来，并首先在核武器的试验与研制中得到了应用。蒙特卡罗方法是一种计算方法，但与一般数值计算方法有很大区别。它是以概率统计理论为基础的一种方法。由于蒙特卡罗方法能够比较逼真地描述事物的特点及物理实验过程，解决一些数值方法难以解决的问题，因而该方法的应用领域日趋广泛。

- **启蒙时期(18世纪---第二次世界二战)**
- **开创时期（第二次世界大战期间。）**
- **发展时期（前期…1980…后期）**

- 启蒙时期

1.蒲丰氏(Buffon)问题



(a) C.D.Buffon(1707~1788)　　(b) 蒲丰随机投针实验

为了求得圆周率π值，在十九世纪后期，有很多人作了这样的试验：将长为$2l$的一根针任意投到地面上，用针与一组相间距离为$2a$（$l<a$）的平行线相交的频率代替概率P，再利用准确的关系式：

$$P=\frac{2l}{\pi a}$$

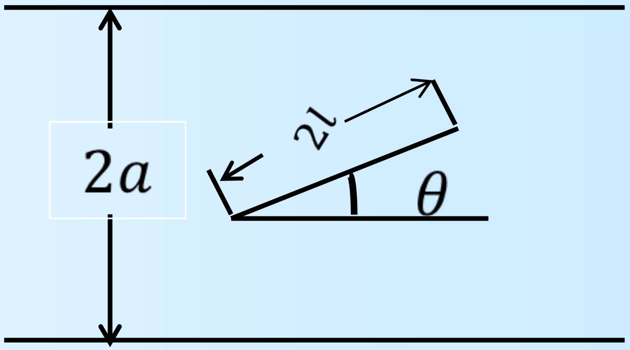
求出π值

$$\pi=\frac{2l}{aP}\approx\frac{2l}{a}\left(\frac{N}{n}\right)$$

其中N为投计次数，n为针与平行线相交次数。这就是古典概率论中著名的蒲丰氏问题。

关于投针实验结果的推导

针与线相交的几率为



$$f(\theta) = \left|\frac{2l\ sin\theta}{2a}\right|$$

假设θ在$[0 \sim \pi]$上分布是均匀的， 密度函数为$\frac{1}{\pi}$, 经过大量实验后，针与水平线相交的概率为

$$P = \int_0^{\pi} \frac{l}{a} |\sin\theta| \frac{d\theta}{\pi} = \int_0^{\pi} \frac{l}{\pi a} \sin\theta\, d\theta = \frac{2l}{\pi a}$$

可以证明，用投针实验测量π值，在进行n次后，测量π值得标准误差为：

$$\delta\pi \approx \frac{2.37}{\sqrt{n}}$$

投针实验的科学意义

1. 用投针实验测量π值，这是一条与传统方法不同的新途径。
2. 开启了用随机性事例的模拟方法，但方差难以减小，使得这种方法难以发展。

投针次数n	方差$\delta\pi$
1,000	0.075
2,000	0.053
3,000	0.043
4,000	0.037
5,000	0.033
10,000	0.0075

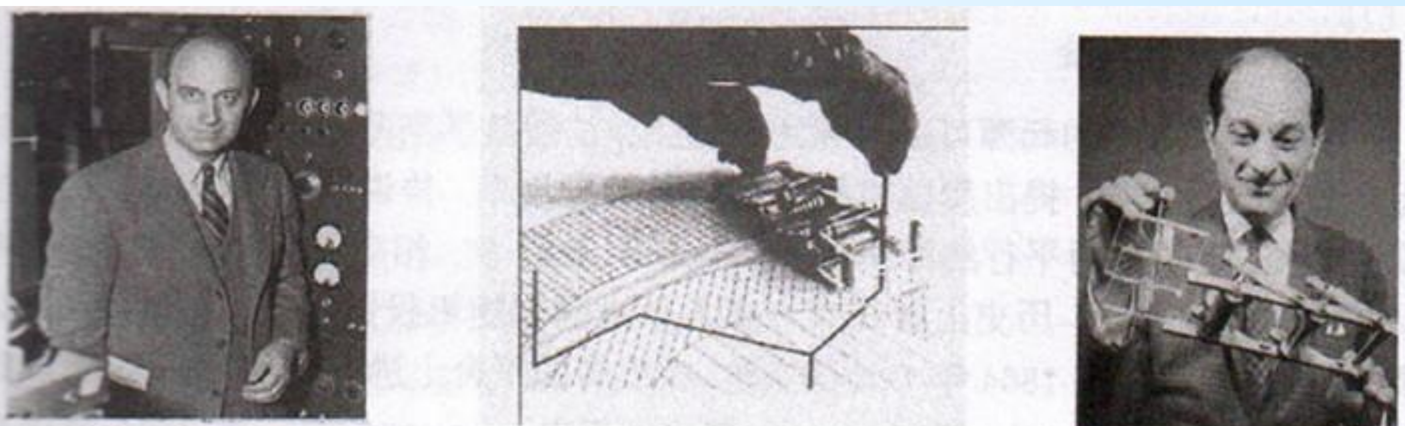
一些人进行了实验，其结果列于下表：

实验者	年份	投计次数	π的实验值
沃尔弗(Wolf)	1850	5000	3.1596
斯密思(Smith)	1855	3204	3.1553
福克斯(Fox)	1894	1120	3.1419
拉查里尼 (Lazzarini)	1901	3408	3.1415929

2. 费米模拟装置

20世纪30年代，费米发明的“FERMIAC”机械模拟实验装置，用来模拟中子在核装置内的随机扩散运动。实验时“FERMIAC”在核装置的二维平面滑动，随机地选取快中子或慢中子，确定中子运动方向和碰撞距离，得到中子随时间的变化情况，这类似蒙特卡罗方法模拟中子的随机运动。

蒙特卡罗方法的启蒙时期长达两个世纪，其主要原因是缺乏高速计算工具。没有现代的电子计算机，不可能进行千百万次的模拟计算。

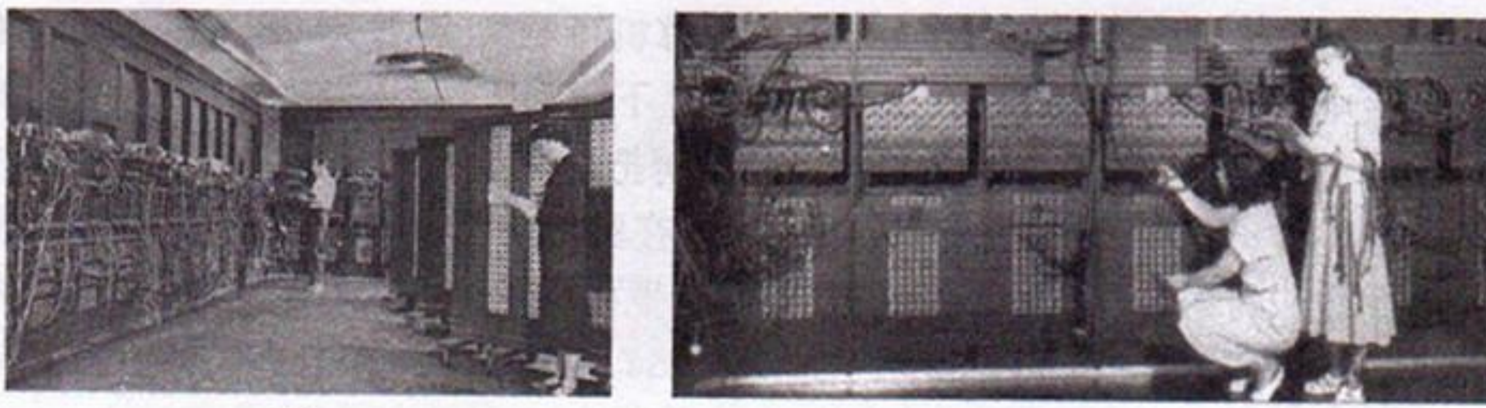


(a) E.Formi(1901~1954)　(b)　(c)

- 开创时期

1.原子弹研制时期

20世纪40年代，是蒙特卡罗的开创时期。当时出现了蒙特卡罗方法的技术条件和应用需求。技术条件是发明了电子计算机（１９４５年，ＥＮＩＡＣ）



(a) (b)

图 世界第一台计算机“ENIAC”和计算员操作

应用需求是美国制造原子弹和氢弹。1942年到1945年是美国研制原子弹时期，实施研制原子弹的曼哈顿计划，1945年美国试验第一颗原子弹。原子弹设计需要需论上计算中子在原子弹内的扩散与增值，与各种材料元素发生碰撞、产生散射、吸收和裂变，都是随机过程。3位科学家：乌拉姆(S.Ulam)、冯.诺依曼(J.von Neumann)和梅特罗波利斯(N. Metropolis)是蒙特卡罗方法的三位开创者。

(a) 蒙特卡罗城　(b) 蒙特卡罗赌场　(c) 轮盘赌

图　蒙特卡罗城、蒙特卡罗赌场和轮盘赌

(a) “MANIAC” 电子计算机　(b) N.Metropolis

图 1.7　“MANIAC” 电子计算机和 N.Metropolis

1.2 随机数产生和检验

伪随机数发展中，出现了两个事件：一是Marsaglia(1968,1972)发现线性同余法产生的随机数具有不均匀性和相关性，出现了高维随机数的降维现象，二是经典的裴波那契方法和反馈移位寄存器方法受到质疑，主要问题是产生的伪随机数序列具有相关性，出现与均匀性和独立性相矛盾的现象，并且在实际使用中出了问题（Ferrenberg,et.,al. 1992；Grassberger,1993）。

　20世纪80年代后，出现了各种伪随机数产生器和统计检验方法。如：麦森变形产生器，G.马莎格利亚提出的各种组合产生器。20世纪80年代美国MathWorks公司推出的Matlab数学软件，采用的伪随机数产生器，基本上跟上了伪随机数产生器的发展。例如7.4版本使用的麦森变形产生器，周期为10^{6001}。

1.2.1. 真随机数产生器

用物理的方法产生的随机数称为真随机数，其最大的特点是独立性和均匀性好，没有周期；其缺点是增加硬件设备和费用，使用中由于随机序列无法重复产生，因此无法进行复算。

- 噪声真随机数产生器

1974年，美国兰德公司用电子旋转轮产生真随机数，做成100万个真随机数表。

1978年，Frigerio，Clark，Tyler用a粒子放射源和高分辨率计数器产生真随机数，每秒产生1.6个真随机数。

2010年，据报道，美国ComScire量子世界公司生产出利用物理噪声源做成的随机数产生器，每秒可产生6.25万个随机数，售价895美元。

网站：http://www.random.org提供真随机数

1.2.2 早期伪随机数产生器

- 伪随机数的定义和性质

  从均匀分布U(0,1)抽样得到的简单子样称为随机数，其概率密度为：

$$f(x)=1 \quad (0\le x \le 1).$$

随机数序列$(U_1,U_2,\ldots)$具有独立同分布。随机数的一个重要性质，高维分布的均匀性。有s个随机数组成的s维空间上的点$(U_{n+1},U_{n+2},\ldots,U_{n+s})$,在s维空间单位立体$G_s$上均匀分布。对任意的$a_i$， $0\le a_i \le 1, i=1,2..,s$, $U_{n+i} \le a_i$的概率为：

$$P(U_{n+i}\le a_i, i=1,2,..s)=\prod_{i=1}^{s} a_i$$

注意区别： 概率密度函数(pdf)，概率

伪随机数的性能要求：

(1)能通过严格的统计检验

(2)产生伪随机数的算法有坚实的数学理论支撑

(3)伪随机数序列可以重复产生，不用存储在计算机内存

(4)速度快，有效，计算机内存占用小

(5)周期长，至少有10^{50}，如果需要N个随机数，则周期不小于$10N^2$。

(6)多流线产生，可以在并行机上实现。

(7)不产生0或者1的伪随机数，避免0溢出或者其他计算困难。

伪随机数产生的数学结构：

(1)定义伪随机数的状态空间S,必须是有限域。

(2)必须有初始态S_0，给定伪随机数的初始值。

(3)包含一个转换函数$S_t=f(S_{t-1})$，一般是递推式，是数学结构的主体。

(4)定义伪随机数的输出空间U，通常是整数。

(5)包含一个输出函数$U = g(St)$，把整数转变成(0,1)随机数。重复(3),(4),(5)部分，产生随机数序列。

早期伪随机数产生方法：

1. 经典的裴波那契产生器

Taussky 和Todd(1966)加同余产生器，递推公式为：

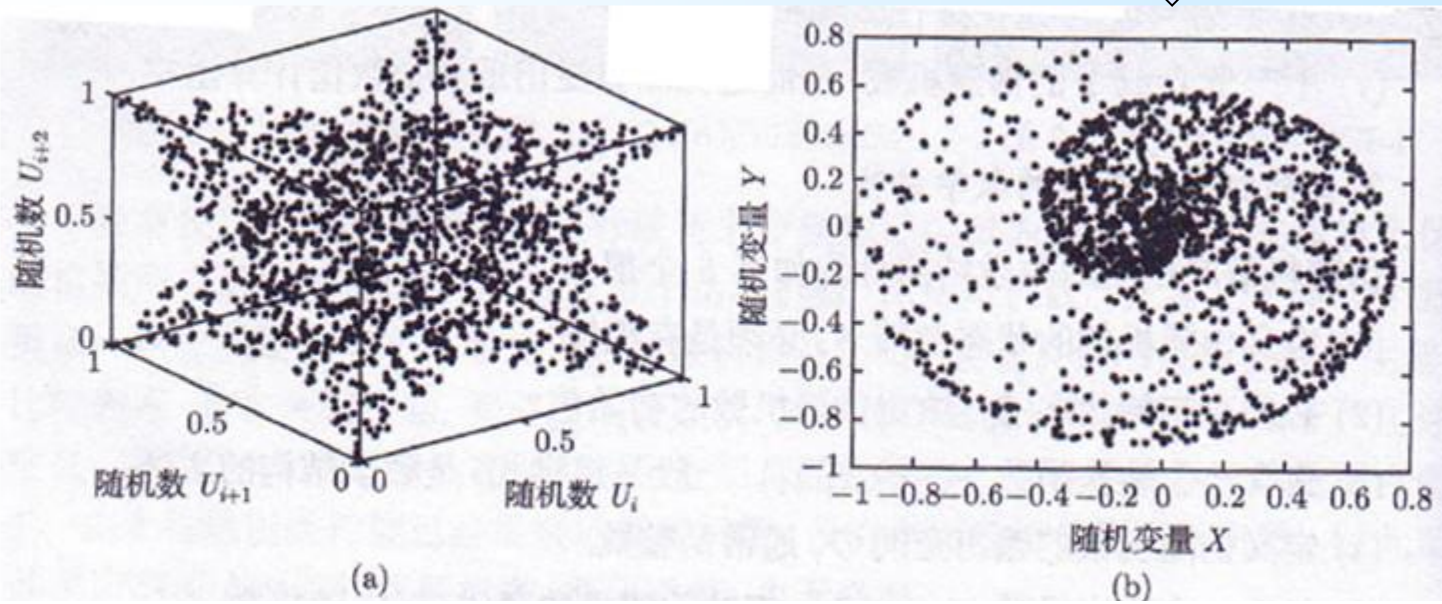
$$X_i \equiv (X_{i-2} + X_{i-1})(\mathrm{mod}\ \ \mathrm{M}),\ i > 2, \qquad U_i = \frac{X_i}{M}$$

当X_0=X_1=1时，产生的随机数系列时间点的裴波那契系列：

1,1,2,3,5,8,13,21，…..

问题：Dieter(1971),Knuth(1981)指出这种随机数序列存在不居中现象，而且产生显著的序列相关，出现随机数只能分布在3维空间的8个等边三角形平面上。

$$X_i = \sqrt{U_i}\cos(2\pi U_{i+1})\sin(\pi U_{i+2}), Y_i = \sqrt{U_i}\sin(2\pi U_{i+1})\sin(\pi U_{i+2})$$



2.反馈移位寄存器产生器

Tausworthe(1965)提出反馈移位寄存器产生器，其数学基础时本原多项式和异或运算(⊕)，递推公式为：

$$X_i \equiv (X_{i-p} \oplus X_{i-q}) \text{ 或}$$
$$X_i \equiv (X_{i-p} \oplus X_{i-p+q})$$
$$(p > q)$$

输出函数：$U_t = \sum_{t=1}^{w} X_{ts+l-1} 2^{-l}$

产生随机数系列需要p个二进制数值的初始值。产生随机数的速度快，而且周期长(2^p-1),例如，$p=521$，$q=32$，周期为6.86×10^{156}

- 1981年，Knuth对反馈移位寄存器产生器提出过异议和警告
- 1992年(Phys.Rev.Lett., 69, 3382)，美国佐治亚大学的三位物理学家(Ferrenberg,Laudau,Wong)，他们发现在伊辛模型模拟中，5个计算程序由于使用了反馈移位寄存器产生的伪随机数，存在着微妙的相关性，导致了模拟结果完全错误。
- 1993年，Grassberger发现在自回避游动模型模拟时也出现了类似的问题。

3.线性同余产生器

Lehmer（1951）年提出了乘同余产生器，Roterberg(1960)提出了混合同余产生器，线性同余产生器产生整数随机数的递推同余式为：

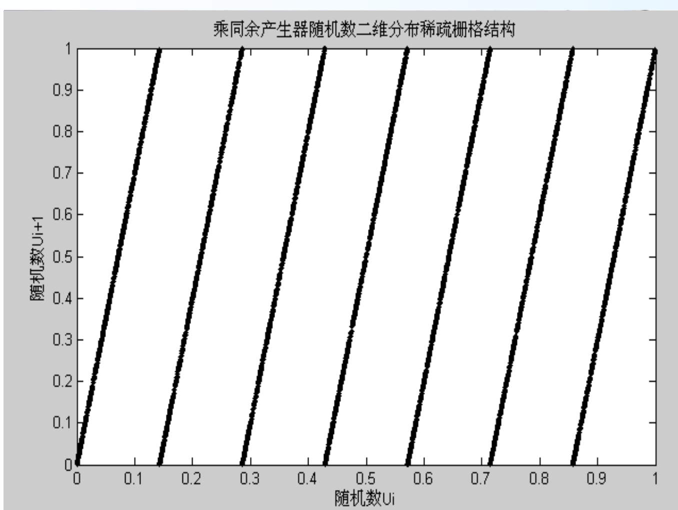
$$X_i=(A*X_{i-1}+C)\ (\mathrm{Mod}\ M),\quad U_i=X_i/M$$

式中，M为模，A为乘子，C为增量。当C=0时，是乘同余产生器，当C>0时，是混合同余产生器，线性同余产生器产生随机数序列需要一个初始值X_0, 参数(M,A,C,X_0)通过线性同余产生器的参数。

线性同余产生器的问题

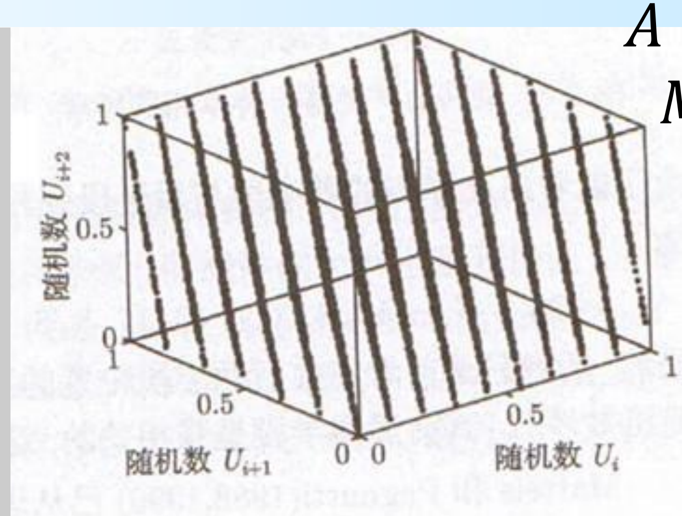
- **降维现象和稀疏栅结构**

Marsaglia(1968,1972)发现, 由线性同一产生器产生的伪随机数系 列，把其相继的s个随机数$(U_{i+1}, U_{i+2}, \dots, U_{i+s})$作为$s$为空间的一个点，这些点只分布在少数几个彼此平行的超平面上。例如，$M = 2^{32}$, s为随机数可能落在s维空间不超过$(s!\,M)^{\frac{1}{s}}$个低维的超平面上，维数$s = 1,5,10,50,100,500,1000$低维超平面的个数为$2^{32}, 220, 41, 30, 47, 193, 377$ 。 随着为数的增加，s维随机数落在少数几十个低维平面上。100位以后，只落在几百个平面上，这就是降维现象，产生稀疏栅格。



$A = 7$
$M = 2^{31} - 1$

乘同余产生器（A=7）的稀疏栅格



$A = 65539$
$M = 2^{31}$

RANDU 的稀疏栅格

- 样本的相关异常

乘同产生器1（A=7, M=2^{32}-1）和乘同产生器2（A=214748630， M=2^{32}-1 ），在(0,1)间隔 出现二重稀疏栅格结构，它们产生的随机数独立性不好。从2维独立正态分布抽样得到的样本值为：

$$X_i = \sqrt{-2\ln U_i}\cos(2\pi U_{i+1}), Y_i = \sqrt{-2\ln U_i}\sin(2\pi U_{i+1}).$$

样本点(X_i,Y_i)落在一条螺旋线上，具有很强的相关性。

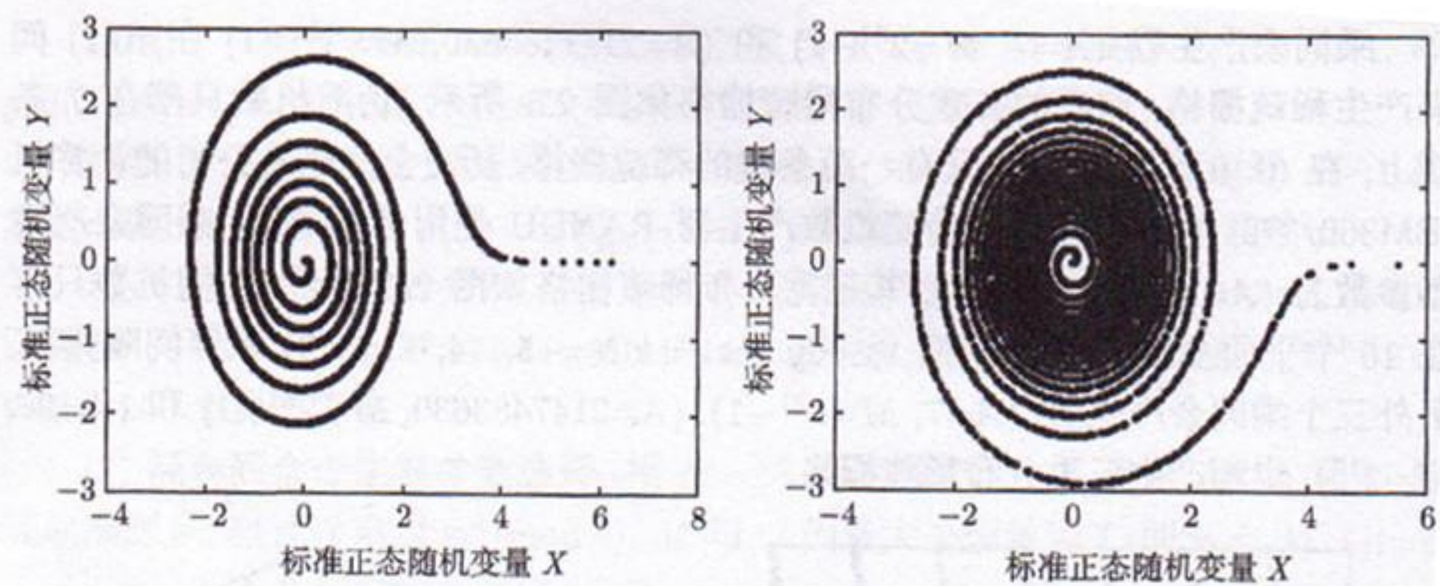


图 2.7 乘同余产生器 1 样本相关异常　　图 2.8 乘同余产生器 2 样本相关异常

- **长周期相关性**

例如：乘同余法产生器$X_i \equiv 15 * X_{i-1}\ (\text{mod } 19), X_0 = 1$

$\{X_i\}$:1,15,16,12,9,2,11,13,5,18,4,3,7,10,17,8,6,14,1

这是周期为18的系列

为了看出此系列的前后半段相关，写出相关系数为-1的另一序列：

$\{19 - X_i\}$:18,4,3,7,10,17,8,6,14,1,15,16,12,9,2,11,13,5,18

Matteis和Pagnutti(1988,1990)从理论上证明，所有的线性、非线性同余序列都存在着长周期相关现象。其他伪随机数序列也可能发生长周期相关现象。

1.2.4 随机数理论检验和统计检验

- **随机数理论检验**

  伪随机数的理论检验方法是一种事前检验方法，它是指在构造伪随机数产生器之前，选取算法和参数时，就行理论检验。为了避免出现降维现象和产生稀疏栅格，理论检验方法有相邻平行超平面之间最大距离检验（也称为谱检验），平行超平面最小数目检验和最接近点距离检验。

- **统计检验**

  不管用什么方法产生的伪随机数，它们能否作为随机数使用，最终都要靠统计检验来确定。检验的内容是随机数的参数是否与理论分布一致，随机数是否有较好的均匀性、独立性和连贯性。按实数随机数10进制数值大小的检验方法我们称为一般检验。

一般检验方法有均匀性检验（包括矩检验、频率检验和累积频率检验），相关性检验（包括相关系数检验、无重复联列表检验、有重复联列表检验和均方相继差检验），组合规律性检验（包括最小距离检验、距离检验、配套检验、扑克检验和空隙检验）。

严格检验方法的检验对象是整数随机数，整数随机数用2进制数字串表示，对于32位计算机，整数随机数用32位2进制位串表示，随机数产生器产生整数随机数序列是以2进制位串形式排列，严格检验方法是对这样的2进制位串序列应具有的随机性能和规律，进行统计检验，所以严格检验方法是按整数随机数二进制位串排列的检验方法。按整数随机数二进制位串排列检验比按(0,1)随机数10进制数值大小排列检验要严格得多。在检验方法设计上，有很多独到之处。比如，Marsaglia (1995, 2008, 2010) 提出的严格检验方法，目前比较流行的Diehard程序，包括2进制秩检验、猴子检验、计数1检验、生日间隔检验、最大公因数检验和大猩猩检验。

1.3 概率分布抽样方法

随机变量的随机抽样指的是由该变量的总体分布产生简单随机子样。在用蒙待卡罗法解题时. 经常会遇到具有不同分布的各种随机变量. 要求产生对应于该随饥变量的随机子样.也即随机数列。这一步骤称为对该随机变量的随机模拟或随机抽样。

为方便起见，用X_F表示由已知分布$F(x)$中产生的简单子样的个体。对于连续型分布，常用分布密度函数$f(x)$表示总体的已知分布，用X_f表示由已知分布密度函数$f(x)$产生的简单子样的个体。另外，在抽样过程中用到的伪随机数均称随机数。

1.3.1 离散型分布随机变量的直接抽样

对于任意离散型分布：

$$F(x) = \sum_{x_i < x} P_i$$

其中x_1，x_2，…为离散型分布函数的跳跃点，P_1，P_2，…为相应的概率，根据前述直接抽样法，有离散型分布的直接抽样方法如下：

$$X_F = x_I, \qquad \text{当} \sum_{i=1}^{I-1} P_i < \xi \le \sum_{i=1}^{I} P_i, \xi \sim U(0,1)$$

该结果表明，为了实现由任意离散型分布的随机抽样，直接抽样方法是非常理想的。

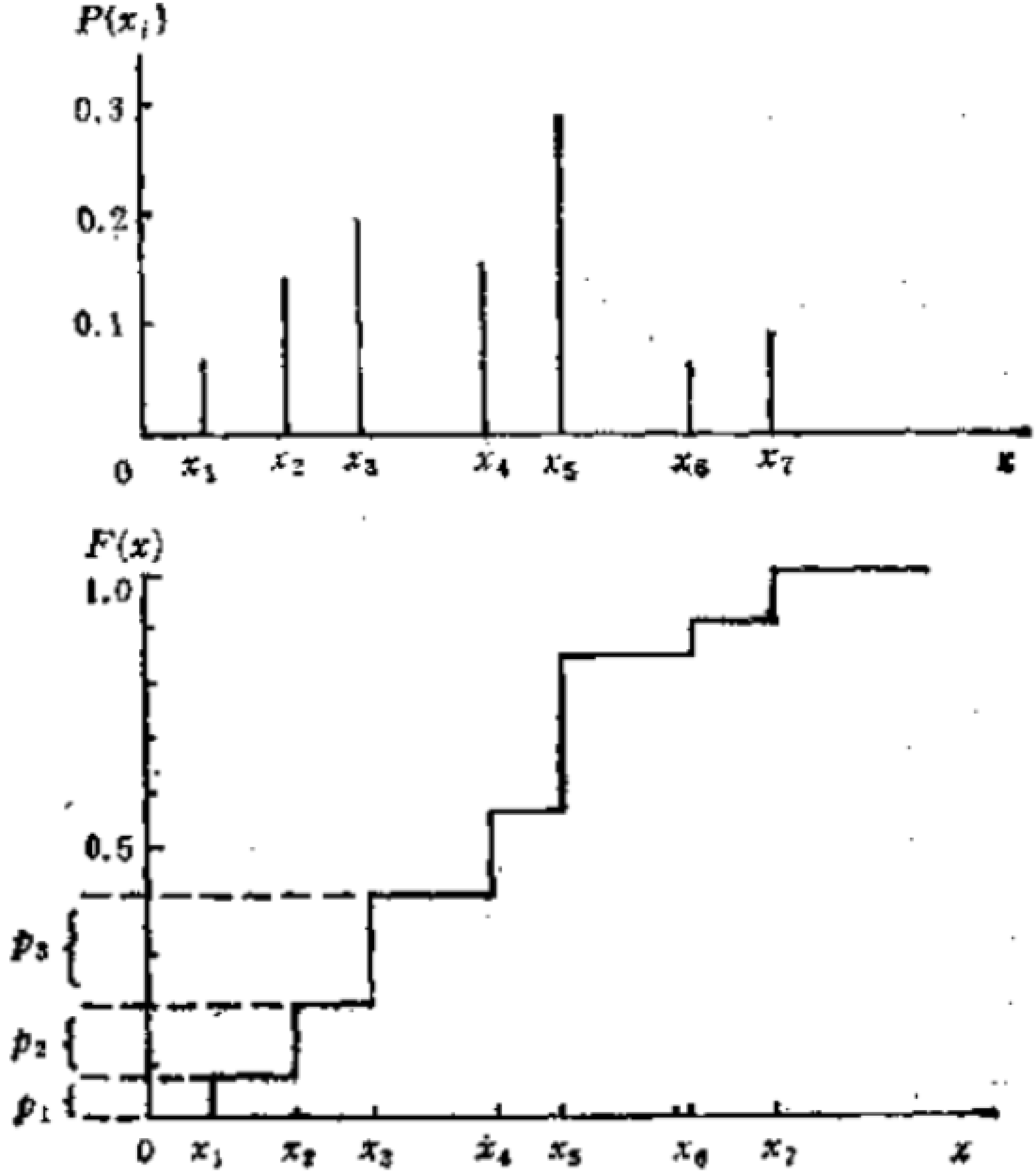


图 13.2　离散随机变量的概率分布和累积分布

例．掷骰子点数的抽样

掷骰子点数X=n的概率为：

$$P(X=n)=\frac{1}{6}$$

选取随机数ξ~U(0,1)，如

$$\frac{n-1}{6}<\xi\leq\frac{n}{6}$$

则

$$X_F=n$$

在等概率的情况下，可使用如下更简单的方法：

$$X_F=[6\cdot\xi]+1$$

其中［］表示取整数。

1.3.3.连续分布的随机变量抽样

- **A: 直接抽样方法(反函数法)**

任意连续随机变量Y的概率密度为g(y)，累积分布函数为 G(y)

$$x = G(y) = \int_{\infty}^{y} g(t)dt$$

为(0,1)区间中均匀分布的随机变量。 若G(y)的反函数存在. 上式可写成

$$y = G^{-1}(x)$$

所以若令ξ_i为随机置量Y 的随机子样(随机数列)，r_i 为(0,1) 区间均匀分布的随机子样(随机数列)，则应在

$$r_i = G(\xi_i) = \int_{-\infty}^{\xi_i} g(y)dy$$

写成反函数形式

$$\xi_i = G^{-1}(r_i)$$

利用这个公式可从r_i求出任意连续随机变量的随机ξ_i系列。

例1：对指数分布的直接抽样

满足指数分布$f(x) = \lambda e^{-\lambda x}(x \geq 0)$的随机数$\xi_i$

$$r_i = \int_0^{\xi_i} \lambda e^{-\lambda x} dx = 1 - e^{-\lambda \xi_i}$$

求得：$\xi_i = -\frac{1}{\lambda}\ln(1 - r_i)$.

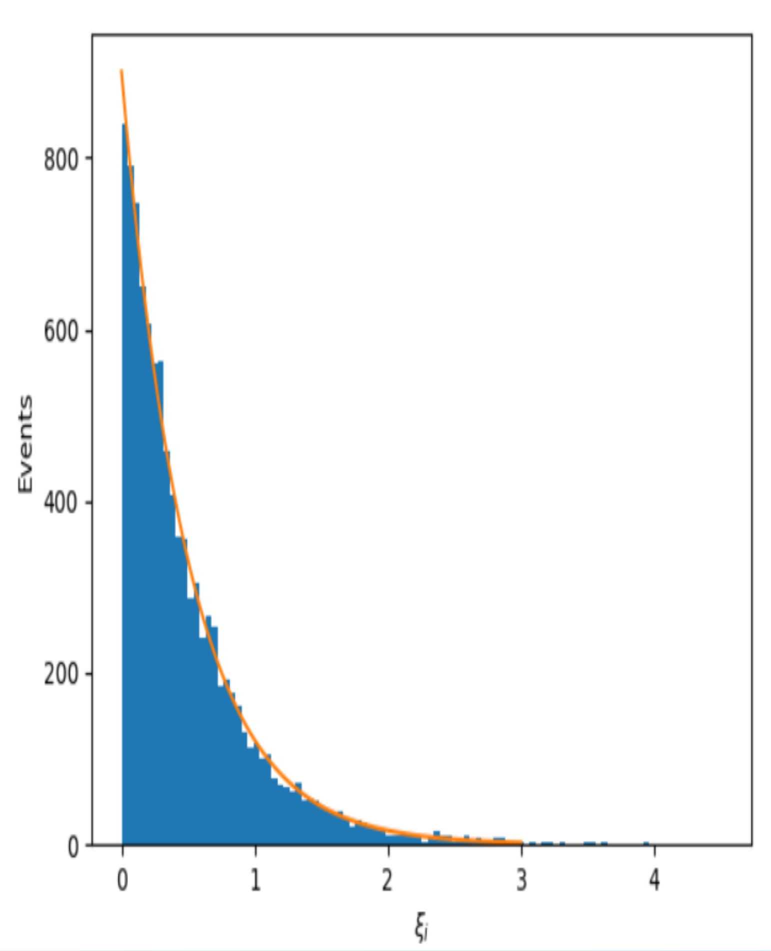
注意到r_i与$1 - \mathrm{r_i}$都是[0,1]区间均匀分布的随机数，故上式可以写成：

$$\xi_i = -\frac{1}{\lambda}\ln r_i$$

Python代码

```
import numpy as np
from matplotlib import pyplot as plt

ri = np.random.rand(10000)
xi = -1/2*np.log(ri)
plt.hist(xi,bins = 100)
x = np.linspace(0,3,100)
y = 2*np.exp(-2*x)*450
plt.plot(x,y)
plt.xlabel(r'$\xi_{i}$')
plt.ylabel('Events')
plt.show()
```



小结：直接抽样方法

1. 检查密度函数$f(x)$是否归一，若不归一，要做归一化处理。

2. 计算

$$r_i = G(\xi_i) = \int_{-\infty}^{\xi_i} g(y)dy$$

3. 反解出ξ_i, 即得到满足分布密度函数$g(y)$的抽样

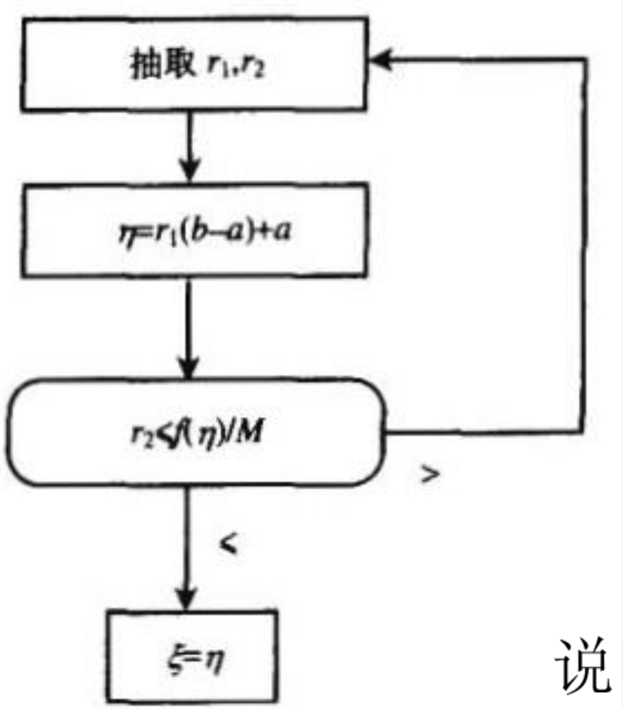
$$\xi_i = G^{-1}(r_i)$$

C.舍选抽样方法

舍选法是冯·诺曼(Von Neumann)为克服直接抽样和变换抽样方法的时难最早提出来的。它抽样的基本思想是按照给定的分布密度函数$\phi(\xi)$对均匀分布的随机数序列$\{\xi_i\}$进行舍选。舍选的原则是在$\phi(\xi)$大的地方，保留较多的随机数ξ_i；在$\phi(\xi)$小的地方，保留较少的随机数ξ_i。使得到的子样中ξ_i的分布满足分布密度函数$\phi(\xi)$的要求。这种方法对分布密度函数$\phi(\xi)$在抽样范围内有界，且其上界是容易得到的情况，总是可以采用的。它使用起来十分灵活，计算也较简单，因而使用也比较广泛。但是这种方法，对$\phi(\xi)$在抽样范围内函数值变化很大的时候，效率是很低的，因为大量的均匀分布的抽样点被舍弃了。由于这个原因，有时我们选择另外一些更有效的方法。

- 第一类舍选法

设随机置量X的取值域为$[a,b]$,概率密度$f(x)$为有界函数,其极大值用M 表示$f(x) \le M$,舍选抽样的流程图:



框图的含义如下:

η是$[a,b]$间内均匀分布随机数,考察$r_2 \le \phi(\eta)/M$ 是否成立. 若成立,则η取为随机变量X的随机数ξ_i; 否则,η被舍弃。重复以上过程可求得X的随机数列$\{\xi_i\}$.

说明:M可以比最大值大,但不能比最大值小。在实际工作中,M通过有限次计算得到时,为安全起见,可以把M适当放大,如10%或更高。

舍选抽样法的证明

讨论：

- 对于无限区间的情形，可进行截尾处理. 即选择有限区间[a,b]时，满足 $\int_a^b f(x)\mathrm{d}x > 1-\varepsilon$.

只要ϵ且够小. 就可以应用上述方法而使计算误差满足要求。

- 舍选抽样的效率

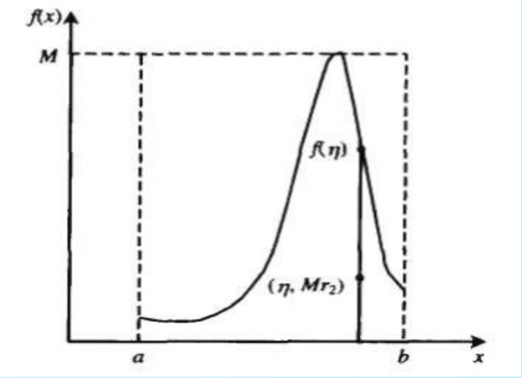
在舍选法中，假定产生n对r_1, r_2(共$2n$个随机数). 其中满足不等式$r_2 \le f(\eta)/M$ 的η的个数m与n之比. 等于曲线$f(x)$下的面积与总面积$(b-a)M$ 之比.

$$\frac{m}{n} = \frac{\int_a^b f(x)\mathrm{d}x}{(b-a)M} = \frac{1}{(b-a)M}$$

$f(x)$可以是不归一的

因此，舍选法的抽样效率为

$$E = \frac{m}{2n} = \frac{1}{2M(b-a)}.$$



例：对随机变量η抽样，使它满足分布函数

$$f(x)=\begin{cases}2x, 0\le x\le 1\\ 0, 其他\end{cases}$$

解：在$x\in[0,1]$,分布函数的最大值$M=2$. 产生两个随机数 $\xi_1,\xi_2\in[0,1]$,判断

$$\xi_2\le f(\xi_1)/2，\quad 或者\ \xi_2\le\xi_1$$

是否成立，若成立，取$\xi=\xi_1$,若不成立，再取两组随机数，重复以上操作。实际上，由于随机数的独立性，ξ_1,ξ_2的值可以互换，使它们满足舍选要求，所以可以取

$\xi=\max(\xi_1,\xi_2)$. (思考题：用反函数法怎么抽样？)

类似地，可以推广到高幂次情况：

$$f(x)=\begin{cases}nx^{n-1}, 0\le x\le 1\\ 0, 其他\end{cases}$$