

实验模拟与数据分析工具

平荣刚

中国科学院高能物理研究所
(pingrg@ihep.ac.cn)

中国科学院大学，2024-3，课程编号：070200M02020H

目 录

第四章 蒙特卡罗模拟

4.1: 蒙特卡罗方法简史

4.2: 随机数产生和检验

4.3: 概率分布抽样方法

4.4: 蒙特卡罗模拟在物理中的应用

第五章 物理事例产生器

5.1: 事例产生器的作用和原理

5.2: 物理事例的运动学描述

5.3: PYTHIA产生子

5.4: 高能物理实验中常用的产生子

教材/参考资料

- **实验物理数据分析（下册）**
朱永生 著 科学出版社(2012)
- **蒙特卡罗方法在实验核物理中的应用**
许淑艳 编著 原子能出版社
- **A Primer for the Monte-Carlo Method**
Iiya M. Sobol, CRC press, Inc. (1994)
- **Exploring Monte Carlo Methods**
William L. Dunn and J. Kenneth Shultis, Elsevier, 2012
- **Introduction to Monte Carlo methods**
Stefan Weinzierl, arXiv:hep-ph/0006269

4.1 蒙特卡罗方法简史

4.1.1: 蒙特卡罗方法产生历史

- 启蒙时期
- 开创时期

4.1.2: 蒙特卡罗方法发展简况

- 发展概要
- 发展动力
- 存在的问题

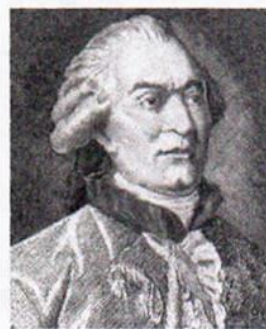
4.1.1: 蒙特卡罗方法产生历史

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益广泛。

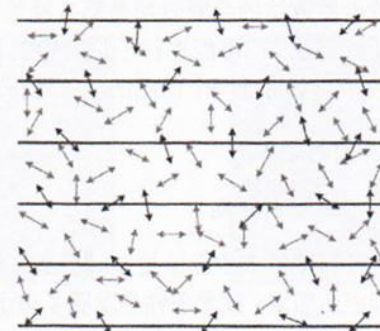
然而这种方法并非新颖，人们在很早以前，就已经加以利用。伯努利在十七世纪的《推算术》中指出，“这个方法并不新鲜，也不特别。每个人都明白，要做这种关于某种现象的推断，做一、二次观察是不够的，需要做大量的观察，……，这种观察越多，则达不到目的的危险就越小。

- 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

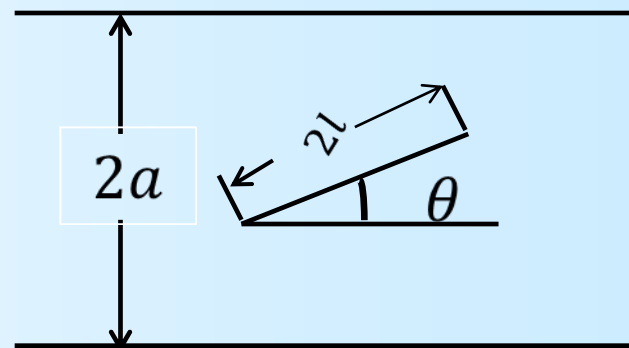
$$\pi = \frac{2l}{aP} \approx \frac{2l}{a} \left(\frac{N}{n} \right)$$

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：

$$\text{如果 } l = a: \delta\pi \approx \frac{2.37}{\sqrt{n}}$$

试求之→

投针实验的蒙特卡洛模拟

- 用投针的方法测量 π 值，这是一条以传统方法不同的新途径

➤ 建模: $\pi = \frac{2l}{aP}$

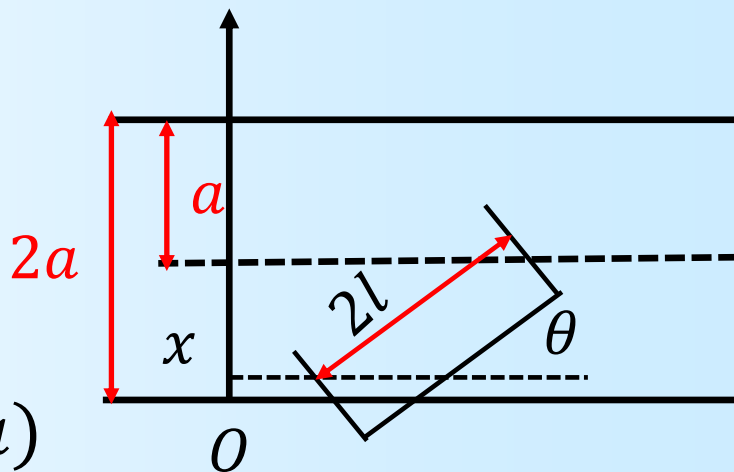
➤ 投针实验:
 x, θ 的密度函数

$$f_1(x) = \begin{cases} 1/a & (0 \leq x \leq a) \\ 0 & (\text{其他}) \end{cases}$$

$$f_2(\theta) = \begin{cases} 1/\pi & (0 \leq \theta \leq \pi) \\ 0 & (\text{其他}) \end{cases}$$

即: $x = a\xi_1, \theta = \pi\xi_2,$

ξ_1, ξ_2 是 $(0,1)$ 中的均匀分布随机数



投针实验的科学意义

➤ 统计结果: 每次投针相交的次数 $s(x, \theta)$

$$s(x_i, \theta_i) = \begin{cases} 1, & \text{当 } x_i \leq l \sin \theta_i, \\ 0, & \text{其他} \end{cases}, \quad s_N = \frac{1}{N} \sum_{i=1}^N s(x_i, \theta_i)$$

- 开启了用随机性事件的模拟方法, 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
100,000	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

有人对Lazzarini结果的质疑

Lazzarini实验：采用 $\frac{L}{D} = \frac{5}{6}$ ，投针 $n_d = 3408$ 次，得到针与水平线相交的次数为 $n_c = 1808$ 次. 得到 $\pi = 3.1415929$ 。

如果 π 用小于16,000的有理数表达，最好的近似是

$$\pi \approx \frac{355}{113} \approx 3.141592920.$$

如果 $P_{cut} = \frac{n_c}{n_d}$ ，那么

$$n_d \approx \pi \frac{D}{2L} n_c = \frac{355}{113} \frac{6}{2 \times 5} n_c = \frac{213}{113} n_c$$

如果 $n_c = 16 \times 113 = 1808$ ，那么 $n_d = \frac{213}{113} \times 16 \times 113 = 16 \times 213 = 3408$.

所以，Lazzarini很幸运！

2. 费米模拟装置

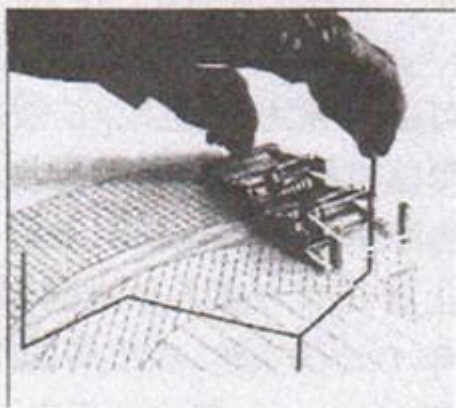
20世纪30年代，费米发明的“FERMIAC”机械模拟实验装置，用来模拟中子在核装置内的随机扩散运动。实验时

“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.Fermi(1901~1954)



(b)

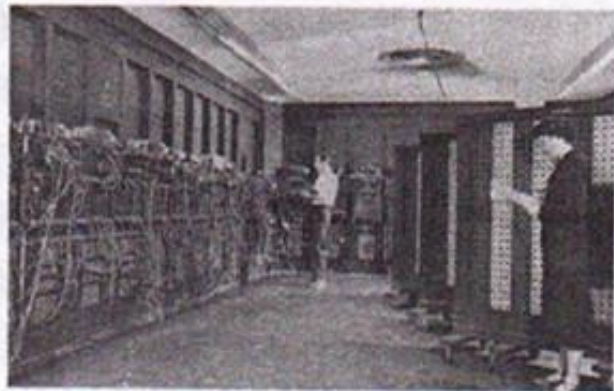


(c)

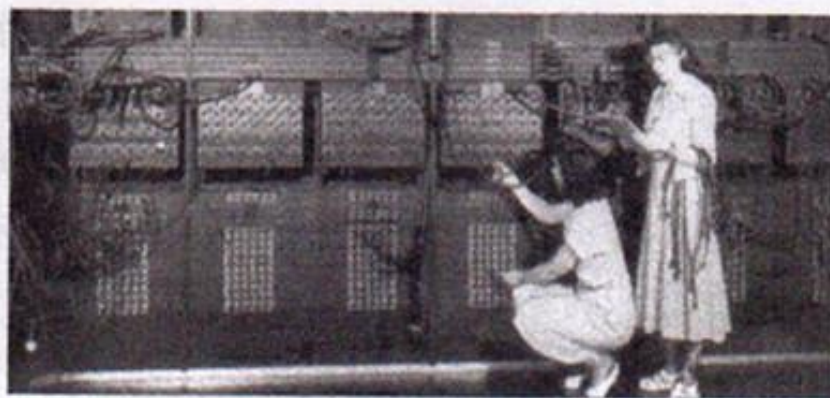
- 开创时期

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），加法5000次/秒，乘法400次/秒，内存：20多个字节



(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) S.Ulam(1909~1984)



(b) J.von Neumann(1903~1957)



(c) N.Metropolis (1915~1999)

图 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

4.1.2:蒙特卡罗方法发展简况

-发展概要

1946年是蒙特卡罗的开创年，蒙特卡罗方法已走过70年，可以1980年为界线，把蒙特卡罗方法的发展历史分为前后两个时期。蒙特卡罗方法理论包括三个方面的内容：随机数的产生和检验方法、概率分布抽样方法、降低方差提高效率的方法。

1. 前一时期发展概要

前一时期，蒙特卡罗理论发展较为缓慢。1978年研制的 α 粒子源真随机数产生器每秒产生2个随机数，4种伪随机数产生器都存在不同缺点，概率分布抽样方法主要是直接抽样方法，Metropolis算法还未发展成为马尔可夫蒙特卡罗方法，蒙特卡罗方法主要解决的是估计值的问题，解决最优化的蒙特卡罗方法还未发展起来。

2. 伪随机数发展的问题

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，二是经典的斐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} 。

3. 后一时期理论发展概要

后一时期蒙特卡洛理论发展较快，蒙特卡洛模拟可以使用真随机数，随机数统计检验出现了马萨格利亚的严格检验方法，Metropolis算法发展成为马尔科夫蒙特卡洛方法，发展了概率抽样多种方法，出现了许多高效的蒙特卡洛方法（互熵方法、稀有事件模拟方法、马尔科夫链蒙特卡洛方法、序贯蒙特卡洛方法.....）

4. 应用和发展

蒙特卡洛方法的应用已经发展到多个领域，传统领域是核科学，主要涉及核粒子的运输。20世纪50年代后，蒙特卡洛方法迅速扩展到其他领域，主要包括科学技术、工程、统计和金融等经济领域。例如：确定性问题、粒子运输、稀薄气体动力学、物理化学核和生物学、粒子滤波核粒子分裂、数理统计学和可靠性、金融经济学和科学实验模拟等。蒙特卡洛方法的发展动力来源于实际应用，很多抽样算法和降低方差提高效率的方法都产生于实际应用。

-蒙特卡罗方法的发展动力

1. 内部发展动力

蒙特卡洛方法的内部发展动力来源于自身理论发展的要求。蒙特卡洛方法的数学性质决定了其收敛速度慢、计算精度低。为了加快收敛，提高精度，要求提高伪随机数的产生速度更快，品质更优，检验方法更为严格，提高概率抽样效率，降低方差。

2. 外部发展动力

蒙特卡洛方法的发展外部动力来源于实际需求。很多高维问题的数值解的误差将随着维数的增加而迅速增加，消耗的计算时间成指数性增长。而蒙特卡洛处理高维问题时，计算误差与维数无关，这成为蒙特卡洛方法的最大优势。此外，蒙特卡洛方法适合处理很多结构或过程复杂的问题，如：粒子的输运问题、截面与能量有关的问题、散射各向异性、介质非均匀性、几何形状复杂和时间有关的问题。

-存在的问题

蒙特卡洛被人们称为“最后的方法”.

- (1)蒙特卡罗方法既可以解决估计值的问题，也可以解决最优化的问题，具有解决问题的超强适应能力，误差容易控制，程序结构简单清晰，应用灵活性强。缺点是收敛速度慢，但可以改善。
- (2)蒙特卡罗的精度问题。
- (3)正确地选择伪随机数产生器。
- (4)拟随机数的“丛聚”问题。
- (5)马尔可夫链的蒙特卡罗抽样问题。
- (6)蒙特卡罗方法的“双重性”挑战。

4.2 随机数产生和检验

4.2.1: 真随机数产生器

4.2.2: 早期伪随机数产生器

4.2.3: 伪随机数产生器的发展

4.2.4: 随机数理论检验和统计检验

4.2.1. 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 α 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32bit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。

- 量子真随机数产生器

- 量子真随机数产生器是利用光学量子技术，根据光子在半透明镜子上反射和透射光子的随机性来产生二进制随机数。这种技术相对成熟。2001年，量子随机数产生器已经成为商品。瑞士ID量子公式的产品：



图 2.1 QRBG121



图 2.2 QUANTIS-OEM

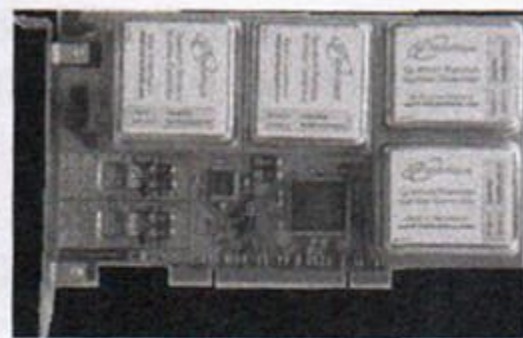


图 2.3 QUANTIS PCI-4

网站：<http://www.random.org>提供真随机数

QRBG121:重370克，支持各种操作系统，每秒可产生37.4万随机数。

QANTIS-OEM:重30克，每秒可产生12.5万随机数。

QANTIS-PCI-4:每秒可产生50万随机数。

当前，使用真随机数模拟器进行蒙特卡罗模拟已经成为现实。目前，在微机上伪随机数产生器平均每秒可产生2000万个伪随机数，比真随机数产生器几乎高2个量级。目前，真随机数产生器价格还比较昂贵，要普及使用真随机数产生器还比较困难。对大规模问题，需要使用巨型机，直接使用真随机数产生器，由于产生速度低，毫无优势可言，因此，在巨型机上使用真随机数产生器进行大规模计算，是不合适的。由于伪随机数性能已经很接近真随机数，没有必要非得使用真随机数。因此，蒙特卡罗模拟完全使用真随机数，既无可能，也无必要。

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：

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

随机数序列 (U_1, U_2, \dots) 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 $(U_{n+1}, U_{n+2}, \dots, U_{n+s})$ ，在 s 维空间单位立体 G_s 上均匀分布。对任意的 a_i ， $0 \leq a_i \leq 1, i=1, 2, \dots, s$ ， $U_{n+i} \leq a_i$ 的概率为：

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

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：

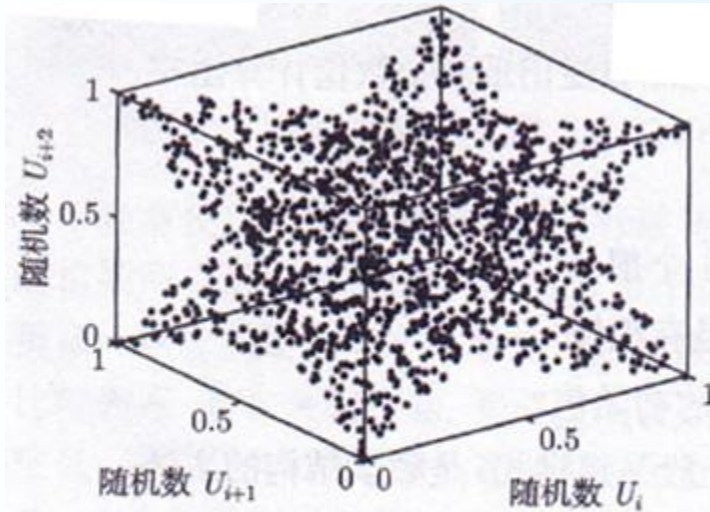
$$X_i \equiv (X_{i-2} + X_{i-1})(\text{mod } M), \quad i > 2, \quad U_i = \frac{X_i}{M}$$

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：

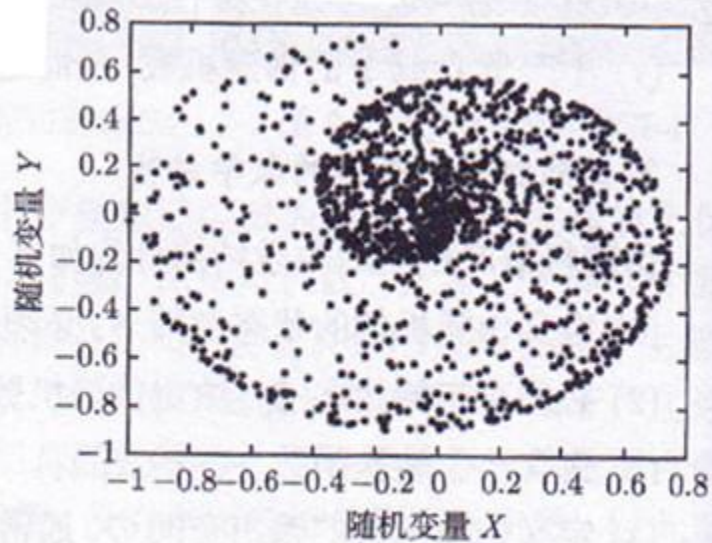
1,1,2,3,5,8,13,21,

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。

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



(a)



(b)

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 (\oplus)，递推公式为：

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

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

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：

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

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。

乘同余参数选择：参数选择是在保证最大周期条件下的参数选择。模 M 由机器字长决定，32位机， $M=2^{32}$ 。初始值 X_0 一般可为任意正数，也可以是小于 M 的正奇数。若模 M 是素数，乘子 A 可选为 M 的**原根**，这时，乘同产生器的全周期为 $M-1$ 。对于32位机，乘同产生器的周期为 $2^{32}-1 \sim 2.1 \times 10^9$ 。

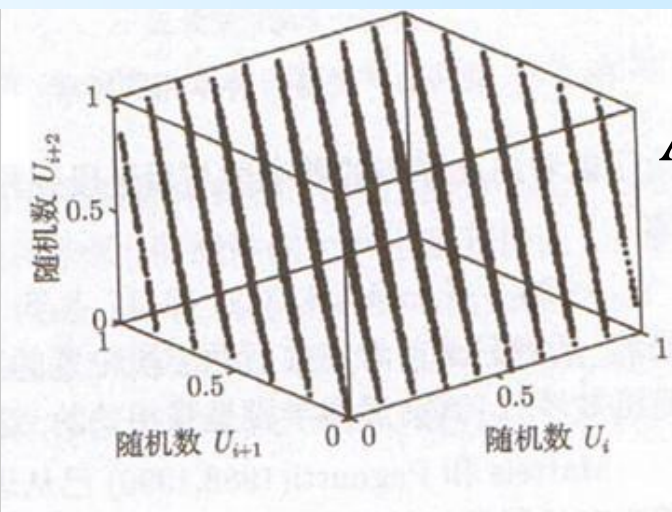
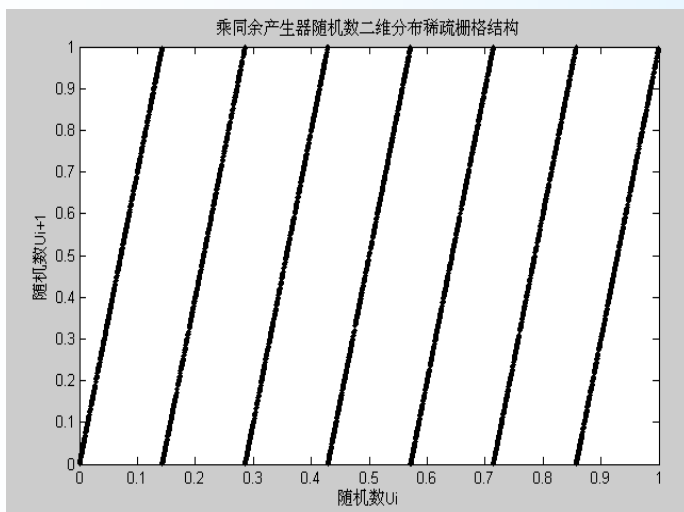
混合同余产生器的参数选择：模 $M = 2^\beta$ ，最大周期为 M ，参数 M, A, C 的选择准则：当且仅当 $A \equiv 1 \pmod{4}$ ， M 与 C 的最大公约数为1。

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。

IBM, RANDU:



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

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

乘同余产生器 ($A=7$) 的稀疏栅格

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。

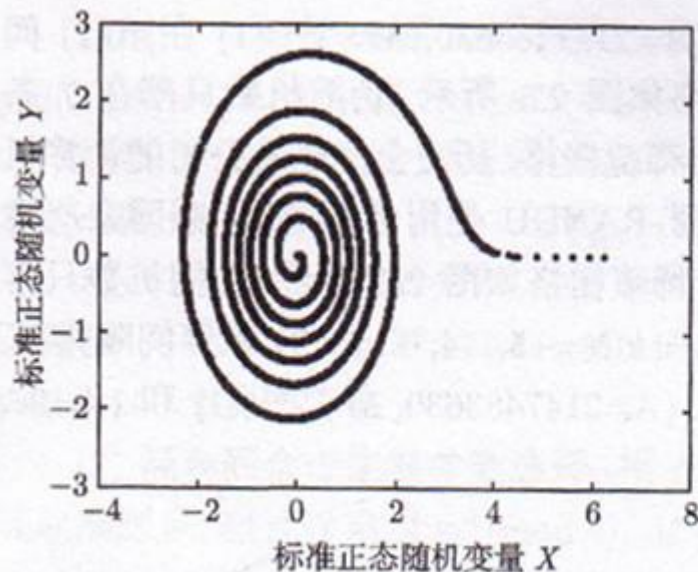


图 2.7 乘同余产生器 1 样本相关异常

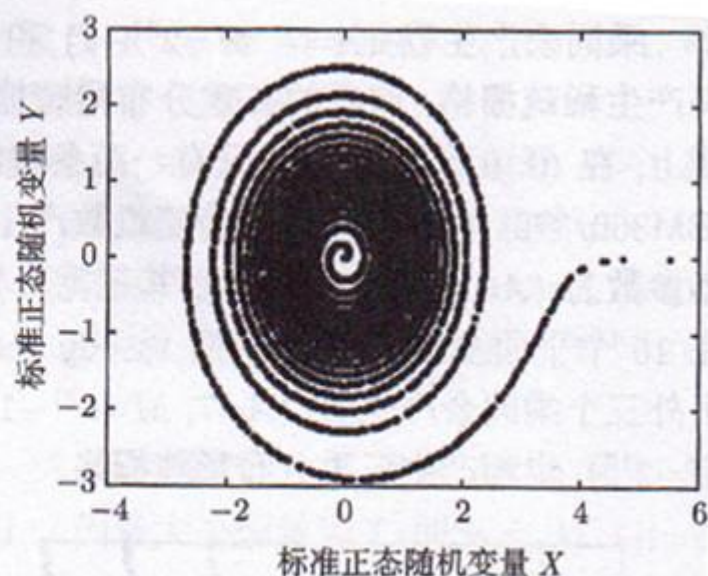


图 2.8 乘同余产生器 2 样本相关异常

■ 长周期相关性

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

$\{X_i\}$: 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。

4.2.3: 伪随机数产生器的发展

● 非线性同余产生器

令 M 是素数, $F_M = \{0, 1, \dots, M-1\}$ 是一个 M 阶Galois域, $g(X_{i-1})$ 是 F_M 上的一个非线性整数函数, 通常是 F_M 上的一个排列多项式, 此时有 $\{g(0), g(1), \dots, g(M-1)\} = F_M$, $X_{i-1} \in F_M$.

$$X_i \equiv g(X_{i-1}) \pmod{M}, i \geq 1.$$

两种非线性同余产生器:

✓ 逆同余产生器

递推公式: $X_i \equiv (AX_{i-1}^{-1} + C) \pmod{M}, i \geq 1.$

$A, C \in F_M$. $AC \neq 0$, 当**本原多项式** $f(x) = x^2 - Cx - A \in F_M$, 产生器的全周期为 M 。

逆余同产生器的主要优点是消除高维稀疏栅格结构。但它仍然存在长周期相关现象, 并且速度明显慢于线性同余产生器。

✓ 二次式同余产生器

若非线性整数函数 $g(X_{i-1})$ 取二次式形式，递推公式为：

$$Xi \equiv (AX_{i-1}^2 + BX_{i-1} + C) \pmod{M}$$

令 p 为素数， $M = p^b$ ， $b \geq 2$ ，如果满足下列条件， $A \equiv 0 \pmod{p}$ ， $B \equiv 1 \pmod{p}$ ， $C \not\equiv 0 \pmod{p}$ ；当 $p=2$ 时， $A \equiv (B-1) \pmod{4}$ ；当 $p=3$ 时， $A \equiv 0 \pmod{9}$ ； $AC \equiv 6 \pmod{9}$ ，则
 $a \in \{1, 2, \dots, b-1\}$ ，最大公因数 $\gcd(A, p^b) = p^a$ ，Kunth(1981)
证明二次式同余产生器的最大周期为 p^b 。

● 进位借位运算产生器

Masagilia, Zaman(1991)首先提出了进位借位运算产生器。

进位借位产生器有进位加法产生器，进位减法产生器，进位乘法产生器和借位减法产生器。

✓ 进位加法产生器

设 b, p, q 为正整数, b 称为基, p, q 称为延迟, $p > q$, 进位加法产生器递推加法器为:

$$X_i \equiv (X_{i-p} + X_{i-q} + C_{i-1}) \pmod{b}, i \geq p$$

式中, C_i 为进位, 若 $X_{i-p} + X_{i-q} + C_{i-1} \geq b$, 则 $C_i = 1$; 若 $X_{i-p} + X_{i-q} + C_{i-1} \leq b$, 则 $C_i = 0$; 进位加法器的初值由 $(X_0, X_1, \dots, X_{p-1})$ 和 C_{p-1} 构成。由于进位加法没有乘法运算, 而 \pmod{b} 运算不会超过一个减法运算, 所以进位加法产生器速度很快, 效率很高。而当 $M = b^p + b^q - 1$ 为素数时, 且 b 是 M 的一个原根时, 进位加法器的最大周期为 $b^p + b^q - 1$ 。例如, $b = 2^{31}$, $p = 20$, 最大周期为 $2^{620} \approx 10^{186}$ 。

✓ 进位减法产生器

进位减法产生器递推加法器为：

$$X_i \equiv (X_{i-p} - X_{i-q} - C_{i-1}) \pmod{b}, i \geq p$$

式中， C_i 为进位，若 $X_{i-p} - X_{i-q} - C_{i-1} \geq b$ ，则 $C_i=1$ ；若 $X_{i-p} - X_{i-q} - C_{i-1} \leq -b$ ，则 $C_i=0$..

✓ 进位乘法产生器

进位乘法产生器递推式为：

$$X_i = A_n X_{i-n} + \cdots + A_2 X_{i-2} + A_1 X_{i-1} + C_{i-1}$$

式中，进位 $C_i \equiv X_{i-1} \pmod{M}$. 例如，一个乘法产生器的参数可取为： $A_1=698769069$, $X_0=521288629$, $C_0=7654321$, $M=2^{32}$.

• 迟延裴那波契产生器

迟延裴那波契产生器是使用序列中更前面的随机数去产生新的随机数。它的递推公式为：

$$X_i \equiv (X_{i-p} \otimes X_{i-q}) \pmod{b}, i \geq p$$

$$q = 0, p = 1, \oplus = + \quad ??$$

式中， $p, q (p > q)$ 称为迟延系数，运算符 \otimes 代表二进制操作符 $+, -, \times, \oplus$ 之一，产生随机数需要 p 个初值： $X_{0,1}, X_{0,2}, \dots, X_{0,p}$ 。

迟延裴波那契产生器的质量依赖于选择的参数，它的最大周期为 $M^p + M^q$ 。例如，对于32位机， $M=2^{31}$ ，最大周期为 $M^{31p} + M^{31q}$ 。选择 $p=1279, q=418$ ，周期为 10^{20169} 。周期几乎达到了无限长。

• 线性同余组合产生器

组合产生器是由几个单一随机数产生器组合而成，它有可能得到比单一产生器性能更好的随机数。例如，Marsaglia(1984)提出的KISS (keep it simple stupid)产生器 KISS84, 它是有线性混合同余、联合移位寄存器和进位乘三个产生器组成的。整数随机数为：

$$X_i = I_i + J_i + K_i$$

其中：

$$I_i \equiv (69069I_{i-1} + 12345) \pmod{2^{32}}$$

搅拌器

$$J_i = J_{i-1} \oplus (J_{i-1} \ll 13), J_i = J_i \oplus (J_i \gg 17), J_i = J_i \oplus (J_i \ll 5);$$

$$K_i = 698769069K_{i-1} + C_{i-1}, C_i \equiv K_i \pmod{2^{32}}.$$

初始值 $I_0 = 123456789, J_0 = 362436, K_0 = 521288629, C_0 = 7654321$, 周期为 $2^{124} \approx 2.12 \times 10^{37}$.

• 麦森变形产生器

Tausworthe反馈移位寄存器产生器曾经受到批评和非议，实际工作中也出现了问题，但由于反馈移位寄存器快速高效和长周期的特点，仍然吸引人们研究探索。Masumoto 等人(1992-2008)提出的变型的广义反馈移位寄存器，周期为 $2^s - 1$ ， s 是素数， $2^s - 1$ 在数论上成为麦森数，所以，这种产生器也成为麦森变形产生器(Mersenne twister)。麦森变形产生器由一个线性广义反馈位移寄存器产生器和一个联合位移寄存器组成，前者是产生快速长周期的随机数，后者的作用是把随机数搅拌得更加均匀，麦森变型产生器为：

$$X_i = X_{i-k+m} \oplus ((X_{i-k} \& p) | (X_{i-k+1} \& q))A,$$

$$Y_i = X_i,$$

$$Y_i = Y_i \oplus (Y_i \gg u),$$

$$Y_i = Y_i \oplus ((Y_i \ll s) \& b),$$

$$Y_i = Y_i \oplus ((Y_i \ll t) \& c),$$

$$Y_i = Y_i \oplus (Y_i \gg l),$$

$$U_i = Y_i / M.$$

式中， p, q, b, c 是16进制数，参数 u, s, t, l 是10进制参数， k 是递推度， $1 \leq m \leq k, m$ 是地推度的中值。

4.2.4 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**。

统计检验的步骤

- 提出要检验的假设，构造统计检验方法
- 选举检验统计量，确定检验统计量所遵从的分布(标准正态分布, χ^2 分布....)
- 给出显著水平，确定检验判别法则
- 根据统计检验方法，计算检验统计量和检验概率值
- 进行统计推断，判定假设成立与否

• 参数检验(矩检验)

检验所产生的随机数系列的分布是否与(0,1)期间均匀分布随机数变量 U 的相应参数一致。对于所产生的 N 个随机数： $r_1, r_2, r_3, \dots, r_N$ ，各阶子样矩定义为：

$$\hat{m}_k = \frac{1}{N} \sum_{i=1}^N r_i^k \quad (k = 1, 2, \dots)$$

若零假设 H_0 为真，各阶子样矩的期望值和方差为：

$$m_k = \frac{1}{k+1}, \quad \sigma_{k,N}^2 = \frac{1}{N} \left(\frac{1}{2k+1} - N m_k^2 \right)$$

提示：可以从MC积分去理解

统计量 $Z_{k,N} = \frac{\hat{m}_k - m_k}{\sigma_{k,N}}$ 渐进地服从标准正态分布 $N(0,1)$ ，
就可以按照给定的显著水平检验 H_0 。

● 均匀性检验

(A) 频率检验

将(0,1)区间划分为 k 个等长的子区间，每个子区间长度为 $1/k$ ，设产生了 N 个伪随机数 r_1, r_2, \dots, r_n ，落在第 j 个子区间的伪随机数个数记为 $n_j (j = 1, 2, \dots, k)$ ，称为经验频数。如果零假设 H_0 为真，这 N 个随机数落在任一区间的频率为 $p_i = 1/k$ 。故落在任意区间内的随机数个数的理论值(理论频数)为：

$$m_j = Np_j = N/k, j = 1, 2, \dots, k$$

统计量

$$\chi^2 = \sum_{j=1}^k \frac{(n_j - m_j)^2}{m_j}$$

渐近地服从自由度 $k-1$ 的 χ^2 分布。因此，对于给定的显著水平 α 。可利用皮尔逊检验来确定随机数列是否满足均匀分布的零假设 H_0 。

(B) 累积频率检验

将所产生的 N 个随机数按数值递增的次序排列 $r_1, r_2, r_3, \dots, r_N$ 。
子样分布函数 $F_N(r)$ 在 r_i 点的数值为

$$F_N(r_i) = i / N.$$

这也就是随机数列的累积频率。而零假设即是 $(0,1)$ 区间的均匀分布。其累积分布在 r_i 点的数值是

$$F_0(r_i) = r_i.$$

所以令 $n(x)$ 是 $r_1, r_2, r_3, \dots, r_N$ 中满足 $r_i < x$ 的随机数个数，则统计量

$$D_N = \max_{0 < x < 1} \left| \frac{n(x)}{N} - x \right|,$$

可用柯尔莫哥洛夫检验方法，在给定的显著性水平 α 下接收还是拒绝零假设 H_0 。

• 独立性检验

• 相关系数检验

设产生了 N 个随机数 $r_1, r_2, r_3, \dots, r_N$, 各随机数之间相互独立的必要条件是它们的相关系数等于0, 若前后相距为 j 的随机数之间的相关系数记为 ρ_j , 按相关系数的定义. ρ_j 的估计量

:

$$\hat{\rho}_j = \left[\frac{1}{N-j} \sum_{i=1}^{N-j} r_i r_{j+i} - (\bar{r})^2 \right] / s^2, \quad j=1, 2, \dots,$$

其中:
$$s^2 = \frac{1}{N} \sum_{i=1}^N \left(r_i - \frac{1}{2} \right)^2.$$

对充分大的 N (如 $N - j > 50$). 当原假设 $H_0: \rho_j = 0$ 为真。统计量

$$u = \hat{\rho}_j \sqrt{N-j},$$

渐近地服从标准正态分布, 因而 u 可作为随机数序列 $r_1, r_2, r_3, \dots, r_N$ 独立性检验的检验统计量。

-联列表独立性检验

在 xy 平面上将 $0 \leq x \leq 1$, $0 \leq y \leq 1$ 的正方形分 $J \times K$ 个矩形, 用任意一种方法将伪随机数列 $\{r_i\} (i = 1, 2, \dots, N)$ 两两组成二维空间上的点列 $\{r_{i1}, r_{i2}\} (i = 1, 2, \dots, N)$, n 为小于等于 $N/2$ 的最大整数。这些点中落入第 j , k 个矩形中的数记 n_{jk} , ($j = 1, \dots, J$; $k = 1, \dots, K$).

$$n_{j\cdot} = \sum_{k=1}^K n_{jk}, \quad n_{\cdot k} = \sum_{j=1}^J n_{jk},$$

得到联列表, 显然

$$\sum_{j=1}^J \sum_{k=1}^K n_{jk} = \sum_{j=1}^J n_{j\cdot} = \sum_{k=1}^K n_{\cdot k} = n.$$

$j \backslash k$	1	2	...	K	合计 $n_{j\cdot}$
1	n_{11}	n_{12}	...	n_{1K}	$n_{1\cdot}$
2	n_{21}	n_{22}	...	n_{2K}	$n_{2\cdot}$
\vdots	\vdots	\vdots		\vdots	\vdots
J	n_{J1}	n_{J2}	...	n_{JK}	$n_{J\cdot}$
合计 $n_{\cdot k}$	$n_{\cdot 1}$	$n_{\cdot 2}$...	$n_{\cdot K}$	N

统计量:

$$\chi^2 = n \sum_{j=1}^J \sum_{k=1}^K \frac{\left(n_{jk} - \frac{n_{j.} n_{.k}}{n} \right)^2}{n_{j.} n_{.k}}$$

渐近地服从自由度 $(J - 1)(K - 1)$ 的 χ^2 分布, 利用 χ^2 检验便可确定在给定显著性水平 α 上, 零假设 H_0 (相互独立的随机数列)是否被接受。

- 连贯性检验

随机数的连贯性检验是按照随机数出现的先后顺序, 检验它的连贯现象是否异常. 将随机数列 $r_1, r_2, r_3, \dots, r_N$ 按某种规律分成两类, 分别称为 a 类和 b 类。属于 a 类的概率为 p , 属于 b 类的概率为 $q = 1 - p$ 。例如, $r_i \leq p$ 的 r_i 属于 a 类; $r_i > p$ 的 r_i 属于 b 类, 即是一种分法。按随机数出现的先后顺序进行排列

$$\underbrace{aabbbaaababb\cdots}_{N \uparrow}$$

相连的同类元素构成游程，令 m, n 分别为 a, b 类元素的个数。显然 $N = m + n$ ，且总游程数记为 R ，它的分布为

$$p(R = 2k) = 2 \binom{m-1}{k-1} \binom{n-1}{k-1} p^m q^n,$$

$$p(R = 2k + 1) = \left[\binom{m-1}{k} \binom{n-1}{k-1} + \binom{m-1}{k-1} \binom{n-1}{k} \right] p^m q^n.$$

R 的数学期望值为：

$$E(R) = p^2 + q^2 + 2Npq,$$

$$V(R) = 4Npq(1 - 3pq) - 2pq(3 - 10pq).$$

当 N 充分大，统计量 Z 渐进地服从标准正态分布：

$$Z = [R - E(R)] / \sqrt{V(R)}$$

因此， Z 可作为检验随机数列连贯性的检验统计量。

作业:

- 1.采用线性同余法产生伪随机数，取 $A=5, C=1, M=16$ 和 $X_0=1$,记录下产生的前20个随机数，它产生数列的周期是多少？
- 2.取 $A=137, C=187, M=256$ 和 $X_0=1$,用线性同余法产生3维随机数和2维随机数，然后分别绘出其3维和2维分布图。
- 3.在Buffon实验中，求 N 次投针实验计算圆周率的方差和标准差。
- 4.在Buffon实验中，如果平行线换成边长为 L 的方格，针的长度为 L ，求投针与方格线相交的概率，并用实验验证。

投针试验的方差

投针N次有M次与横线相交，这个分布满足二项式分布定理。
它的方差为： $Np(1 - p)$ ，在投针试验中，

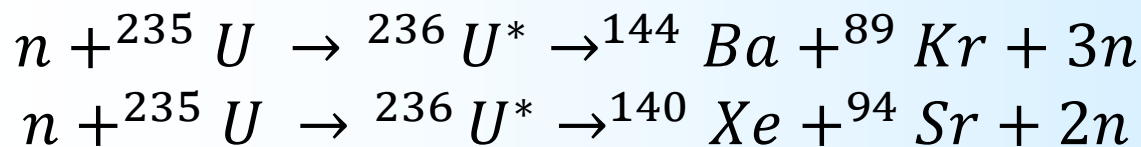
$\frac{M}{N} = \frac{2}{\pi}$ ， $\therefore \frac{\Delta M}{N} = \left| \Delta \left(\frac{2}{\pi} \right) \right| = \frac{2}{\pi^2} \Delta \pi$ ，把 $\Delta M = \sqrt{NP(1 - p)}$ 代入得到：

$$\Delta \pi = \frac{1}{\sqrt{N}} \pi \sqrt{\frac{\pi}{2} - 1} \approx \frac{2.37}{\sqrt{N}}$$

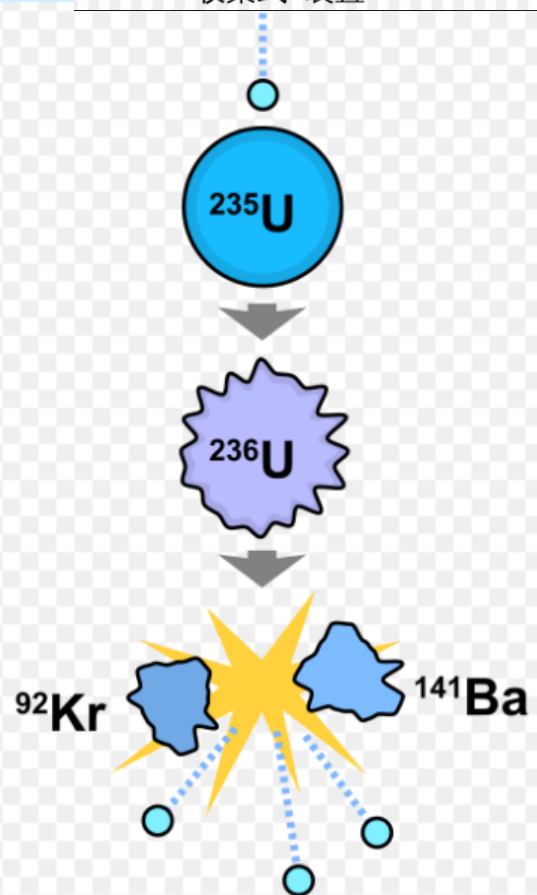
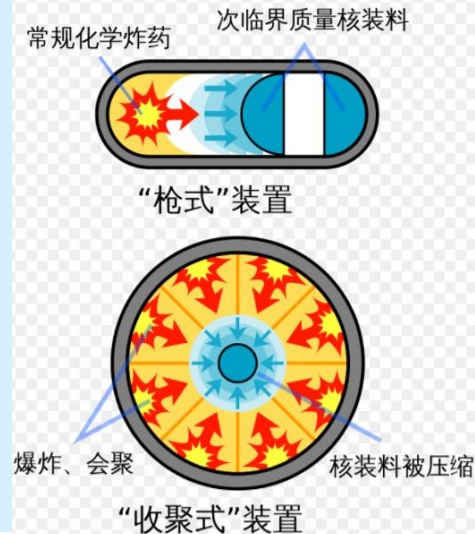
返回

[返回](#)

原子弹爆炸原理



铀-235和钚-239此类重原子核在中子的轰击后，通常会裂变变成两个中等质量的核，同时再放出2到3个中子和200兆电子伏的能量。在裂变中放出的中子，一些在裂变系统中损耗了，而一些则继续进行重核裂变（继续轰击重原子核）反应。只要在每一次的核裂变中所裂变出的中子数平均多余一个（即中子的增值系数大于1），那么核裂变即可以继续进行，一次一次的反应后，裂变出的中子总数以指数形式增长，而产生的能量也随之剧增。如果不加控制，最终，这个裂变系统会变为一个剧烈的链式裂变反应。



本原多项式

本原多项式是近世代数中的一个概念，是唯一分解整环上满足所有系数的最大公因数为1的多项式。本原多项式不等于零，与本原多项式相伴的多项式仍为本原多项式。

定义：如果 $f(x) = \sum_{i=1}^n a_i x^i$ 是唯一分解环 D 上的多项式，如果 (a_0, a_1, \dots, a_n) 的最大公约数为1，则称 $f(x)$ 是 D 上的一个本原多项式。

多重迭代产生器：

$$X_t = (a_1 X_{t-1} + \dots + a_k X_{t-k}) \bmod m, \quad t = k, k+1, \dots,$$

输出函数：

$$U_t = \frac{X_t}{m}$$

[返回](#)

异或运算 $a \oplus b$

a	1	0	1	0
b	1	0	0	1
$a \oplus b$	0	0	1	1

如果 a, b 两个值相同，异或结果为0

如果 a, b 两个值不同，异或结果为1

$$0 \oplus 0 = 1 \oplus 1 = 0$$

$$0 \oplus 1 = 1 \oplus 0 = 1$$

异或结果可以看成两个数相加取2的模，即

$$a \oplus b = (a + b) \pmod{2}$$

[返回](#)

原根

定义： p 为素数, $i \neq j$, i, j 介于 $(1, p-1)$ 之间。若

$$g^i(\bmod p) \neq g^j(\bmod p),$$

则称 g 是 p 的一个原根。

枚举法求解：

例如： 2 不是模 7 的原根， 因为

$$2^3 = 8 \equiv 1(\bmod 7),$$

$$2^3 \equiv 2^6(\bmod 7),$$

$$2^6 = 64 \equiv 1(\bmod 7)$$

3 是模 7 的一个原根， 因为：

$$3^1 \equiv 3(\bmod 7)$$

$$3^2 \equiv 2(\bmod 7)$$

$$3^3 \equiv 6(\bmod 7)$$

$$3^4 \equiv 4(\bmod 7)$$

$$3^5 \equiv 5(\bmod 7)$$

$$3^6 \equiv 1(\bmod 7)$$

游程

一串位置相连，数值相同的元素构成一个游程，其中元素的个数称为游程的长度。

例如：有0和1两种元素构成的系列中，只有0游程和1游程两种类型。

系列：	<u>0 0</u>	1	0	1	<u>0 0 0</u>	<u>1 1 1</u>	0
游程：	1	2	3	4	5	6	7

共有7个游程，4个0游程（两个长度为1,1个长度为2,1个长度为3），3个1游程（两个长度为1，1个长度为3）

[返回](#)