

一种新的随机数组合发生器的研究

王萍¹, 许海洋²

(1. 大连理工大学 应用数学系, 辽宁 大连 116024;

2. 首都师范大学 信息工程学院, 北京 100037)

摘要:提出基于 Mersenne Twister 法和素数模乘同余法的随机数组合发生器。针对目前组合发生器理论多是对线性同余类中不同方法的组合的情况, 结合两种类型的发生器的优点, 得到一种新的随机数发生器。实验表明得到的组合发生器具有相对优越的统计性质, 均匀性和独立性都有很大的提高。

关键词:Mersenne Twister; 随机数; 计算机模拟; 组合随机数发生器

中图分类号:O242.1; TP391.9

文献标识码:A

文章编号:1005-3751(2006)04-0079-03

Study on a New Kind of Combined Random Number Generator

WANG Ping¹, XU Hai-yang²

(1. Department of Applied Mathematics, Dalian University of Technology, Dalian 116024, China;

2. Information Engineering College, Capital Normal University, Beijing 100037, China)

Abstract: In the paper, a combined number generator based on the method of Mersenne Twister and prime modulo multiplicative linear congruence generators is proposed. At present, combined generator theory is in-depth only for linear congruence. Try to combine two type of generators, then a new method of combined generator can be achieved. The experiments show that the combined generator obtained has more superior statistical characteristics. Good uniformity and independence will be achieved.

Key words: Mersenne Twister; random number; computer simulate; combined random number generator

0 引言

随机数序列在仿真工作中使用相当普遍。在计算机中广泛利用数学方法模拟随机过程来产生随机数, 同物理方法产生的随机数相比, 其产生的随机序列容易受到初始值的影响, 且存在周期性, 不是真正的随机数, 所以称之为伪随机数(Pseudo-random number)。

常见的产生伪随机数的方法有多种^[1], 如线性同余法(LCG, Linear Congruent Generators)、Fibonacci 法、Tausworthe 法。为了克服经典方法的缺陷, 近十多年来人们研究了多种新的随机数产生方法, 如非线性同余法、进位加-错位减法(AWL-SWB)及乘子和增量也在递推中变化的复合素数发生器法等。但是单个发生器产生的随机数还存在或多或少的缺陷, 为此, 人们提出组合单个的发生器来产生随机数的观点。文中尝试把不同类型的发生器进行组合, 即用属于线性同余类的素数模乘同余发生器和属于模 2 递推类的 Mersenne Twister 发生器构造组合发生器。实验证明新发生器产生的序列具有更优越的统计性质。

1 理论基础

将两个随机数发生器组合是为了用另一个数列来改变原有随机数序列的排列顺序, 以减小数列元素之间的相关性。文中用于组合的发生器属于两种不同的类型, 在产生随机数方面有着各自不同的优点。

1.1 素数模乘同余法

素数模乘同余法(也称为素数模积式 LCG, 记为 PMMLCG), 是由 Hutchinsom 提出的, 也是目前应用最广泛的均匀随机数发生器^[2]。素数模乘同余法是线性同余发生器的一种, 其算式如下:

$$\begin{cases} X_n = aX_{n-1} \pmod{M} \\ r_n = X_n / M \end{cases} \quad (1)$$

其中 X_n, M, a 均为非负整数。 M 为模数, 取为小于 2^L 的最大素数; a 为乘子, 取为 M 的素元, 而且要尽量大, 这样取法可以保证周期最大为 $M-1$ 。

素数模乘同余法在计算机上较容易实现, 而且运算量小。目前很多计算程序都使用素数模乘同余法产生随机数, 但是由于其随机数产生的局限性, 相邻随机数间存在长周期相关现象^[3], 影响了计算结果的可信度, 另外, 线性同余序列的最大缺陷是高维稀疏网格结构。文中素数模乘同余算法中选取的参数为: $X_n = 764261123X_{n-1} \pmod{M}$, 其中 $M = 2^{31} - 1$, 此时周期为 $T_1 = 2^{30} - 2$ 。

收稿日期: 2005-07-23

作者简介:王萍(1981-), 女, 山东烟台人, 硕士研究生, 研究方向为随机数、概率论、蒙特卡罗模拟; 张宏伟, 副教授, 硕士生导师。

1.2 Mersenne Twister 法

以往的模 2 类的发生器的特征多项式大都为本原三项式,因此产生的随机序列的随机性比较差,有很大的相关性,例如 Ferrenberg 在统计物理学的 Ising 模型的模拟试验中就因为 GFSR 序列间的相关性而得到了错误的结果^[4]。文中用到的 Mersenne Twister 法是 GFSR 法的一种推广的方法。

Mersenne Twisters(简称 MT)是 1998 年由 Makoto 和 Takuji 提出的^[5],递推公式为:

$$X_{k+n} = X_{k+m} \oplus (X_k^r \mid X_{k+1}^r)A, (k = 0, 1, \dots) \quad (2)$$

它是模 2 类发生器的一种,即产生二进制的随机数位,从而构成随机数。

MT 算法产生随机数序列 $\{X_i\}$, 其中 $X_i = (X_{i,w-1}, X_{i,w-2}, \dots, X_{i,0})$ 是 F_2 上的 w 位的字向量。然后将 $\{X_i\}$ 除以 $2^w - 1$, 就生成了 $[0, 1]$ 上的均匀随机序列 $\{r_i\}$ 。对于这个算式要做的解释就是: 整数 n , 递推式的阶数; 整数 r (隐藏于 X_k^r 的定义中), $0 \leq r \leq w - 1$; 整数 m , $1 \leq m \leq n$; A 为 F_2 域上 $w \times w$ 阶常数矩阵。 X_k^r 代表取 X_k 的前 $w - r$ 位, X_{k+1}^r 表示取 X_{k+1} 的后 r 位。 $(X_k^r \mid X_{k+1}^r)$ 代表将 X_k 的前 $w - r$ 位与 X_{k+1} 的后 r 位连接, 从而组成一个新的 w 位的字向量。例如, $(X_k^r \mid X_{k+1}^r) = (X_{k,w-1} \dots X_{k,r}, X_{k+1,r-1} \dots X_{k+1,0})$ 。 \oplus 表示模 2 的位加法运算。

n, m, w, A 都是该序列的周期参数, 为了得到更长的周期, 也为了提高计算速度, 一般都取 A 为:

$$A = \begin{bmatrix} & & 1 & & \\ & & & \ddots & \\ & & & & 1 \\ a_{w-1} & a_{w-2} & \dots & a_0 & \end{bmatrix}$$

如此, XA 只是相当于一个位移运算, 即:

$$XA = \begin{cases} \text{shiftright}(X) & \text{if } X_0 = 0 \\ \text{shiftright}(X) \oplus a & \text{if } X_0 = 1 \end{cases}$$

其中 $X = (X_{w-1}, X_{w-2}, \dots, X_0)$, $a = (a_{w-1}, a_{w-2}, \dots, a_0)$ 。可以看出, 当 $r = 0$ 且 $A = I$ 时, 即为 Tausworthe 序列, 所以 MT 是它的扩展。

MT 算法的特征多项式的展开式大约有 100 多项, 且其本原性易被证明, 因此它拥有很好的随机性。而且 MT 运算速度快、周期长, 可达 $T_2 = 2^{19937} - 1$ 。但是 MT 都有其不足之处, 它产生的随机序列均匀性、独立性就不是很完美。

2 组合发生器

早在 20 世纪 70 年代, Nance, Overstreet 和 Brown, Solomon 就提出了用组合发生器来产生高质量的伪随机数列。Deng 和 George 等人对组合发生器给出了严格的理论分析^[6], 指出: 几个独立且近似均匀的随机变量的线性组合也是一个近似均匀的随机变量, 但其分布比组成它的任一个变量更接近 $U(0, 1)$ 。最重要而且最有实用价

值的组合随机数理论工作归功于 L'Ecuyer 和 Tezuka。

2.1 组合发生器算法

组合发生器是把几个独立的发生器以某种方式组合在一起产生随机数, 常常先用一个随机数发生器产生的随机数列作为基础, 再用另一个发生器对随机数列进行重新排列得到的新数列作为实际使用的随机数, 希望能够比任何一个单独的随机数发生器统计性能更好。该算法的步骤^[2]如下:

1) m 个不同随机数发生器生成随机序列 $\{X_i^{(j)}\} (i = 1, 2, \dots; j = 1, 2, \dots, m)$ 。

2) 令 $c^{(1)}, c^{(2)}, \dots, c^{(m)}$ 为 m 个任意的非零整数, $M^{(j)} (j = 1, 2, \dots, m)$ 为每个发生器的模, 并且为互异的素数。

3) 定义:

$$\begin{cases} U_i = \sum_{j=1}^m \frac{c^{(j)} X_i^{(j)}}{M^{(j)}} \\ u_i = U_i \bmod 1 \end{cases} \quad i = 1, 2, \dots \quad (3)$$

则 $\{u_i\}$ 为 $U(0, 1)$ 随机数。

文中取的是 $m = 2$ 时的情况, 将 MT 算法与乘同余法按照上式进行组合生成新的序列。文中采用将 MT 算法所产生的随机数列作为基础, 用素数模乘同余发生器所产生的随机数列对其进行组合, 新的随机数列统计性质更好。

2.2 组合发生器的周期

假设被组合的各序列 $\{X_i^{(j)}\}$ 的周期为 $\text{Period}(X_i^{(j)})$, $(j = 1, 2, \dots, m)$ 。如果它们两两互素, 且 $\text{Period}(X_i^{(j)})$ 与 $c^{(j)}$ 互素, 这时组合发生器的周期是组合它的 m 个序列周期的最小公倍数 (LCM), 即:

$$\text{Period}\{u_i\} = \text{LCM}(\text{Period}(X_i^{(1)}), \text{Period}(X_i^{(2)}), \dots, \text{Period}(X_i^{(m)}))$$

所以文中的组合发生器的周期 T 为: $T = \text{LCM}(T_1, T_2) = \text{LCM}(2^{19937} - 1, 2^{30} - 2)$, 可以视为无穷大。

3 统计检验

在计算机上用数学方法产生随机序列, 能否有效地应用于仿真中, 主要取决于该序列能否通过随机数的各类统计检验^[7]。一般进行独立性检验、均匀性检验、参数检验, 另外还有组合规律检验等。

3.1 均匀性检验和独立性检验

在随机数检验中, 通常应先进行独立性检验, 因为其他许多统计检验是以独立性为前提的。而且希望得到的是 $(0, 1)$ 区间均匀随机数, 所以均匀性检验也起着决定性的作用。

图 1, 图 2, 图 3 分别是 MT 序列, PMMLCG 序列, 组合发生器序列的二维散布图。不难看出, 图 2 具有线性同余类发生器的缺陷: 高维稀疏网格结构; 图 3 与图 1 相比, 点的分布更为均匀。

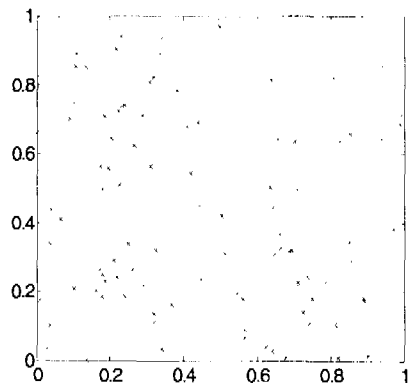


图 1 MT 序列

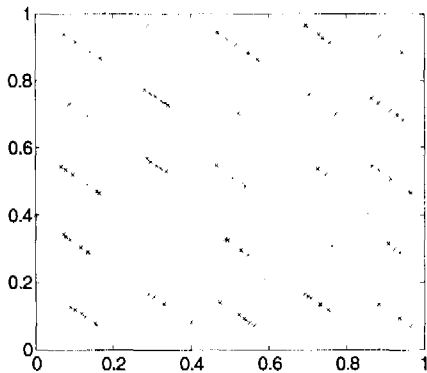


图 2 PMMLCG 序列

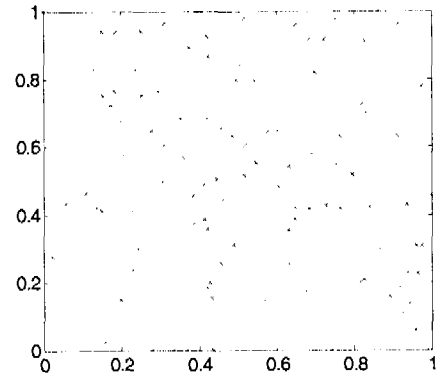


图 3 组合发生器序列

此外,组合发生器和原 MT 各自生成 30000 个数,对其统计性质进行了数据比较。所作的检验均是在显著水平 $\alpha = 0.05$ 时进行的。均匀检验中的卡方检验, $\chi^2(500) > 552.78$ 时可拒绝均匀假设;K-S 检验, $D_n > C_\alpha = 1.35$ 时可拒绝均匀假设。独立性检验中的游程检验, $R_n > \chi^2(6) = 12.59$ 时可拒绝独立假设;相关检验, $|\rho_1 \sqrt{n-1}| > u_{\alpha/2} = 1.96$ 时可拒绝检验。相关数据见表 1。

表 1 两种随机数发生器的均匀性检验和独立性检验

	Mersenne Twister $n = 624, m = 397$	组合发生器 $c_1 = 1, c_2 = 2$
$\chi^2(500)$	528	510
K-S 检验(D_n)	0.998994	0.992880
相关系数(ρ_1)	0.000854	0.000314
游程检验(R_n)	5.168730	1.842598

3.2 参数检验和组合规律检验

参数检验即检验随机数的总体分布参数与均匀分布参数是否一致;组合规律检验是检验数列的排列从数值上看是否有规律性。正负连检验, $\chi^2(\text{正连数}) > \chi^2(7) = 14.067$, $\chi^2(\text{负连数}) > \chi^2(7) = 14.067$ 时可拒绝该检验。取组合发生器和原 MT 各自生成 30000 个数进行数据比较,相关数据见表 2。

表 2 两种随机数发生器的参数检验和组合规律检验

	Mersenne Twister $n = 624, m = 397$	组合发生器 $c_1 = 1, c_2 = 2$
最大值	0.999985	0.999986
最小值	0.000022	0.000020
一阶矩	0.498409	0.501104
与 1/2 的差值	0.001591	0.001104
二阶矩	0.331914	0.333762
与 1/3 的差值	0.001419	0.000339
二阶中心矩	0.083533	0.082660
与 1/12 的差值	0.000200	0.000673
$\chi^2(\text{正连数})$	4.807467	4.1568
$\chi^2(\text{负连数})$	3.749867	11.9208

4 结 论

计算机模拟能够成功的关键是在于计算机上实现真正的随机抽样,而随机抽样产生的基础是随机数。

从以上的统计测试的结果可见,文中提出的组合发生器的性能是比较优越的。组合发生器生成的随机数列保持了原 MT 算法生成的随机数列良好的无连贯性,甚至有所提高。与此同时,还大大提高了均匀性和独立性,其周期在实际应用中可以看作是无穷大。由此可见,该组合发生器具有良好的统计性质,符合随机模拟的要求,有利于提高计算机仿真结果的有效性。

参考文献:

[1] 杨自强,魏公毅.综述:产生伪随机数的若干新方法[J].数值计算与计算机应用,2001(3):201-216.
[2] 高惠璇.统计计算[M].北京:北京大学出版社,1995.92-108.
[3] 罗 平.线性同余发生器的缺陷及改进[J].计算机工程,1995,21:295-297.
[4] 张传林,林立东.伪随机数发生器及其应用[J].数值计算与计算机应用,2002(3):188-208.
[5] Matsumoto M, Nishimura T. Mersenne Twister: A 623-Dimensionally Equidistributed Uniform Pseudo-Random Number Generator[J]. ACM Transactions on Modeling and Computer Simulation, 1998, 8(1): 3-30.
[6] Deng L Y, George E O. Generation of Uniform Variate from Several Nearly Uniformly Distributed Variables[J]. Comm Statist Simu, 1990, 19: 145-154.
[7] 周 燕.关于线性同余组合发生器的周期性和统计性质[J].重庆大学学报,2000,23(6):67-70.