文章编号:1000－582x(2000)06－0067－04

关于线性同余组合发生器的周期性和统计性质

周　燕
(重庆三峡学院 电子工程系,重庆　404000)

摘　要:通过数值试验研究线性同余组合发生器产生的伪随机数列的周期性质和统计性质,并提供一种周期很长、独立性和均匀性都很好的伪随机数发生器。

关键词:线性同余发生器;组合发生器;周期;均匀性;独立性
中图分类号:O 211.9　　**文献标识码**:A

线性同余组合发生器是以一个线性同余发生器产生的随机数列为基础,再用另一个线性同余发生器对该随机数列进行重新排列,得到新的数列作为实际使用的随机数。文献[1,2]认为,两个满周期的混合同余发生器构成组合发生器,两个混合发生器的周期都为 $2p$ 时,组合发生器的周期达到 $2^p(2^p-1)$,并且文献[1]中还认为随机数表的长度 $K=2$ 与长度更大的随机数表具有相同的功能。作者在研究这种随机数发生器的过程中,通过具体的数值试验,否定了文献[1,2]的结论,找到了线性同余组合发生器的周期所遵循的规律,研究了线性同余组合发生器的统计性质与随机数表的长度之间的关系。

1　线性同余发生器

目前应用最广泛的随机数发生器之一是线性同余发生器[1,2],简称 LCG(Linear Congruence Generator)或称 LCG 方法。此方法利用数论中的同余运算来产生随机数,故称为同余发生器。同余发生器的算法为:

$$\begin{cases} x_n = (ax_{n-1} + c)(\mathrm{mod}\ M) \\ r_n = x_n/M \qquad (n = 1,2,3,\cdots) \\ \text{初值 } x_0 \end{cases} \tag{1}$$

其中 M 为模数,a 为乘子(乘数),c 为增量,且 xn、a、c、M 均为非负整数。

显然由(1)式得到的 $x_n(n=1,2,\cdots)$ 满足:$0 \leqslant x_n < M$,从而 $r_n \in [0,1)$。应用(1)式产生均匀随机数时,应适当选取参数 a、c、x_0、M,才能得到周期长且随机性好的数列。

在(1)式中,若 $c=0$,则称相应的算法为乘同余法;若 $c \neq 0$,则称相应的算法为混合同余法。

混合同余发生器达到满周期 $T = M = 2^L$ 的算法为

$$\begin{cases} x_n = ((4a+1)x_{n-1} + (2\beta+1))(\mathrm{mod}\ M) \\ r_n = x_n/M \\ x_0 \text{ 为任意非负整数} \end{cases} \tag{2}$$

其中 $n = 1,2,\cdots$,a、b 为任意正整数。

为使统计性质优,参数 a、c、x_0 的选取准则为:

1) x_0 为任意非负整数;

2) a 满足:(i) $a\ (\mathrm{mod}\ 8) = 5$

(ii) $\dfrac{M}{100} < a < M - \sqrt{M}$

(iii) a 的二进制没有明显规律;

3) c 为奇数,且

$$\frac{c}{M} = \frac{1}{2} - \frac{\sqrt{3}}{6} \approx 0.211\ 324\ 865$$

乘同余法达到最大周期的充要条件为:

1) 当 $M = 2^L(L \geqslant 4)$,x_0 为奇数时,则取 $a = 3$ 或 $5(\mathrm{mod}\ 8)$ 且最大周期 $T = 2^L - 2$;

2) 当 $M = 10^s(s \geqslant 5)$,x_0 不是 2 或 5 的倍数时,则取 $a(\mathrm{mod}\ 200)$ 等于以下 32 个值之一:3,11,13,17,21,27,29,37,53,59,67,69,77,83,91,109,117,123,131,133,139,141,147,163,173,179,181,187,189,197 且 $T = 5 \times 10^{s-2}$;

收稿日期:1999-09-17
作者简介:周燕(1958-),女,重庆市人,重庆三峡学院电子工程系讲师,从事计算物理方面研究。

3) 当 $M = p$(p 为素数)时,则取 a 为 M 的素元,且可得到最大周期 $T = M - 1$。

为使乘同余法产生的数列 $\{x_n\}$ 统计性质优,应选择参数 a 值大($a < M$)。

2 组合同余发生器

2.1 算法说明

用线性同余法(LCG 方法)产生随机数,有一些缺陷,主要是该方法产生的均匀随机数作为 m($m > 1$)维均匀随机向量时相关性大;其次是 LCG 方法产生的均匀随机数列的周期 T 与计算机的字长有关。在整数的尾数字长为 L 位的计算机上,不可能得到 $T > 2^L$ 的均匀随机数列。

为克服 LCG 方法的上述缺陷,得到周期更长随机性更好的随机数,利用组合发生器所依据的原则,即打乱数列的次序使之排列不规则,先用一个 LCG 产生的数列为基础,再用另一个 LCG 产生的数列对它进行重新排列,由此得到的新数列作为实际使用的随机数。具体算法为:

① 用第一个 LCG 产生 K 个随机数,组成随机数表,这 K 个随机数被顺序存放在矢量 $\boldsymbol{A} = (t_1, t_2, \cdots, t_K)$ 中;

② 用第 2 个 LCG 产生一个随机整数 j,要求 $1 \leqslant j \leqslant K$;

③ 令 $r_n = t_j$,然后再用第一个线性同余发生器产生一个随机数 y,令 $t_j = y$;

④ 重复 ②、③,得随机数序列 $\{r_n\}$,即为组合发生器产生的数列。

2.2 周期性研究

按线性同余发生器中所述的算法构成多个不同的 LCG,每两个 LCG 构成一个组合发生器,整理出来的数据见表 1。其中:$T1$ 是第一个 LCG 的周期,$T2$ 是第二个 LCG 的周期,K 为矢量 A 的长度,T 为组合发生器的周期。

分析表 1 可得到如下结论:

1) 如果 $K > T1$,且 $T1 < T2$,则 $T = T2$(表 1 中 1 ~ 4 组数据)。

2) 如果 $T1 = T2$,则无论 K 多大,组合发生器的周期都不会大于单独一个同余发生器的周期,且 $T = T1 = T2$(表 1 中 5 ~ 14 组数据)。

3) 如果 $K < T1$,在这种情况下,根据大量的实验数据可总结出其周期遵循如下规律:即组合后的周期 T 等于 $T1$,$T2$ 的最小公倍数除以 G,G 是与 K 或 $T1$ 有关的常数,其表达式为:

表 1 线性同余组合发生器的周期性研究

序号	K	$T1$	$T2$	T
1	600	512	50 000	50 000
2	821	512	50 000	50 000
3	513	512	50 000	50 000
4	600	512	131 069	131 069
5	101	50 000	50 000	50 000
6	13	50 000	50 000	50 000
7	300	512	512	512
8	121	512	512	512
9	600	512	512	512
10	513	512	512	512
11	97	131 069	131 069	131 069
12	98	131 069	131 069	131 069
13	251	131 069	131 069	131 069
14	1 000	131 069	131 069	131 069
15	100	2 048	50 000	1 600 000
16	101	2 048	50 000	6 400 000
17	102	2 048	50 000	3 200 000
18	103	2 048	50 000	6 400 000
19	104	2 048	50 000	1 600 000
20	105	2 048	50 000	6 400 000
21	106	2 048	50 000	3 200 000
22	107	2 048	50 000	6 400 000
23	108	2 048	50 000	1 600 000
24	109	2 048	50 000	6 400 000
25	110	2 048	50 000	3 200 000
26	97	512	131 069	67 107 328
27	251	512	131 069	67 107 328
28	118	512	131 069	33 553 664
29	120	512	131 069	16 776 832
30	500	512	50 000	400 000
31	100	50 000	2 048	6 400 000
32	101	50 000	2 048	6 400 000
33	102	50 000	2 048	6 400 000
34	103	50 000	2 048	6 400 000
35	104	50 000	2 048	6 400 000
36	105	50 000	2 048	6 400 000
37	106	50 000	2 048	6 400 000
38	107	50 000	2 048	6 400 000
39	108	50 000	2 048	6 400 000
40	109	50 000	2 048	6 400 000
41	110	50 000	2 048	6 400 000
42	97	131 069	512	67 107 328
43	120	131 069	512	67 107 328

T = $T1$、$T2$ 的最小公倍数 / G

G 的取值为：

① 当 $T1 \neq 2^p$（p 为正整数）时，G 取 1（表 1 中 31 ~ 43 组数据）。

② 当 $T1 = 2^p$（$p \geqslant 4$）时，又分为下面几种情形：

a. 如 $K(\mathrm{mod}\ 4) = 0$，则 G 取 4（表 1 中 15、19、23、29、30 组数据）。

b. 如 $K(\mathrm{mod}\ 4) \neq 0$，$K(\mathrm{mod}\ 2) = 0$，则 G 取 2（表 1 中 17、21、25、28 组数据）。

c. K 取其它数值时，G 取 1（表 1 中 16、18、20、22、24、26、27 组数据）。

根据以上讨论，只要 K 为小于 $T1$ 的奇数，$T1$、$T2$ 互素，就可以得到最大周期：$T = T1 \cdot T2$。

2.3 统计性质研究

在试验中选用了目前常用的两个用乘同余法产生的随机数序列 RAND1 和 RAND2 进行组合。对于 RAND1，其算法为[1]：

$$x_n = 1\ 220\ 703\ 125 x_{n-1} (\mathrm{mod}\ M),\ M = 2^{31}$$

对于 RAND2，其算法为：

$$x_n = 16\ 807 x_{n-1} (\mathrm{mod}\ M),\ M = 231$$

表 2 中分别列出了两个乘同余发生器 RAND1 和 RAND2 以及两个组合同余发生器 RANDCOM1 和 RANDCOM2 的周期和统计检验[3,4]结果。其中 RANDCOM1 是以 RAND1 作为组合发生器算法中的第一个 LCG，RANDCOM2 是以 RAND2 作为组合发生器算法中的第一个 LCG。在试验中由于 RANDCOM1 和 RANDCOM2 的周期太长，其值是按 2.2 中总结的规律计算出来的。

由表 2 可以看出，打乱后的随机数列和打乱前的随机数列相比，独立性变化很大，其它性质变化较小。经过一系列的数值计算表明，独立性是否得到改善，显著的依赖于向量长度 K 的选取。K 对线性同余组合发生器独立性的影响见表 3。

由表 3 可以看出，原来独立性较好的随机数序列 RAND1，一般来说，打乱后的随机数序列独立性会变差；原来独立性较差的随机数序列 RAND2，一般来说，打乱后的随机数序列独立性会变好。

表 2　线性同余组合发生器的统计性质

	RAND1	RAND2	用 RAND2 打乱 RAND1 得到 RANDCOM1（向量长度 K = 108）	用 RAND1 打乱 RAND2 得到 RANDCOM2（向量长度 K = 129）
初 值	1	1	1	1
周 期	536 870 912	2 147 483 646	144 115 187 941 638 144*	576 460 751 766 552 576**
随机数个数	10 000	10 000	10 000	10 000
分组区间数	500	500	500	500
最大值	9 999 646E－01	9 999 999E－01	9.999 646E－01	9.999 999E－01
最小值	4.941 085E－05	3 903 639E－06	4 941 085E－05	3 903 639E－06
一阶矩	5 041 382E－01	5.018 268E－01	5.042 887E－01	5.018 648E－01
二阶矩	3 380 003E－01	3.354 743E－01	3 380 862E－01	3 355 567E－01
二阶中心矩	8.386 207E－02	8.364 746E－02	8.379 757E－02	8.369 187E－02
正连数	5 054	5 043	5 056	5 048
正连数所占百分比	5.054 000E－01	5 043 000E－01	5.056 000E－01	5.048 000E－01
χ^2	534 30	447 90	538.50	448.50
独立性检验 ρ	3.589 257E－02	1 443 892	1.973 430E－03	2 109 616E－02

*，**：组合发生器的周期是通过前述规律计算得到的。

表 3 向量长度 K 对线性同余组合发生器独立性的影响

K	RANDCOM1 独立性检验 ρ	RANDCOM2 独立性检验 ρ
102	1.87 独立性大大变坏	0.316 独立性大大改善
108	0.00197 独立性大大改善	0.300 独立性大大改善
110	0.013 独立性稍有改善	0.396 独立性大大改善
117	0.313 独立性变坏	1.76 独立性变坏
121	0.759 独立性变坏	0.592 独立性改善
124	0.802 独立性变坏	2.34 独立性变坏
129	0.149 独立性变坏	0.0211 独立性大大改善

3 结论和展望

1）数值计算结果表明，与原来的 LCG 数列相比，线性同余组合发生器产生的数列的周期是否会增大、依赖于两个 LCG 序列的周期 T1、T2 和向量长度 K 三者之间的关系，与文献[1,2]中的结论不同。

2）数值计算结果表明，与原来的 LCG 数列相比，线性同余组合发生器产生的数列的独立性是否会改善，与向量长度有很大的关系，与参考文献[1]中的结论不同。

3）以上结论仅仅是通过数值计算归纳总结而成，有待于进一步从理论上加以探讨。

4）在 2.3 节中提到的线性同余组合发生器 RANDCOM2 通过统计检验（表 2），表明具有很好的统计性质，它克服了 LCG 算法的缺陷，均匀性好，独立性高，周期特别长，对所有的实用场合来说，周期可以认为是无穷长，具有较大的实用价值。

参考文献：

[1] 高惠璇. 统计计算[M]. 北京：北京大学出版社，1996. 85-108.

[2] 程兴新、曹敏. 统计计算方法[M]. 北京：北京大学出版社，1989. 26-45.

[3] 中国科学院计算中心概率统计组[M]. 概率统计计算. 北京：科学出版社，1983. 391-399.

[4] 刘德贵. FORTRAN 算法汇编（第二分册）[M]. 北京：国防工业出版社，1983. 481-493.

On the Research of the Periodicity and Statistical Characteristics by Linear Congruence and Combined Generator

ZHOU Yan

(Electronic Engineering Department, Chongqing Three Gorges College, Chongqing 404000, China)

Abstract: In this paper the author researches the periodicity and statistical characteristics produced by linear congruence and combined generator through numerical value tests, and puts forward a random numeral generator with a long periodicity, a good independence and homogeneity.

Keywords: Linear congruence generator; combined generator; periodicity; homogeneity; independence

（责任编辑 吕幕英）