一类长周期的伪随机数序列[*1)]

皮新明

(武汉理工大学)

A SORT OF PSEUDORANDOM NUMBER SEQUENCES WITH EXTREMELY LONG PERIOD

Pi Xinming

(*Wuhan University of Technology, Wuhan*)

Abstract

This paper deals with the period T of pseudorandom number sequence generated by subtract-with-borrow (SWB). For the selected base b, legs $r > s$ it is shown that T is dependent only on b, r and s and equal to the order of b with respect to $m = b^r - b^s + 1$ provided that m and A, which is determined by initial values $x_1, \cdots, x_r$ and initial borrow value c_{r+1}, are coprime. For $b = 2^{31} - 1$ this paper searches the field $0 < r \leq 300$, $r - s \leq 10$ and finds out all the pairs of (r, s) for which m is prime. In addition, the periods of SWB corresponding to $b = 2^{31} - 1$ and these pairs (r, s) are also calculated. Besed on data obtained, generators of SWB with extremely long period could be designed.

Key words: pseudorandom number sequence, subtract-with-borrow, period, prime number, order

§ 1. 引　　言

近年来, 随着计算技术的迅猛发展, 在计算机上利用 Monte-Carlo 方法 (亦称随机模拟) 解决科学及工程实际问题已经成为现实. 不过对此还要涉及到随机数 $\xi \sim (0,1)$. 然而生成真正的 $\xi \sim (0,1)$ 却十分困难. 一般都是依照一定的规则在计算机上生成伪随机数序列取代 $\xi \sim (0,1)$、以用于随机模拟工作.

为获取品质优良的伪随机数序列以提高随机模拟结果的有效性, 人们不断地提出多种伪随机数发生器即生成方法 [1]. 特别地, 九十年代初由 Marsaglia 等 [2] 提出的进位加法 (add-with-carry) 与借位减法 (subtract-with-borrow) 更是因其既具有有效性又具有极长的周期而引起人们的关注 [3].

* 1999 年 9 月 30 日收到.

1) 湖北省自然科学基金资助项目.

伪随机数发生器的周期就是该发生器 (即生成方法) 生成的伪随机数序列的周期, 长周期是评价伪随机数发生器品质的重要指标之一 [4], 周期的探讨对伪随机数发生器的研究极具价值.

在 $s<r, m=b^r-b^s+1$ 是素数, 并且 b 是 m 的原根的条件下, Marsaglia 等 [2] 正确地指出了借位减法的周期是 $m-1$. 不过从实用的角度而言, b 是 m 的原根这一条件是过于苛刻了.

本文讨论了借位减法的周期问题, 在宽松得多的条件下导出了借位减法的周期. 文中还给出了数字实例, 所得数据可供用于设计具有极长周期的借位减法发生器.

§2. 关于借位减法

在本文中记号 $\mathcal{N}$ 表示自然数集. 借位减法的基本步骤如下:

1. 选定基数 $b\in\mathcal{N}$, 步长 $r,s\in\mathcal{N}$, 其中 $r>s$.

2. 选定 $[0,b-1]$ 中的 r 个整数 $x_1,\cdots,x_r$ 作为初值. 一般而言, 这 r 个初值不全选取为 0, 也不全选取为 $b-1$.

3. 选定一个 $c_{r+1}\in\{0,1\}$ 作为初始借位值.

4. 对 $n>r$, 令

$$t_n=x_{n-s}-x_{n-r}-c_n, \tag{1}$$

$$\begin{cases} \text{若 } t_n\geq 0, & \text{则取 } x_n=t_n, c_{n+1}=0, \\ \text{若 } t_n<0, & \text{则取 } x_n=t_n+b, c_{n+1}=1. \end{cases} \tag{2}$$

重复步骤 4 可依次得到 $x_{r+1},x_{r+2},\cdots$. 由此即生成借位减法伪随机整数序列

$$\{x_n,\ \ n\in\mathcal{N}\}. \tag{3}$$

基于序列 (3) 即易生成借位减法伪随机数序列.

对于序列 (3), 若有 $t\in\mathcal{N}$, 满足

$$x_n=x_{n+t},\ n=n_0+1,\ n_0+2,\cdots,$$

则称 t 为 (3) 的一个周期. 当 $n_0>0$ 时, 称 n_0 为暂态值 (transient). 序列 (3) 的周期之最小者称为它的最小周期, 记为 T. 在不会发生误会的情况下, 最小周期亦可简称为周期.

关于序列 (3) 的周期, 文 [2] 中提出一个结论, 如下面的定理 1 所述.

定理 1. 如果 $m=b^r-b^s+1$ 是素数且 b 是 m 的一个原根, 则序列 (3) 的周期 $T=m-1$.

在本文中我们将在宽松得多的条件下扩展上述结果. 首先, 由式 (1) 及 (2) 知

$$x_n=x_{n-s}-x_{n-r}-c_n+c_{n+1}\cdot b,\ n=r+1,r+2,\cdots. \tag{4}$$

§3. 基本结果

为说明方便，先证明一个引理.

引理. 序列 (3) 的周期存在.

证明. 对于 $j \in \mathcal{N}$, 称 (3) 中的自第 j 项起的连续 r 项

$$x_j, x_{j+1}, \cdots, x_{j+r-1} \text{ 以及 } c_{j+r}$$

为一个 r- 序段，记为 $s(j)$. 显然 $s(j_1) = s(j_2)$ 等价于

$$x_{j_1+k} = x_{j_2+k}, k = 0, \cdots, r-1, \text{ 且 } c_{j_1+r} = c_{j_2+r}.$$

因此 (3) 中不相等的 r- 序段至多 $2 \cdot b^r$ 个，故必有 $0 < j_1 < j_2$ 使 $s(j_1) = s(j_2)$. 根据借位减法基本步骤 4 知

$$x_{j_1+k} = x_{j_2+k}, \ k = 0, 1, \cdots. \tag{5}$$

记 $t = j_2 - j_1$, $n_0 = j_1 - 1$, 则式 (5) 可表达为

$$x_n = x_{n+t}, \ n = n_0 + 1, \ n_0 + 2, \cdots. \tag{6}$$

式 (6) 表明 t 为 (3) 的一个周期. 引理证毕.

由引理知序列 (3) 的最小周期 T 存在. 以下约定：序列 (3) 的周期即指它的最小周期 T. 另外，不失一般性可以认为 $n_0 = 0$. 此至多相当于屏弃 (3) 中前面的有限项，并不对 (3) 的基本性质产生任何影响. 且对于 $k < r+1$, 令 $c_k = c_{k+T}$, 对于 $k < 0$, 令 $x_k = x_{T+k}$, 则式 (4) 的成立范围可以扩展，即有

$$x_n = x_{n-s} - x_{n-r} - c_n + c_{n+1} \cdot b, n = 1, 2, \cdots. \tag{4'}$$

对于序列 (3), 记

$$P_n = \sum_{k=1}^{n} x_k b^{k-1}, n \in \mathcal{N}.$$

本文的基本结果如以下定理 2 所示.

定理 2. 假定 $m = b^r - b^s + 1$ 与 $A = b^s P_{r-s} - P_r + c_{r+1} b^r$ 互素，以 m 为模 b 的次数 (*order*) 为 L. 如果初值 $x_1, \cdots, x_r$ 不全取为 0, 也不全取为 $b-1$, 则序列 (3) 的周期为 $T = L$.

证明. 由于 T 为序列 (3) 的周期，故由式 (6) 及 $n_0 = 0$ 知

$$P_{T+n} = P_T + b^T P_n, \ n \in \mathcal{N} \tag{7}$$

经简单的代数运算可以得到

$$b^s (P_{T+r-s} - P_{r-s}) = \sum_{k=r+1}^{r+T} x_{k-s} b^{k-1} \tag{8}$$

以及

$$b^r P_T = \sum_{k=r+1}^{r+T} x_{k-r} b^{k-1}. \tag{9}$$

从式 (8) 减去式 (9) 并利用 (4′) 与 (7) 可得到

$$(b^s - b^r)P_T + b^s(b^T - 1)P_{r-s} = P_{T+r} - P_r - c_{r+1}b^r(b^T - 1),$$

或表达为

$$mP_T = (b^T - 1)A. \tag{10}$$

因初值 $x_1, \cdots, x_r$ 不全为 0, 也不全为 $b-1$, 由式 (10) 知 $0 < A < m$. 注意到 $(m, A) = 1$, 由 (10) 可导出 $m|(b^T - 1)$, 从而成立

$$L|T. \tag{11}$$

下面往证 $T|L$. 用记号 $0.X_1 \cdots X_n$ 表示 b 进位制下的小数, 即 $0.X_1 \cdots X_n = \sum_{k=1}^{n} X_k \cdot b^{-k}$, 这里 $X_k \in [0, b-1]$ 为整数, $k = 1, \cdots, n$, $n \in \mathcal{N}$. 特别对于序列 (3) 有

$$0.x_n \cdots x_1 = b^{-n} P_n, \ n \in \mathcal{N} \tag{12}$$

因 T 是 (3) 的周期, 所以在 b 进位制下的循环小数 $0.\dot{x}_T \cdots \dot{x}_1$ 的循环节的长度必定为 T.

利用式 (10) 和 (12) 可以得到

$$0.\dot{x}_T \cdots \dot{x}_1 = 0.x_T \cdots x_1 \cdot \frac{b^T}{b^T - 1} = \frac{A}{m} \tag{13}$$

另一方面, 既然 L 是 b 关于模数 m 的次数, 既约真分数 $\dfrac{A}{m}$ 就可以展开为 b 进位制下的纯循环小数 [5], 即有

$$\frac{A}{m} = 0.a_1 \cdots a_L a_{L+1} \cdots, \text{ 而且 } a_k = a_{k+L}, \ k \in \mathcal{N}. \tag{14}$$

比较 (13) 与 (14) 知

$$T|L. \tag{11′}$$

由 (11) 和 (11′) 即可得到 $T = L$, 当然还有 $a_k = x_{T-k+1}$, $k = 1, \cdots, T$. 不过它们对于定理 2 的证明并不重要. 到此定理 2 证毕.

推论. 假定 $m = b^r - b^s + 1$ 是素数, 且 b 对模 m 的次数是 L. 如果初值 $x_1, \cdots, x_r$ 不全为 0, 也不全为 $b-1$, 则序列 (3) 的周期为 $T = L$.

实际上, 定理 2 还可进一步推广.

定理 3. m, A 如定理 2 所述, 假定 $(m, A) = d$, $m' = m/d$, 以 m' 为模 b 的次数为 L. 如果初值 $x_1, \cdots, x_r$ 不全取为 0, 也不全取为 $b-1$, 则序列 (3) 的周期为 $T = L$.

对定理 2 的证明略加改动，即得到定理 3 的证明，故此处从略.

十分明显，定理 1 只是定理 2 推论的一个特例.

顺便指出，在利用上述借位减法生成伪随机数序列时，出于某种考虑人们往往取基数 b 为素数. 例如文 [2] 中取 $b = 2^{32} - 5$, 并得出 $r = 43$, $s = 22$ 时 m 为素数且 b 为 m 的一个原根. 然而从定理 2 的证明可知，b 不必是素数.

§4. 初值组与序列 (3) 的关系

现讨论在定理 1 与定理 2 的不同条件下初值组的选取对序列 (3) 结构的影响. 对选定的步长 r, 初值组共有 b^r 种不同的选取方式.

在定理 1 的条件下，由于有 $T = m - 1$, 序列 (3) 中进入了周期状态的那部分 (即循环部分) 共存在 $m - 1$ 种不同形式的 r- 序段. 如果 $c_{r+1} = 0$, 则有以下三种情况:

1. 当 r 个初值全被取为 0 时，有 $x_n = 0, n \in \mathcal{N}$, 序列 (3) 实为零序列;

2. 有 $b^r - b^s$ 种初值组将作为 r- 序段出现在序列 (3) 的循环部分中;

3. 有 $b^s - 1$ 种初值组将不满足或不完全满足周期规律，因而将不会以 r- 序段的形式出现在序列 (3) 的循环部分中.

不难理解，不同的初值组对应的序列 (3) 的差异仅在于循环部分的起点不同. 因此当 $c_{r+1} = 0$ 时，序列 (3) 的结构与初值组的选取无实质关系. 这实际上就是文 [2] 所说的 "理想序列".

如果 $c_{r+1} = 1$, 则有类似的情况出现. 唯一的不同点是当 $x_1 = \cdots = x_r = b - 1$ 时，序列 (3) 成为 $x_n = b - 1$, $(n \in \mathcal{N})$ 的常数序列.

如果 m 为素数，则在定理 2 的条件下，有 $T|(m-1)$. 故有 $k \in \mathcal{N}$ 使 $T = (m-1)/k$. 于是存在 k 类不同的序列 (3), 每一类称为一条 "轨道", 同一轨道中的序列仅循环的起点不同，它们所含的 r- 序段及其次序实际上完全相同; 不同的轨道互不相交. 在所有的轨道中共有 $m - 1$ 种不同的 r- 序段出现.

§5. 算法及数字实例

在定理 2 、定理 3 中，并不强求 m 为素数，不过当 m 为合数时，在应用上存在两方面的困难. 一是当 $n_0 > 0$ 时，不能直接从初值 $x_1, \cdots, x_r$ 及 c_{r+1} 求出 A; 二是即使求出了 A, 为了得出 b 对模 m 或 m' 的次数，要用到 m 或 m' 的完全分解. 即使 m 或 m' 仅为百余位的数，到目前为止，这也往往是十分困难甚至完全不可能的. 因此在使用借位减法时还是希望选取 m 为素数. 为此就需要对模数 m 的素性作出判定. 下面的定理 4 及定理 5 对于自然数的素性判定十分有用.

定理 4[5] (费尔马 (Fermat) 小定理). m 为素数的必要条件是若 $(a, m) = 1$, 则 $a^{m-1} \equiv 1 \pmod{m}$.

定理 5[6]. 假设 $m - 1 = \prod\limits_i p_i^{\alpha_i}$, 其中 p_i 为素数，$\alpha_i \in \mathcal{N}$. 如果对于每一个 p_i 都存在 $a_i \in \mathcal{N}$, 使 $a_i^{m-1} \equiv 1 \pmod{m}$, $a_i^{(m-1)/p_i} \not\equiv 1 \pmod{m}$, 则 m 是素数.

为了应用上的方便，本文取基数 $b=2^{31}-1$，根据定理 4 和定理 5 在范围 $r-s\leq 10$，$r\leq 300$ 中系统地搜寻了使 $m=b^r-b^s+1$ 为素数的全体 (r,s) 数对.

搜寻素数 m 的基本算法如下：

1. 取定 $k=r-s=1,\cdots,10$，依次对 $r=k+1,\cdots,300$ 检验 $b^{m-1}\equiv 1\ (\mathrm{mod}\ m)$ 是否成立. 若不成立，则 m 为合数，放弃该对 (r,s).

2. 若上式成立，则 m 是以 b 为底的概素数. 此时需要作出 $m-1$ 的完全分解：$m-1=\prod_i p_i^{\alpha_i}$. 若对每个 p_i 均有 $b^{(m-1)/p_i}\not\equiv 1\ (\mathrm{mod}\ m)$，则 m 是素数，并且 b 是 m 的一个原根. 记录下该对 (r,s).

3. 若至少存在一个 p_i，使 $b^{(m-1)/p_i}\equiv 1\ (\mathrm{mod}\ m)$，则利用定理 4 与定理 5 继续检验 m 的素性. 通常的做法是：

I. 记 $P=\{p|b^{(m-1)/p}\equiv 1\ (\mathrm{mod}\ m),\ p\ \text{为素数}\}$；

II. 取 $a=2$，检验

$$a^{m-1}\equiv 1\ (\mathrm{mod}\ m) \tag{15}$$

及

$$a^{(m-1)/p}\not\equiv 1\ (\mathrm{mod}\ m) \tag{16}$$

是否对每个 $p\in P$ 都成立. 若式 (15) 不成立，则 m 为合数，中止运算；否则删去 P 中使式 (16) 成立的那些 p，所剩下的集仍记为 P；

III. 若 P 不是空集，再依次取素数 $a=3,5,\cdots$，转入 II，直到 P 为空集或有某一个素数 a，使式 (15) 不成立为止. 当 P 为空集时，由定理 5 知 m 为素数，记录下该对 (r,s)；若出现后一种情况，则 m 为合数，放弃该对 (r,s).

若 m 是素数，则 b 不是 m 的原根，然后求出 b 对 m 的次数.

现将搜寻的结果列于表 1.

表 1　部分数对 (r,s) 与周期 T 之值

数对 (r,s)	周期 T	数对 (r,s)	周期 T	数对 (r,s)	周期 T
(5, 4)	$m-1$	(40, 31)	$m-1$	(136, 127)	$m-1$
(8, 2)	$(m-1)/8$	(73, 68)	$m-1$	(178, 169)	$(m-1)/3$
(22, 16)	$(m-1)/84$	(78, 70)	$(m-1)/4$	(276, 275)	$(m-1)/3$

顺便指出，检验 m 是否为概素数的运行时间 $t=O(\log_2^3 m)=O(r^3\log_2^3 b)$. 作者在 K6-233 机上对任一指定的 $k=r-s$，$1\leq k\leq 10$，$r\leq 300$ 搜寻概素数 m 耗费约 10 小时，因而总共耗费约 100 小时得出了以上结果.

例如，为了验证 $m=b^{136}-b^{127}+1$ 是素数且 b 是 m 的一个原根，需要用到 $m-1$ 的完全分解. 由于这里 b 是素数，只要得出 b^9-1 的完全分解即可. 作者得出

$$b^9-1=2\cdot 3^4\cdot 7\cdot 11\cdot 31\cdot 151\cdot 331\cdot 613\cdot 529510939\cdot 2903110321\cdot p,$$

此处 $p=5333\cdots 6127$ 是一个 53 位的素数.

容易看出，随着 $r-s$ 的增大， $m-1$ 的完全分解越来越困难. 尤其是 $r-s$ 为素数或具有较大的素因子时，这种困难将变得更加严重.

§6. 结　论

借位减法是进入九十年代以来引起广泛关注的伪随机数生成方法之一. 本文在非常宽松的条件下导出了借位减法的周期，如定理 2 、定理 3 所述.

对 $b=2^{31}-1$, 本文得出了 9 对使 $m=b^r-b^s+1$ 为素数的 (r,s), 其中有 4 对 (r,s) 使得 b 是 m 的原根. 所得结果表明：即使 b 不是 m 的原根，序列 (3) 的周期 T 仍可以表达为 $T=(m-1)/k$, 而且 k 也常为相当小的自然数. 对于数对 $r=178, s=169$, 有 $T\approx 4.04012\times 10^{1660}$, 对于数对 $r=276, s=275$, 有 $T\approx 1.366223\times 10^{2575}$.

在定理 2 的条件下，序列 (3) 的周期 T 与初值 $x_1,\cdots,x_r$ 和初始借位值 c_{r+1} 无关. 但 $x_1,\cdots,x_r$ 和 c_{r+1} 对序列 (3) 的统计性质和结构的影响以及 b,r,x 的选取对序列 (3) 品质的影响却需要作进一步探讨，以期为随机模拟工作提供品质优良的伪随机数序列.

作者对罗平教授的帮助表示衷心的感谢.

参 考 文 献

[1] P.L′ Ecuyer, Random numbers for simulation, Commun. ACM33., 10 (1990), 86–97.

[2] G.Marsaglia, B.Narasimhan, A.Zaman, A random number generator for PC's, Comput. Phys. Commun., 60 (1990), 345–349.

[3] B.Couture, P.L′ Ecuyer, Distribution properties of multiply-with-carry random number generators, Math. Comput., 66 (1997), 591–607.

[4] B.D. Ripley, Thoughts on pseudorandom number generators, J.Comput.App.Math., 31 (1990), 153–163.

[5] 陈景润, 初等数论, 科学出版社, 1980.

[6] J.Brillhart, D.H. Lehmer, J.L. Selfridge, New primality criteria and factorizations of $2^n\pm 1$, Math. Comput., 29 (1975), 620–647.