Available online at www.sciencedirect.com

SCIENCE @ DIRECT•

ELSEVIER

Mathematical and Computer Modelling 42 (2005) 809–816

MATHEMATICAL
AND
COMPUTER
MODELLING

www.elsevier.com/locate/mcm

A Generator of Pseudo-Random Numbers Sequences with a Very Long Period

S. SÁNCHEZ AND R. CRIADO
Dpto. de Matemáticas y Física Aplicadas y CC. de la Naturalcza
Universidad Rey Juan Carlos
E-28933 Móstoles, Spain
<ssanchezg><rcriado>@escet.urjc.es

C. VEGA
Dpto. de Matemática Aplicada a las Tecnologías de la Información
Universidad Politécnica de Madrid
E-28040, Madrid, Spain
cvega@mat.upm.es

Abstract—We show how it is possible to choose, in an adequate manner, the parameters a, b, and m of the linear congruential generator,

$$x_{n+1} = (a \cdot x_n + b) \bmod m,$$

in order to maximize the period of the generated series of pseudo-random numbers, in such a manner that the period obtained is close to the theoretical upper bound of $m!$. © 2005 Elsevier Ltd. All rights reserved.

Keywords—Pseudo-random number sequences, Linear congruences.

1. INTRODUCTION

In many scientific fields, it is necessary to use sequences of pseudo-random numbers to develop techniques such as location of primary sound by means of radars, cryptographic procedures, or Monte Carlo simulations. There are many procedures to generate such sequences, see [1] for a review, which apart from appropriate randomness properties should have a long enough period for the incumbent application. The sequence obtained from the linear congruential generator,

$$x_{n+1} = (a \cdot x_n + b) \bmod m,$$

has properties that make it specially attractive for an extensive use, both from the statistical point of view and due to the possibility of obtaining sufficiently long periods if the parameters

0895-7177/05/$ - see front matter © 2005 Elsevier Ltd. All rights reserved. Typeset by $\mathcal{A}_{\mathcal{M}}\mathcal{S}$-TEX
doi:10.1016/j.mcm.2005.09.009

a, b, m are conveniently chosen, see [1]. A key property resides in the fact that the number of operations needed per bit generated is small. On the other hand, its main disadvantage is that the generated sequences are predictable [2,3], suggesting that this generator is not adequate for cryptographic purposes. However, in those applications where random behavior is simulated, such as the search of big prime numbers, where it is necessary to check if a number is composed or "probably" prime, this generator can be certainly used. In this paper, we shall present a generator that allows us to overcome such faults, because its configuration is capable of breaking the natural order in the generation of random numbers and allows us to approach a large period length, close to the theoretical upper bound of $m!$ and, hence, higher than that of combined generators, [4]. Moreover, as it is a nonlinear generator, it defies analytic attack.

2. BASIC RESULTS AND PRELIMINARIES

Let us recall that if F is a set with a finite number of elements, a generator of F is an algorithm that obtains a sequence of elements of F. A sequence $\{x_n\}_{n\geq 0}$ is periodic if there exists k, such that $x_{n+k} = x_n$ for all $n \in \mathbb{N}$. For a periodic sequence, the set of numbers k satisfying the above condition constitutes a subset of $\mathbb{N}-\{0\}$. If λ is the smallest element of that subset, the subsequence $x_0, x_1, \ldots, x_{\lambda-1}$ is called a period of $\{x_n\}_{n\geq 0}$ and λ is called the length of that period. We say that $\{x_n\}_{n\geq 0}$ is almost periodic if there exists $m \in \mathbb{N}$, such that the sequence $\{x_n\}_{n\geq m}$ is periodic. In this case, the smallest number μ satisfying this condition is called the entrance index to the period. It is said that $\{F, f, x_0\}$ is of maximum period if the period length is equal to $|F|$.

In the sequel, we shall concentrate in single-step generators, that is, those that can be written as $x_{n+1} = f(x_n)$, where f is a mapping $f : F \longrightarrow F$ and F is a finite set. We shall interpret the generator of F as an algorithm that produces the sequence $\{x_n\}_{n\geq 0}$ of elements of F and we shall denote it by $\{F, f, x_0\}$.

As the set F is finite, there are h, k, with $h < k$, such that $x_h = x_k$. Applying f, we have that $x_{h+r} = x_{k+r}$, therefore, $k - h$ is the period of the sequence $\{x_i\}_{i\geq h}$. If $x_i = x_j$ for $j > i$, as $\{x_l\}_{l\geq i}$ is periodic for $i \geq \mu$, then $j - i \equiv 0 \pmod{\lambda}$. We have the following simple result.

PROPOSITION 1. *Let $\{F, f, x_0\}$ be a generator, where F is a finite set. The sequence $\{x_n\}_{n\geq 0}$ defined through $x_{n+1} = f(x_n)$ is almost periodic.*

In our case, $F = \mathbb{Z}_m = \{0, 1, \ldots, m-1\}$ is a commutative ring with unit. We are interested in knowing under what conditions the affine mapping $x \longmapsto a \cdot x + b$ is of cycle length m. The mapping f is bijective if and only if a is an invertible element of $\mathbb{Z}_m$. The mapping f^k provides $x \longmapsto a^k x_0 + (1 + a + a^2 + \cdots + a^{k-1}) \cdot b$. Using the notation, $S_k(a) = (1 + a + a^2 + \cdots + a^{k-1})$, we may rewrite f^k as $x \longmapsto a^k x_0 + b \cdot S_k(a)$. We now provide some properties related with $S_k(a)$.

For that purpose, we define the following polynomial with integer coefficients for $k \geq 1$,

$$S_k(x, y) = \frac{x^k - y^k}{x - y} = x^{k-1} + x^{k-2}y + \cdots + y^{k-1}.$$

PROPOSITION 2. *Let $k = p$ be a prime number and let $x, y \in Z_k$. If $x \equiv y(\operatorname{mod} k)$, then $S_k(x, y) \equiv 0(\operatorname{mod} k)$.*

PROOF. Indeed, if $x \equiv y(\operatorname{mod} k)$, then $S_k(x, y) = x^{k-1} + x^{k-1} + \cdots + x^{k-1} = kx^{k-1}$ and as $k = p$, according to Fermat's theorem $x^{k-1} \equiv 1(\operatorname{mod} k)$. Finally, $S_k(x, y) = k \equiv 0(\operatorname{mod} k)$. If $y = 1$, then $x \equiv 1(\operatorname{mod} k)$ and $S_k(x) \equiv 0(\operatorname{mod} k)$. On the other hand $S_k(a) = (1 + a + a^2 + \cdots + a^{k-1}) = (1/a - 1)(a - 1)(1 + a + a^2 + \cdots + a^{k-1}) = (1/a - 1)(a^k - 1)$. Therefore, $S_k(a) = (a - 1)^{-1}(a^k - 1)$ and $S_k(x) \equiv 0(\operatorname{mod} k)$, so $a^k \equiv 1(\operatorname{mod} k)$. ∎

It is important to remark that for f to be of length cycle m, fixed points should not exist. For the sake of simplicity, let $b = 1$. Then, the equation $x = a \cdot x + a$ has a unique solution if $a \not\equiv 1 (\operatorname{mod} m)$. Hence, in order to f will not have a fixed point, it is necessary that $a \equiv 1 (\operatorname{mod} m)$. But, in this case, $S_k(x) \equiv 0 (\operatorname{mod} k)$ and $f^m = e$, where e is the identity element. Following this reasoning, it is not difficult to prove the following.

PROPOSITION 3. *If m is a prime number, $a \equiv 1 (\operatorname{mod} m)$ and b is not congruent with $0 (\operatorname{mod} m)$, the mapping $f_{a,b} : \mathbb{Z}_m \to \mathbb{Z}_m$ defined by $f_{a,b}(x) = \overline{a \cdot x + b}$, where $\overline{a \cdot x + b}$ is the equivalence class modulo m corresponding to the number $a \cdot x + b$, is a cyclic permutation of order m in $\mathbb{Z}_m$ (and hence, a maximum length generator).*

The single-step generator that we have seen cannot generate sequences with period longer than $|F|$. Our objective is to design a generator whose period length is close to the theoretical boundary $m!$, the order of a symmetric group with m elements. It is well known that if the order of the cyclic group $\langle f \rangle$ generated by f satisfies $|\langle f \rangle| = s$, given $x \in \mathbb{Z}_m$, we have that either x is a fixed point of certain element of $\langle f \rangle$ (in this case, we denote x by x_F), or the set $H_x^s = \{x, f(x), \dots, f^{s-1}(x)\}$ contains exactly s elements. Moreover, it is also possible to find a sequence $x_1, x_2, \dots, x_l \in \mathbb{Z}_m$, such that

$$\mathbb{Z}_m = H_0 \cup H_{x_1}^s \cup \cdots \cup H_{x_l}^s,$$

where $H_0 = \{x_F\}$, $|H_{x_1}^s| = \cdots = |H_{x_l}^s| = s$, and $l \cdot s = m - 1$.

3. GENERATION OF PSEUDO-RANDOM NUMBERS AND LINEAR CONGRUENCES

Generation based on the expression,

$$x_{n+1} = (a \cdot x_n + b) \operatorname{mod} m, \tag{1}$$

depends on four parameters, a, b, m, and x_0. The parameters in expression (1) can be divided in two categories,

1. parameters a, b, m, providing a maximum length generator,
2. parameters a, b, m, which do not provide a maximum length generator.

The first case is thoroughly considered in [5]. Here, we shall carry out the analysis of the second option, concentrating on the case where m is a prime number. To do so, we shall use the following well-known results that summarize part of what was mentioned in the previous section.

THEOREM 1. *If $a \equiv 1 (\operatorname{mod} m)$, then for any sequence produced according to (1) there exists a fixed point x_F, such that*

$$x_F = (a \cdot x_F + b) \operatorname{mod} m.$$

REMARK. If m is a prime number, then $m - 1$ is composed, so the period formed by the previous $m - 1$ numbers is divided in l subsets of s elements each, so that

$$1 \cdot s = m - 1 = \phi(m),$$

where $\phi(m)$ is the Euler phi function, which counts the number of integers in $\{1, \dots, n\}$ that are relatively prime to n. The number of possible values of s is

$$\tau(\phi(m)) - 1,$$

where $\tau(p_1^{\alpha_1}, p_2^{\alpha_2}, \dots, p_n^{\alpha_n}) = (\alpha_1 + 1), (\alpha_2 + 1), \dots, (\alpha_n + 1)$ for $(m - 1) = p_1^{\alpha_1}, p_2^{\alpha_2}, \dots, p_n^{\alpha_n}$. The number of subsets l is

$$l = \frac{m-1}{s}.$$

Now, for a given value of s, the valid values of parameter a are the solutions of the following equation,

$$a^s \equiv 1 \,(\operatorname{mod} m)\,. \tag{2}$$

The number of solutions of this equation is $\phi(s)$, so that, if $s = \phi(m)$, we have Euler theorem. To be able to choose the coefficient a, we must keep in mind the limited possibilities of programming languages such as FORTRAN or C. In the case of FORTRAN, in order for the intermediate results of calculation with expression (1) to not surpass the 31-bit threshold, it is possible to use Schrage's decomposition [6]. This is based on the fact that any integer number of 31 bits can be expressed as

$$\alpha \cdot 2^{16} + \beta,$$

where α and β are numbers of 15 and 16 bits, respectively. If we use C with the option of minimum representation of 32 bits, parameters a, b, and m must satisfy the condition,

$$a\,(m-1) + b \leq 2^{32} - 1. \tag{3}$$

In order to select a, we follow the recommendations of [5] and impose that a and m satisfy the conditions,

$$0.01m < a < 0.99m. \tag{4}$$

Also, in order to satisfy the serial correlation criterion, see [5], parameters a and m should be such that the partial quotients of a/m must be small, so that the Dedekind sums $\sigma(a, b, m)$ and the serial correlation coefficient,

$$C \approx \sigma\,(a, b, m)\,/m,$$

is small. This property also favors the fulfillment of the series criterion for pairs of numbers.

If the parameter a is chosen properly, the parameter b ($k = b/m$) may be chosen arbitrarily, for example satisfying the condition (related with Knuth requirements),

$$\lfloor k \cdot m \rfloor \leq b \leq \lceil k \cdot m \rceil\,, \tag{5}$$

where

$$k = \left(\frac{1}{2} - \frac{1}{6}\sqrt{3}\right) \approx 0.21132486540518711774 54.$$

Substituting in (3) the lower bounds of (4) and (5) we obtain the following equation for m,

$$m^2 - m + \lfloor k \cdot m \rfloor * 10^2 - (2^{32} - 1) * 10^2 \leq 0.$$

Solving this equation for integer numbers, we can obtain $m = 655339$ as a valid value for this parameter. In a similar way, we obtain an equation for the upper bounds for the parameters a and b. In consequence, the choice of m is restricted. For a generator of maximum period, the only way to increase the period length is to increase the value of the parameter m, which is bounded, or use a combination of generators.

In our case, the following procedures can be followed in order to achieve a long period. As we have l subsets, the simplest way to form the series is the following. Introduce a seed x_0 for the i^{th} subset, generate all the numbers x belonging to this i^{th} subset, take a seed x_0 from the following subset $i + 1$ and so forth until all the x_0 from all the l subsets have been used. In this manner, we can obtain all the m values of x in the interval from 0 to $m-1$, including the value of x_F. Therefore, it is possible to produce a sequence whose generation depends on the following.

1. The order in which we go through the subsets. The number of possible variations is $l!$, hence, the period length will be $l! \cdot m$.
2. The choice of seed x_0 in each subset. The number of combinations for the choice of the initial value is s^l, hence, the period length will be $l! \cdot m \cdot s^l$.
3. The type of process we use to choose the number of elements in each subset. This process is carried out so that the uniform law of distribution of x in the interval between 0 and $m-1$ holds. In consequence, the period length will be $l! \cdot m \cdot m!/(s!)^l$.

In this way, for the chosen parameters a, b, m, we can generate not only a sequence (as in the generator of complete period case) but many different sequences. Therefore, it is possible to come closer to the theoretical upper bound of $m!$ elements.

Following the generation procedures above, and if we take initially $m = 655339$, its closest prime number is 655337, so taking $m = 655337$, we have that $m - 1 = 655336 = (2^3) \cdot (11)^2 \cdot 677$. If, for instance, $l = 2 \cdot 11 = 22$, we determine $s = (655337 - 1)/22 = 29778$. The period length for each generation from 1 to 3 will be the following.

1. $22! \cdot 6,55337 \cdot 10^5 \approx 10^{21} \cdot 6,55337 \cdot 10^5 \approx 10^{26}$;
2. $22! \cdot 6,55337 \cdot 10^5 \cdot (29778)^{22} \approx 10^{21} \cdot 6,55337 \cdot 10^5 \cdot 10^{98} \approx 10^{124}$;
3. $22! \cdot 6,55337 \cdot 10^5 \cdot (6,55337 \cdot 10^5!)/(29778!)^{22} \approx 10^{21} \cdot 6,55337 \cdot 10^5 \cdot 10^{3527102}/10^{2647414} \approx 10^{879714}$.

To see the degree of approximation, we can apply Stirling's formula to approximate the value 655337!,

$$\ln(m!) \approx \left(m + \frac{1}{2}\right) \cdot \ln(m) - m + \ln(\sqrt{2\pi}),$$

from where

$$(6,55337 \cdot 10^5)! \approx 10^{3.527.102}.$$

Let us see some specific examples where m is prime. If we want a generator with m of the order of $6 \cdot 10^5$, the first prime is $m = 600011$, whereas the previous prime is $m = 599999$, and $m - 1 = 599998 = 2 \cdot 7 \cdot 17 \cdot 2521$. The possible values of l are 2, 7, 14, 17, 34, ... Let us take, for instance, $l = 17$. Then, $s = (599999 - 1)/17 = 35294$. We proceed in the same way for the prime number $m = 600011$. The partial quotients of a/m are, for 7133/599999, [0, 84, 8, 1, 1, 1, 2, 103], and for 7091/600011, [0, 84, 1, 1, 1, 1, 1, 1, 12, 14, 3].

The results of the calculations made for this value m in four examples are gathered in Table 1, where $x_0[i]$ are the solutions of (2) previously ordered.

The generator with the parameters mentioned above has been subject to a series of statistical tests. They do not suffice to judge statistical properties, therefore, in order to use the the generator in an extensive way in practical cases it would be more appropriate to try it with batteries of 15 or more significant tests, see [7], to allow for a more justified choice of the parameters a, b. Each criterion is applied to the following sequence of real numbers between 0 and 1,

$$x_0, x_1, \ldots, \tag{6}$$

which are assumed to be statistically independent and uniformly distributed. Some criteria are suitable for integer numbers but they are not so for sequence (6). In this case, instead of the above sequence, we can use

$$y_0, y_1, \ldots, \tag{7}$$

defined by

Table 1. Examples.

Example 1	Example 2	Example 3	Example 4
$m = 599999$	$m = 599999$	$m = 600011$	$m = 600011$
$a = 7133$	$a = 7133$	$a = 7091$	$a = 7091$
$b = 126795$	$b = 122939$	$b = 126798$	$b = 126798$
$l = 17$	$l = 14$	$l = 10$	$l = 29$
$x_F = 373930$	$x_F = 373930$	$x_F = 307858$	$x_F = 307858$
$s = 35294$	$s = 42857$	$s = 44884$	$s = 20690$
$x_0[i]$	$x_0[i]$	$x_0[i]$	$x_0[i]$
26006	26006	26007	26007
26010	26010	26013	26013
26017	26017	26017	26017
26030	26030	26022	26022
26035	26035	26029	26029
26065	26065	26031	26031
26080	26080	27037	27037
26096	26096	27046	27039
26117	26117	27050	27046
26138	26138	27059	27050
26163	26163	27071	27059
26233	26233		27071
26320	26320		27085
26334	26334		27089
26603			27093
26864			27102
27977			27105
			27113
			27121
			27131
			27143
			27148
			27151

$$y_n = \lfloor d \cdot x_n \rfloor , \tag{8}$$

which is a sequence of uniformly distributed integer numbers between 0 and $d - 1$. d is chosen so that it is easy to use in one way or another.

For each example, a file of 2 MB has been generated according to the generation Strategies 1–3. Each value x was constrained to the interval $(0, 1)$ and multiplied by a multiple $d = 2^8$, as indicated in [5]. Thc obtained files were subjected to a series of statistical tests for a significance level $p = 0.99$.

Table 2. Statistical test for the examples.

Example	χ^2	Series	Inversions
Example 1			
Strategy 1.	0.0034	2.51	1.85
Strategy 2.	0.0034	3.06	2.20
Strategy 3.	0.0102	2.59	2.00
Example 2			
Strategy 1.	0.0034	2.03	2.13
Strategy 2.	0.0034	3.23	2.15
Strategy 3.	0.0102	2.30	2.01
Example 3			
Strategy 1.	0.0228	2.74	1.85
Strategy 2.	0.0228	4.18	2.25
Strategy 3.	0.0684	2.62	1.94
Example 4			
Strategy 1.	0.0228	2.43	1.70
Strategy 2.	0.0228	5.18	2.37
Strategy 3.	0.0684	2.50	1.94

4. CONCLUSIONS

Our generator is, in some sense, a model for certain approaches. The restrictions for parameters a, b, m are very general and can be substituted by others or simply eliminated in future computer platforms. If, for instance, parameter m is a large prime number, then the factorization of the number $m - 1$ would become complex. Consequently, all the estimates and reconstructions of other parameters depending on parameter m may be inefficient. The key is the vector of seeds of each interval,

$$\overline{x_0} = \{x_0^1, x_0^2, \ldots, x_0^l\}.$$

These are primitive roots or generators of the cyclic groups. If the factorization of $m - 1$ is not known, then computing these roots is inefficient, even more when there also exist $\phi(m-1)$ primitive roots. Hence, if we have a sufficiently big sample of the generated sequence, it is possible to determine the parameters a, b and m. However, we lack an efficient method to determine the intervals if one does not know the factorization of $m - 1$, being specially complex to determine the number of elements within each interval and the parameter s in the equation $a^s \equiv 1 (\operatorname{mod} m)$.

REFERENCES

1. P. L'Ecuyer, Random number generation, In *Handbook of Simulation: Principles, Methodology, Advances, Applications and Practice*, (Edited by J. Banks), John Wiley and Sons, (1998).
2. J. Boyar, Inferring sequences produced by pseudo-random number generator, *J.A.C.M.* **36**, 129–141, (1989).
3. A.M. Frize, R. Hastad, J.C. Lagarias and A. Shamir, Reconstruction truncated integer variables satisfying linear congruences, *Journal on Computing* **17** (2), 262–280, (1988).
4. B. Schneier, *Applied Cryptography*, Second Edition, John Wiley and Sons, New York, (1996).
5. D.E. Knuth, *The Art of Computer Programming V.2*, Third Edition, Addison-Wesley, Reading, MA, (1997).
6. L. Schrage, A more portable Fortran random number generator, *ACM. Trans. Math. Software* **5**, 132–138, (1979).
7. G. Marsaglia, *The Marsaglia Random Number CDROM, Including the Diehard Battery of Tests of Randomness*, Department of Statistics, Florida State University, Tallahassee, FL, (1995).

8. J. Banks, *Handbook of Simulation: Principles, Methodology, Advances, Applications and Practice*, John Wiley and Sons, Inc., (1998).
9. P. Bratley, B.L. Fox and L.E. Schrange, *A Guide To Simulation*, Springer-Verlag, Berlin, (1983).
10. R. Kumanduri and C. Romero, *Number Theory with Computer Applications*, Paramus, Prentice Hall, (1998).