On the Initial Seed of the Random Number Generators

Tae Soo Kim* and Young Hae Lee**

Abstract

A good arithmetic random number generator should possess full period, uniformity and independence, etc. To obtain the excellent random number generator, many researchers have found good parameters. Also an initial seed is the important factor in random number generator. But, there is no theoretical guideline for using the initial seeds. Therefore, random number generator is usually used with the arbitrary initial seed. Through the empirical tests, we show that the choice of the initial values for the seed is important to generate good random numbers.

Key Words: random number, multiple recursive generator, empirical tests

1. Introduction

The ability to generate satisfactory sequences of random numbers is one of the key links between Computer Science and Statistics. Standard methods may no longer be suitable for increasingly sophisticated uses, such as in precision simulation studies. A simulation of any system or process in which there are inherently random components requires a method of generating or obtaining numbers that are random, in some sense. All the randomness required by the simulation model is simulated by various random number generators whose output is assumed to be a sequence of independent uniform random variables, which is denoted "U(0,1)". These random numbers are then transformed as needed to simulate random variables from different probability distributions.

But, the random variable in U(0,1) is an mathematical abstraction. In practice, there are no true random variables. As of today, from a prescribed mathematical formula but satisfy different requirements as if they were true random numbers, we gain the sequence. Such a sequence is called the pseudo-random and the program or procedure that produce such a sequence is called pseudo-random number generator. The most popular algorithm for generating pseudo-random numbers was suggested by Lehmer in 1949. It is called the congruential method. The method relies on a sequence of integers that are computed by one formula

$$m_i = g(m_{i-1}, m_{i-2}, \cdots) (\mathrm{mod}\, M), \tag{1}$$

where a fixed deterministic function g of previous given elements $m_{i-1}, m_{i-2}, \cdots$, the modulo M are prescribed integers. As pseudo-random numbers, the fractions m_i / M are used. In particularly, if g is a linear function of $m_{i-1}, m_{i-2}, \cdots$, we called it as a linear congruential generator(LCG). In general the LCG is probably the most widely used and best understood kind of random-number generator. Turning to small M, the length of period reduces. On the other hand, if a long period generator is implemented, then the generation is slow. So there are many alternative types. In order to the formula (1) have the full period and good statistical properties, the values of the parameters in a function g must be carefully chosen[1,4,8]. In this paper, we think of the Multiple Recursive Generator[3,4,9,10] and the Combined Generator[5,9.11]. In particular, we studied two combined multiple recursive generators which were designed by L'Ecuyer[2]. We have interest to the statistical properties of generators.

In the formula (1), when $g(m_{i-1}, m_{i-2}, \cdots, m_{i-q}) = a_1 m_{i-1} + a_2 m_{i-2} + \cdots + a_q m_{i-q}$, where a_i's are constants and the initial values $m_{i-1}, m_{i-2}, \cdots, m_{i-q}$ are not all zero. We called them the qth-order multiple recursive generators (MRGs). From the finite field theory, the qth-order MRGs can produce

* Research Professor, Department of Industrial Engineering, Hanyang University, Ansan, Kyunggi-do, 425-791, South Korea
E-mail: tskim@hanyang.ac.kr
** Professor, Department of Industrial Engineering, Hanyang University, Ansan, Kyunggi-do, 425-791, South Korea

random numbers of full period M^q-1 if and only if the polynomial $f(x)=x^q-a_1x^{q-1}-\cdots a_q$ is a primitive polynomial modulus M. Knuth[7] describes the following conditions for testing the primitiveness modulo M:

(i) $(-1)^{q-1}a_q$ is a primitive root modulo M,

(ii) $[x^r \bmod f(x)] \bmod M=(-1)^{q-1}a_q$,

(iii) degree$\{[x^{r/s} \bmod f(x)] \bmod M\}>0$,, for each prime factor s of r, where $r=\frac{(M^q-1)}{(M-1)}$.

Theoretically, there are exactly choices of $(a_1,a_2,\cdots a_q)$ which satisfy these conditions, where $\phi(M^q-1)$ is the Euler function defined as number of integers which is smaller than and relatively prime to M^k-1. For the simplest case of $q=2$ and the very popular modulus $M=2^{31}-1$, there are around 5.74E17's candidates[4]. Hence a significant amount of computation is involved in searching for $(a_1,a_2,\cdots a_q)$ which are able to produce random numbers of full period.

To increase the period and try to get rid of the regular patterns displayed by LCGs, it has often been suggested that different generators be combined to produce a hybrid one. Such combination is often viewed as completely heuristic and is sometimes discouraged. But besides being strongly supported by empirical investigations, combination has some theoretical support. First, in most cases, the period of the hybrid is much longer than that of each of its components, and can be computed. Second, there are theoretical results suggesting that some forms of combined generators generally have better statistical behavior. In this paper, we think about the combination of two MRGs, which was developed and studied by L'Ecuyer, is defined by

$$m_{1,i}=(a_{1,1}m_{1,i-2}-a_{1,2}m_{1,i-3})[\bmod(2^{32}-209)],$$

$$m_{2,i}=(a_{2,1}m_{2,i-1}-a_{2,2}m_{2,i-3})[\bmod(2^{32}-22853)],$$

$$Y_i=(m_{1,i}-m_{2,i})[\bmod(2^{32}-209)],$$

$$U_i=\frac{Y_i}{2^{32}-209},$$

where $a_{1,1}=1403580$, $a_{1,2}=810728$, $a_{2,1}=527612$, $a_{2,2}=1370589$, and has period of approximately 2^{191}(which is about 3.1×10^{57}) as well as excellent statistical properties through dimension 32[2]. The advantage of the above generator is a brief program, simple computations and a huge period. In order to use this algorithm, likewise using any other random generators, we need the seed vector with 6-elements $\{m_{1,0},m_{1,1},m_{1,2},m_{2,1},m_{2,2},m_{2,3}\}$.

The choice of the initial seed vectors in random number generator could not be determined by the theoretical basis. The recommendation to select initial values at random is doubtful. In general, the initial seed vectors could be chosen by empirical methods. To be sure, the careful selection of the seeds is important to generate the pseudo-random numbers. So, L'Ecuyer gave the 10,000's seeds vector as related header-file and asserted that the results have excellent statistical properties. But, for the empirical test to see the uniformity and independence of the two combined-MRGs, we obtained the different results. The test results will be given in the next section.

2. The Empirical Tests

In this section, we practice the various simulation to test the uniformity and independence of distribution of the corresponding pseudo-random numbers. And all tests are related to the deterministic interpretation of goodness-of-fit tests. In facts, d-dimensional random points with independent Cartesian coordinates $(\gamma_1,\cdots,\gamma_d),(\gamma_{d+1},\cdots,\gamma_{2d}),(\gamma_{2d+1},\cdots,\gamma_{3d}),\cdots$ are uniformly distributed in the d-dimensional unit cube at any d. This property is necessary and sufficient for a successful implementation of Monte Carlo algorithms with constructive dimension d. To test whether the null hypothesis H_0: the above d-tuples sequences are distributed uniformly on $[0,1]^d$, is true or not, divide [0,1] into k subintervals of equal

size and let $f_{j_1,j_2,\cdots,j_d}$ be the number of γ_i's having first component in subinterval j_1, second component in subinterval j_2, etc. If we let $\chi^2_N = \frac{k^d}{N}\sum_{j_1=1}^{k}\cdots\sum_{j_d=1}^{k}(f_{j_1,j_2,\cdots,j_d} - \frac{N}{k^d})^2$, then χ^2_N will have an approximate chi-square distribution with degree of freedom $k^d - 1$, under the null hypothesis H_0 is true. The smaller is χ^2_N the better is the agreement of empirical values with theoretical ones. Large values χ^2_N correspond to small p-values. So, too small values of p-values indicate that the experimental data contradicts to our uniformity hypothesis. Firstly, for the uniformity, we have tested for the case d=1, which is called the frequency test, and d=2,3,4, which are called the serial tests. For modeling different problems, different quantities of pseudo-random numbers are necessary. Therefore, we have simulated various initial seeds of a sequence with lengths $N = N_d \times 2^s$, where $s = 0,1,2,\cdots,14$, N_d =600, 300, 250, 150, according to the d=2,3 and 4, respectively. And let k the number of subintervals of [0,1] be as 16, 8, 5, and 4 with respect to the d=1, 2, 3, and 4.

Secondly, for the test of independence, we have proceeded the run test. Let n_i be the number of runs of length i in a sequence of $N = 600 \times 2^s$, where $s = 0,1,2,\cdots,14$. For an independent sequence, the expected values of n_i for runs up and down are given by

$$E(n_i) = \begin{cases} \frac{2}{(i+3)!}[N(i^2+3i+1) - (i^3+3i^2-i-4)], & i \le N-2, \\ \frac{2}{N!}, & i = N-1, \end{cases}$$

Under the null hypothesis H_0: the pseudo-random numbers which are generated by the two combined MRG is distributed independently. We know $\chi^2_N = \sum_{i=1}^{4}\frac{(n_i - np_i)^2}{np_i} + \frac{(n_i' - np_5)}{np_5}$, where n_i' is the number of runs with lengths ≥ 5 and $n = n_1 + n_2 + n_3 + n_4 + n_5'$ means the total number of runs, and the probabilities $p_i = E(n_i)$, for $i = 1,2,\cdots,N-1$, will have an approximate chi-square distribution with degree of freedom 4.

For all tests, we use $\Phi_i = \max_s \chi^2_N$, for $i = 1$, which means the frequency test, for i=2,3 and 4, which means the 2, 3, and 4 dimensional serial tests, respectively, for i=5, which means the run test as the criteria. When all values of Φ_i are less than the quantiles Φ_i^{**} with respect to p-values as 0.1, we will say that the pseudo-random numbers generated by two-combined MRG are distributed uniformly and independently. The recommendation of L'Ecuyer was arbitrarily to select an initial value in 10,000's seed vectors was proposed in his header-file. We have tested arbitrary 100 sequences initial seed vectors among 10,000. And we selected the seed vectors meets criteria in all five tests at the same time. The results of the above tests are terrible. The only one 5230th seed vector (1338960199, 3947731640, 1058186044, 1875415108, 1948201518, 3217931286) passed the all five tests. And the results $\Phi_i = \max_s \chi^2_N$ and $P_i = \min_i P(\chi^2_N)$ of each tests are described in Table 1.

Table 1. The results of test with the 5230th initial seed vector

Tests	$\Phi_i = \max_s \chi^2_N$ ($\min_i P(\chi^2_N)$)	Φ_i^{**} withp − value0.1
Frequency test	19.6557(0.19)	22.3
Serial : 2-dim	75.4642(0.14)	77.7
Serial : 3-dim	144.329(0.10)	145
Serial : 4-dim	283.04(0.15)	284
Run Test	7.6133(0.11)	7.78

Continuously, we proceed with the five empirical tests for all given 10,000's seed vectors. It required the very expand test time. We found out the only 44 of 10,000 passed all five tests. Table 2 shows the result of tests.

Table 2. The list of numbers among 10,000 which passed the all five test in the L'Ecuyer's header-file

The Number of the seed vectors										
74	256	315	420	1007	1373	1385	1561	2069	2495	2744
2859	3139	4081	4255	4416	4950	5147	5214	5230	5376	5798
6020	6090	6105	6118	6123	6154	6246	6537	6921	6934	7389
7900	8372	7374	8983	8990	9329	9424	9542	9568	9718	9998

3. Conclusions

In simulation studies, the quality of the random number generator adopted has a major effect on the results derived. An ideal random number generator should possess at least the properties of long period, good lattice structure, and sound statistical properties. The arbitrary selections of the initial seed values in the random number generators would be not a suitable results. So, we select the initial conditions with attention. As a future theme, we would find the theoretical condition for good random number generator in various cases.

References

[1] Ana Proykova, How to improve a random number generator, Computer Physics Communications 124, pp125-131 (2000)

[2] Averill M.Law & W. David Kelton. Simulation modeling and Analysis, 3rd ed, McGraw-Hill (2000)

[3] Chiang Kao & Huey-Chin Tang, Upper Bounds in Spectral Test for Multiple Recursive Random Number Generators with Missing Terms, Computers Math. Applic. Vol. 33, No 4, pp 119-125 (1997)

[4] Chiang Kao & Hui-Chin Tang, Several Extensively Tested Multiple Recursive Random Number Generator, Computers Math. Applic. Vol. 36, No 6, pp 129-136 (1998)

[5] I. M. Sobol & Yu. L. Levitan, A Pseudo-Random Number Generator for Personal Computers, Computers and Mathematics with Applications, 37, 33-40 (1999)

[6] Jerry Banks. Handbook of simulation : Principles, Methodology, Advances, Applications , and Practice, John ilely & Sons (1998)

[7] Knuth, D.E. The Art of Computer Programming, Vol.2 : Seminumerical Algorithms, 3d ed., Addison-Wesley, Reading, Massachusetts (1998)

[8] Pierre L'Ecuyer, Efficient and Portable Combination Random Number Generators, Communications of the ACM, Vol.31, No.6, pp742-749, 774 (1988)

[9] Pierre L'Ecuyer, Random Numbers for Simulation, Communications of the ACM, Vol.33, No.10, pp85-97 (1990)

[10] Pierre L'Ecuyer, Francois Blouin & Raymond Couture, A Search for Good Multiple Recursive Random Number Generators, ACM Transactions on Modeling and Computer Simulation, Vol.3, No.2, pp87-98 (1993)

[11] Pierre L'Ecuyer & Terry H. Andres, A random number generator based on the combination of four LCGs, Mathematics and Computers in Simulation, 44, pp99-107 (1997)