다중귀납적생성기의 조합에 기초한 난수생성기

김 태수*·이 영해**

A random number generator based on the combination of the Multiple Recursive Generators

Tae Soo Kim*, Young Hae Lee**

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

Abstract

The Multiple Recursive Generator(MRG) has been considered by many scholars as a very good Random Number generator. For the long period and excellent statistical properties, the method of the combination with random number generators are used. In this paper, for two-combined MRGs, we examine the statistical properties and show the importance of the seeds likewise other random number generators. And we modify the two-combined MRGs and verify the statistical superiority.

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 is 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 methods relies on a sequence of integers that are computed by one formula

$$m_i \equiv g(m_{i-1}, m_{i-2}, \cdots) (\mathrm{mod} M), \qquad (1.1)$$

where a fixed deterministic function g of previous $m_{i-1}, m_{i-2}, \cdots$, the modulus M are

* 한양대학교 산업공학과 연구교수
** 한양대학교 산업공학과 연구교수

prescribed integers. As pseudo-random numbers, the fractions m_i/M are used. In particularly, if $g(m_{i-1},m_{i-2},\cdots)=am_{i-1}+c$, where a,c are the given constants, we called it as linear congruential generator(LCG). In general the LCG are probably the most widely used and best understood kind of random-number generator. One one hand, 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 these are many alternative types. In order to the formula (1.1) have the full period and good statistical properties, the values of the parameters in a function g must be chosen very carefully. In Section 2, we give the Multiple Recursive Generators and the combined generator. In particular, we two combined multiple recursive generators which was designed by L'Ecuyer. We have interest to the statistical properties of generators. So we have the empirical test for the two combined multiple recursive generators in Section 3. We state an alternative of the referenced generators to avoid defective and show the effective results.

2. A Combination of Multiple Recursive Generators

In the formual (1.1), when

$$g(m_{i-1},m_{i-2},\cdots)=a_1m_{i-1}+\cdots+a_qm_{i-q},$$

where a_i's constants, we called it the qth-order multiple recursive generators (MRGs). From finite field theory, the qth-order MRG can produce 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 modulo m. Knuth[*] describes the following conditions for testing the primitivity of f(x):

(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 $r=(m^q-1)/(m-1)$.

Futhermore, one very appealing class of generators is obtained by the combining MRGs. The combined generators can increase the period length and improve the uniformity and independence properties of their individual generators. In this paper, we think about the combination of two MRG, which was developed and studied by L'Ecuyer, is defined by

$$\begin{aligned} m_{1,i} &= (a_{1,1}m_{1,i-2}-a_{1,2}m_{1,i-3})[\bmod(2^{32}-209)] \\ m_{2,i} &= (a_{2,1}m_{2,i-1}-a_{2,2}m_{2,i-3})[\bmod(2^{32}-22853)] \\ Y_i &= (m_{1,i}-m_{2,i})[\bmod(2^{32}-209)] \\ U_i &= \frac{Y_i}{2^{32}-209}, \end{aligned}$$

where $a_{1,1}=1,403,580, a_{1,2}=810,728,$

$a_{2,1}=527,612, \quad a_{2,2}=1,370,589,$

and has period of approximately 2^{191} (which is about 3.1×10^{57}) as well as excellent statistical properties through dimension 32 (Averill M. Law & W. David Kelton 2000). 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 $\{m_{1,0},m_{1,1},m_{1,2},m_{2,0},m_{2,1},m_{2,2}\}$.

To be sure, the careful selection of the seeds is important to generate the pseudo-random number. On the other, L'Ecuyer gave the 10,000's seeds vector as related header-file and assert that the results have an excellent statistical properties. But, for the empirical test to see the uniformity and independence of the two combined-MRG, we obtained the different results. The test results will be given in the next section.

3. 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

$$(\mathrm{v}_1,\cdots,\mathrm{v}_d),(\mathrm{v}_{d+1},\cdots,\mathrm{v}_{2d}),(\mathrm{v}_{2d+1},\cdots,\mathrm{v}_{3d}),\cdots$$

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_o : the above d-tuples sequences are distributed uniformly on [0,1], is true or not, devide [0,1] into k subintervals of equal size and let $f_{j_1 j_2 \cdots j_d}$ be the number of v_i's having first component in subinterval j_1, second component in subinterval j_2, etc. If we let

$$\mathrm{x}_N^2 = \frac{k^d}{N}\sum_{j_1=1}^{k}\cdots\sum_{j_d=1}^{k}\left(f_{j_1 j_2\cdots j_d}-\frac{N}{k^d}\right)^2,$$

then x_N^2 will have an approximate chi-square distribution with degree of freedom k^d-1 under the null hypothesis H_o is true. The smaller is x_N^2 the better is the agreement of empirical values with theoretical ones. Large values x_N^2 correspond to small p-values. So, too small values of p-values indicate that the experimental data contradict 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 $d = 1,2,3$, 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 proceed the run test. Let n_i be the numbers 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 is given by

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

Under the null hypothesis H_o : the pseudo-random numbers which is generated by the two combined MRG are distributed independently, we know

$$\mathrm{x}_N^2 = \sum_{i=1}^{2}\frac{(n_i-np_i)^2}{np_i} + \frac{(n_5'-np_5')^2}{np_5'},$$

where n_5' 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 are $p_i = E(n_i)$, for $i = 1,2,\cdots,N-1$, will have an approximate chi-square distribution with degree of freedom 4.

For all tests, we use $\Phi_i = \max_s \mathrm{x}_N^2$, for i = 1, which means the frequency test, for i = 2,3, and 4, which mean the 2, ,3, and 4 dimensional serial tests, respectively, for i=5, which means the run test as the criterions. When all values of Φ_i are less than quantiles Φ_i^{**}, $i = 1,2,\cdots,5,$ with respect to the three different p-values in Tables 1 previously, we will say that the pseudo-random numbers generated by two-combined MRG are distributed uniformly and independently on [0,1].

The recommendation of L'Ecuyer was arbitrary to select initial value in 10,000's seed vectors which was proposed in his header-file. In this paper, we choose p-value as 0.1. We have

tested arbitrary 100 sequences initial seed vectors among 10,000. And we selected the seed vector which meet criterion 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.

Tests	p-value		
	0.1	0.05	0.01
Frequency	22.3	25.0	30.6
Serial:2-dim	77.7	82.5	92.0
Serial:3-dim	145	151	163
Serial:4-dim	284	293	310
Run	7.78	9.49	13.28

Table 1. χ^2 quantiles.

For each sequences with 5230th seed vector, the results $\Phi_i = \max_s \chi^2_N$ and $P_i = \min_i P(\chi^2_N)$, where the probabilities $P(\chi^2_N)$, where $P(\chi^2_N) = \int_{\chi^2_N}^{\infty} f(x)dx$, $f(x)$ is a probability density χ^2 with degree of freedom are described in Table 2. For all $i = 1,2,\cdots,5$, we see that $\Phi_i \leq \Phi_i^{**}$.

4. Conclusions

Tests	$\Phi_i = \max_s \chi^2_N$ ($\min_s P(\chi^2_N)$)
Frequency	19.6557 (0.19)
Serial : 2-dim	75.4642 (0.14)
Serial : 3-dim	144.329 (0.10)
Serial : 4-dim	283.04 (0.15)
Run Test	7.6133 (0.11)

Table 2. Results of Test with 5230th initial seed vector

To generate random numbers of long period, one method recommended by many scholars is the multiple recursive generator which is essentially the extension of the usual prime modulus multiplicative linear congruential generator from one term to k terms. And to obtain the longer period and better statistical properties, we use the combination of the previous generators. But the choose of the initial seed vectors in random number generator could not be determined by the theoretical basis. So the initial seed vectors could be chosen by empirical methods. Out tests mean that the statistical properties are depend to the selection of the initial seed vectors. The 10000's seed vectors which was given by L'Ecuyer is not enough to use as a proper initial seed vectors. For the future theme, we consider the more combined MRGs and examine the appropriateness of the given initial seed vectors.

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 Spectrial 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)