*Pak. J. Statist*
1986, **2** (2), A, 1-9

# ON GENERATING AND TESTING OF RANDOM NUMBERS

**A. Khalique and A. M. Abouammoh**
King Saud University
Saudi Arabia

[*Received : July 20, 1985. Revised : February 8, 1986*]

## ABSTRACT

A brief historical review of the development of uniform random sample generation is given and a modification to the linear congruential method is made so that the parameters are not always fixed. Numerical comparison of various methods by testing randomness of the sample is provided.

*Key Words* : Linear congruential method, uniformity test, frequency test, gap test, K-S test. coupen collector test, serial correlation test, run test.

## 1. INTRODUCTION

Broadly speaking, simulation techniques and Monte Carlo studies have an increasing role in almost every field of human activity. These techniques and studies arise in many fields such as business, marketing, humanities, sociology, housing policies, science and genetics, etc., see Dudewicz (1975) and references therein. Simulation techniques have proved to be quite indispensable tools in solving many complex problems in statistical theory and its applications, see for example Learnmouth and Lewis (1973). Applications of statistical methods such as testing of hypotheses, estimation, analysis of time series, forecasting, goodness of fit tests, queueing models, linear models, experimental designs and reliability theory etc., are heavily dependent on simulation methods. Simulation and Monte Carlo methods in turn depend heavily on random numbers or pseudo-random numbers.

There are many methods for generating random numbers or a random sample from uniform (0, 1) distribution. In fact, generation of random numbers

AMS 1980 Subject Classification : 62E25, 65C05, 65C10 and 68J05.

(or digits), by whatever procedure or method, has to observe three main conditions (*i*) the method is reasonably fast and economical so that it requires less computer time, (*ii*) the numbers are really random and may be taken as a sample from uniform (0, 1) distribution, and (*iii*) the random numbers are independent, that is their period is far beyond the sample size. The literature in the context of generation of random numbers (or samples) and tests of randomness is too vast to be surveyed in this paper. Extensive and almost exhaustive papers, books and bibliographies in this area are given by Tocher (1963), Janson (1966), Nance and Overstreet (1972), Sowey (1972, 1978), Dudewicz (1975), Kennedy and Gentle (1980). Peskun (1980), Knuth (1981), Dudewicz and Ralley (1981) and Ripley (1983).

## 2. HISTORICAL DEVELOPMENT

The very early procedures for generating random digits used to depend mainly on mechanical or physical methods. One of these methods is the Buffon needle problem [Kuo (1965)] which was motivated by an experiment of Hall in 1873. Student (1908 ; a, b) in his sampling experiment for finding the distribution of correlation coefficient is considered an early user of the concept of randomness. Other interesting methods for generating random sample were rolling dice, spinning roullette wheel, top hat and drawing numbered wooden chips or beads from an urn. These methods are sometimes known as observational, [see Hammersley and Handscomb (1964)]. Physical (or observational) methods are used in order to avoid any known or unknown regularities in the generated sequence of random digits. The first table of random numbers may be due to Tippett (1927) which consists of digits that were collected from census reports at the suggestion of Karl Pearson. Tippett's table includes 41,600 digits which were tested by a variety of randomness tests. Later on Kendall and Babington-Smith (1951) and Fisher and Yates (1938) published the tables of random digits. The effort in this direction decreased substantialy when the Rand Corporation (1955), produced its monumental million random digits. It is noticed that most of random numbers generated by some observational methods fail to pass statistical tests of randomness and they are not handy for storage purposes. Other physical sources of random numbers such as radioactivity and noise in electronic circuit are difficult to debug, [see Hull and Dobell (1962)]. Miyatake et al. (1981) have designed an apparatus that uses the probablilistic nature of gamma ray from radioactive nuclei to generate random digits. Their method gives $2.6 \times 10^6$ random digits per hour and has passed some tests of randomness.

The first mathematical or internal numeric method is the mid-square method due to Von Neumann and Metropolis in 1946 [see Hammer 1951]. Hull and Dobell (1962) use $\pi n^2 - [\pi n^2]$, $n - 1, 2, \ldots$ where $[x]$ refers to the integral part of x. The most recent and frequent method is the congruential method [see Lahmer (1951) and Kurita (1977)]. Another new method introduced by Tausworthe (1965) known by feedback shift register method assumes a linear recurrence equation for generating random numbers and depends mainly on the techniques of feedback shift registers developed by Golomb (1967) in communication theory.

## 3. RANDOM NUMBER GENERATION

This section includes detailed explanation of two main internal numeric methods namely the mid-square and the congruential methods. We propose a modification to one of these methods in order to produce more efficient generator. The theory behind the longer cycle of the numbers is not discussed.

### 3.1. Mid-Square Method

This method is one of the earliest and simplest for random number generation which was proposed by Von Neumann and Metropolis in the 1940's [see Von Neumann (1951)].

The mid-square, like most random number generators, determines a sequence of random numbers $x_i$ according to a recurrence relation of the form

$$x_{i+1} = f(x_i)$$

where $f(x_i)$ denotes the operation of squaring and extracting that middle digits. Since there are only finitely many values that $x_j$ can assume, eventually $x_j = x_{j+d}$ where d is called the period of sequence.

### 3.2. Linear Congruential Method

The linear congruential random number generator is considered as the most popular generator with comparatively longer cycle. It depends on an algebraic relation of a recurrence form that is $x_{i+1} = f(x_i)$. The usual recursive rule is

$$x_{i+1} = (a\, x_i + c) \bmod m \qquad (3.1)$$

At the start of the generation the seed number $x_0$ and constants a, c and m are given. The choice of these inputs plays a significant role in the cycle length which has a maximum m.

Marsaglia (1972) has given some rules for finding the period of the sequence for various choices of a, c and m. In case $c = 0$ in (3.1), the

generator is known as the multiplicative congruential method. This method is due to Lehmer (1951) and its period is found by a method due to Kurita (1977).

### 3.3. Modified Linear Congruential Method

The modified linear congruential generator differs from the linear congruential generators in the choice of the parameters a and c. In the present method we use (3.1). Whenever the newly generated random number is zero or is the start of a cycle we use the mid-square method to obtain new constants a and c.

## 4. STATISTICAL TESTS OF RANDOMNESS

It has been noted in section 1 that the infinite sequence of random numbers is appoximated by some finite sequence of (pseudo) random numbers which in turn is generated by a non-random process (phenomenon). Therefore some reliable statistical tests for randomness of the generated sequence must be applied. Knuth (1971) noticed that many random number generators are not random. The Academic Computing Service Bulletin (1975, p. 10-11) describes the popular RANDU generator of the IBM scientific subroutine package as "inaccurate, obselete and downright dangerous to use". Tests of randomness are, therefore, most essential before the numbers are used. Tests of randomness are generally devided into two kinds, *viz.*, theoretical and empirical. We shall give some of the famous forms used for testing the randomness of the generated sequence of random numbers. In order to test the randomness of numbers generated by the modified linear congruential method, we employ the following :

**4.1. Uniformity Test :** Here, we consider

$$\chi^2_9 = \sum_{i=1}^{10} (0_i - e_i)^2 / e_i \qquad (4.1)$$

where $0_i$ and $e_i$ are the observed and the expected frequencies, respectively in the ith of the equi-distant intervals on (0, 1).

**4.2. Frequency Test :** We use the same $\chi^2$-formula as in equation 4.1 but now the frequencies relate to the ith digit $i = 0, 1, \ldots, 9$.

**4.3. The Gap Test :** The probability of a gap of length k betweeu two identical digits is given by

$$P(k) = (0.1)\,(0.9)^k \qquad k = 0, 1 \ldots$$

we use this probability to obtain expected frequencies and then use Chi-square to test the goodness of fit.

**4.4. Kolmogorov–Sinirnov Test :** Here we use the usual K-S statistic given by

$$D_n = \sup_x | F_n(x) - F(x) |$$

the theoritical distribution of $D_n$ is well-known and tabulated for instance in Table VI of Lindgren [1962) p. 486].

**4.5. Coupon Collectors Test :** Greenwood (1955) has obtained the distribution of sequence required to contain a complete set of all digits and this is given by

$$g_i = 10^{n+1} \sum_{j=1}^{8} {}^9C_j (-1)^j (9-j)^{n-1}, \; i = 12, 13, \ldots 20$$

we use

$$\chi^2_8 = \sum_{i=12}^{20} (0_i - e_i)^2 / e_i$$

where $e_i = Mg_i$ and M is the number of counted sets.

**4 6 Serial Correlation Test :** Considering the sequence $x_i$ , i = 1, 2, ... n of independently uniformly distributed random variables, we calculate

$$c_k = \sum_{i=1}^{n-k} (u_i \, u_{i+k}) / n$$

where k is the lag and k = 0, 2, ... n − 1. Standardization

$$z_k = (c_k - 0.25) (\sqrt{12 (n-1)} \quad \text{gives } z_k \sim N(0, 1)$$

[see Kennedy and Gentle (1980)]. Taking k = 1, ..., 5. we calculate

$$\chi^2_5 = \sum_{i=1}^{5} z_i^2$$

**4.7. Run Test :** This is one of the most powerful tests of randomness. It depends on the length of runs in a monotone sequence of random numbers. For details, [see Leven and Wolowitz (1944)].

Using these tests we show the superiority for our method.

## 5. COMMENTS AND CONCLUDING REMARKS

In the linear congruential model, $X_{i+1}$ is generated by the formula

$$X_{i+1} = (a X_i + c) \bmod m$$

m is taken to be a prime number like $2^{16} + 1$ or $2^{31} - 1$ etc., and is bigger than N, the number of random numbers to be generated. The value m is observed to play an important role in the properties of the random numbers

generated. It should be noted that the modified linear congruential method can generate more than N random numbers even when m is less than N and this ensures better randomness. We have, however, used a value of m which is slightly bigger than N. The values of Chi-square statistics for ordinary and modified linear congruential methods have been computed in Table 1, 2 and 3 using of aforementioned tests which show the superiority of modified linear congruential method. It is advisable, in general, that m must be at least equal to $N^2$, though we have used $2^{10}-3$ to generate 1000 random numbers.

## 6. ACKNOWLEDGEMENT

We are thankful to the editor and the referees whose comments have helped to improve the paper.

**Table 1**

Results from Ordinary Linear Congruential Method

| | Uniformity $\chi^2_9$ | Frequency $\chi^2_9$ | Gap $\chi^2_{19}$ | Coupon Collector $\chi^2_8$ | K S Statistics D | Run Test $\chi^2_6$ | Serial Correlation Test $\chi^2_5$ |
|---|---|---|---|---|---|---|---|
| | 3.2800 | 9.1450 | 21.3750 | 137.636 | 0.0197 | 65.0883 | 0.1640 |
| | 14.5000 | 7.0450 | 17.0177 | 2.5489 | .0453 | 75.4122 | 33.6012 |
| | 11.7000 | 3.1000 | 13.5538 | 3.7917 | .0226 | 22.5565 | 0.5280 |
| | 14.5000 | 9.0000 | 11.9854 | 1.8304 | .0196 | 67.5500 | 0.6619 |
| | 9.7000 | 5.1550 | 33.3140 | 8.8204 | .0424 | 74.4073 | 8.6040 |
| | 14.7600 | 17.9000 | 22.5798 | 8.8187 | .0396 | 23.0230 | 14.9647 |
| | 13.3000 | 2.3900 | 20.6694 | 7.7630 | .0257 | 22.9171 | 0.2787 |
| | 17.7000 | 14.3150 | 30.0543 | 4.9134 | .0270 | 65.8047 | 0.4911 |
| | 14.6000 | 9.4650 | 18.3346 | 2.1348 | .0256 | 17.8507 | 2.3944 |
| | 1.8600 | 5.9550 | 12.0811 | 13.1840 | .0156 | 17.1596 | 0.7501 |
| Minimum value | 1.8600 | 3.1000 | 12.0811 | 1.8304 | .0156 | 17.1596 | 0.1640 |
| Maximum value | 17.7000 | 17.9000 | 33.3140 | 13.7636 | .0453 | 75.4122 | 33.6012 |

**Table 2.**—Results from GGUBS of Subroutine of IMSL

| | Uniformity $\chi^2_9$ | Frequency $\chi^2_9$ | Gap $\chi^2_{19}$ | Coupon Collector $\chi^2_8$ | K S Statistics D | Run Test $\chi^2_6$ | Serial Correlation Test $\chi^2_5$ |
|---|---|---|---|---|---|---|---|
| | 8.5199 | 18.0599 | 21.9778 | 3.4146 | 0.0330 | 74.6113 | 6.0901 |
| | 7.6599 | 1.9549 | 20.6152 | 6.5983 | 0.0254 | 79.9064 | 0.3807 |
| | 9.1999 | 5.0549 | 23.5961 | 7.5608 | 0.0206 | 66.4822 | 1.7332 |
| | 8.9999 | 7.4099 | 10.3687 | 3.3951 | 0.0215 | 24.8761 | 1.0794 |
| | 10.8199 | 11.3099 | 19 9413 | 12.4620 | 0.0179 | 167.3919 | 0.3360 |
| | 8.6399 | 2 8249 | 11.8375 | 10.4413 | 0.0306 | 20.9070 | 1.8592 |
| | 10.7799 | 9.8549 | 19.9540 | 3.7903 | 0.0078 | 64.9810 | 2.7641 |
| | 9.1199 | 8.8849 | 12.1260 | 5.0782 | 0.0318 | 22.4850 | 9.3465 |
| | 10.9998 | 7.7949 | 30.0733 | 6.6392 | 0.0214 | 69.5408 | 0.0835 |
| | 10.3999 | 18.1749 | 23.2075 | 9.3210 | 0.0319 | 36.1837 | 1.0572 |
| Minimum value | 7.6599 | 1.9549 | 10.3687 | 3.3932 | 0.0179 | 20.9070 | 0.0835 |
| Maximum value | 10.9993 | 18.1749 | 30.0733 | 12.4620 | 0.0330 | 167.3919 | 9.3465 |

**Table 3.**—Results from Modified Linear Congruential Method

| | Uniformity $\chi^2_9$ | Frequency $\chi^2_9$ | Gap $\chi^2_{19}$ | Coupon Collector $\chi^2_8$ | K S Statistics D | Run Test $\chi^2_6$ | Serial Correlation Test $\chi^2_5$ |
|---|---|---|---|---|---|---|---|
| | .2800 | .1900 | 14.8471 | 11.3988 | .0035 | 24.4088 | 0.7881 |
| | .2400 | .0900 | 14.9420 | 9.6922 | .0041 | 75.6139 | 0.4293 |
| | .1000 | .3450 | 14.3843 | 6.0524 | .0027 | 165.3054 | 0.5878 |
| | .4600 | .3150 | 15.6748 | 16.2397 | .0036 | 20.8740 | 0.5696 |
| | .3000 | .2050 | 20.6432 | 13.7810 | .0042 | 172.1182 | 1.4594 |
| | .2400 | .1450 | 17.2525 | 4.6264 | .6032 | 19.2076 | 0.5822 |
| | .1800 | .1550 | 12.6771 | 7.2356 | .0058 | 23.9380 | 1.4586 |
| | .7400 | .4050 | 11.7977 | 5.9722 | .0045 | 67.6091 | 0 6280 |
| | .1400 | .2550 | 27 0823 | 8.3717 | .0030 | 20.5639 | 0.0534 |
| | .1400 | .1500 | 9.0438 | 14.0375 | .0059 | 20.4113 | 0.6135 |
| Minimum value | .1000 | .0900 | 9.0438 | 4.6264 | .0027 | 20.4113 | 0.0534 |
| Maximum value | .7400 | .4050 | 27.0823 | 16.2397 | .0058 | 172.1182 | 1.4594 |

## REFERENCES


1. Dudewicz, E. J. (1975). Random numbers : The need, the history, the generators. *Statistical Distribution in Scientific Work*, G. P. Patil el al. eds., Vol. 2. D. Reidel, Dodrecht. 25-36.
2. Dudewicz, E. J. and Ralley, T. C. (1981) The Handbook of Random Number Generation and Testing with TESTRAND Computer Code. Columbus, Ohio. American Science Press.
3. Fisher, R. A. and Yates, F. (1938). *Statistical Tables and Biological, Agricultural and Medical Research.* Oliver & Boyd, Edinburgh.
4. Colomb, S. W. (1967). *Shift Register Sequences.* Holden — Day. San Francisco.
5. Greenwood, R. E. (1955). *Coupon Collector's test for random digits.* MTAC 9, 1-5.
6. Hammer, P. C. (1951). *The mid-square method of generating digits.* Nat. Bur. Stand Appl. Math. Ser., 12—33.
7. Hammersley, J. M. and Handscomb, D. C. (1964). *Monte Carlo Methods.* Wiley, New York.
8. Hull, T. E. and Dobel, A. R. (1962). *Random number generators.* STAM Rev., 4, 230—253.
9. Jansson, B. (1966). *Random Number Generator.* Victor Pettersons, Stockholm.
10. Kennedy, Jr. W. J. and Gentle, J. E. (1980). *Statistical Computing.* Marcel Dekker, Inc. New York.
11. Kendall, M. G. and Babigton—Smith, B. (1951). *Tables of Random Samplying Numbers.* Cambridge Univ. Press, Cambridge.
12. Kuo, S. S (1965). *Numerical Methods and Computer.* Addison—Wesley, Reading, Mass.
13. Knuth, D. E. (1981). *The Art of Computer Programming*, 2, Semi-numerical Algorithms, 2nd edition, Reading, Mass : Addison—Wesley.
14. Kurita, Y. (1977). *Choosing parameters for congruential random numbers generators.* In Recent Development in Statistics. Barra, J. R. etal. eds., North—Holland, Amsterdam, 697—704.
15. Lehmer, D. H. (1951) *Mathematical method in large scale, computing units.* Proc. of the 2nd Symposium on Large Scale Computing Machinery. Harvard Univ. Press. Combridge 141—146.

16. Learmouth, G. P. and Lewis, P. A. W. (1973). *Statistical tests of some widely used and recent proposed uniform random number generators.* Proc. of Computer Sci. and Statistics 7th Annual Symposium on the interface, Kennedy W. J. ed., Statistical Laboratory Iowa State Univ. Ames, 163—171.

17. Lindgren, B. W. (1962). *Statistical Theory.* Macmillan Pub. Co., New York.

18. Miyatake, 0, Yoshizawa, Y., Inove, H. and Ichimura, M. (1981). *Random number generated by a physical device.* Proc. of Contributed Papers of the 43rd Session of ISI, Buenos Aires, Argentina, 291—294.

19. Metropolis, N. and Ulam, S. (1949). *The Monte Carlo method,* J. Amer Statistic. Assoc., 44, 335—341.

20. Marsglia, G. (1972). The structure of Linear congruential sequences. *Applications of Number Theory to Numerical Analysis,* ed. S. K. Zaremba Academic Press, New York 249-286.

21. Nance, R. E. and Overstreet, C. (1972). *A bibliography on random number generation.* Comp. Rev., 13, 495—508.

22. Peskun, P. H. (1980). Theoretical tests for choosing the parameters of the general mixed linear congruential psuedo-random numbers generator. *J. Stat. Comp. Simmulation,* II, 281—305.

23. Rand Corporation (1955). *A Million Random Digit with 100,000 Normal Deviates.* The Free Press Glencoe, Illinois.

24. Ripley, B. D. (1983). Computer generation of random variables. A tutorial, *Inter. Stat Rev.*, 51, 301—319.