# RANDOM NUMBERS: IS CONGRUENTIAL BEST

John Huseby
California State Univ., Chico

With the recent publication of several computer algorithms for generating random numbers some researchers might wonder if the ubiquitous congruential generator is defective. In a search for an answer this paper examines the concept of randomness and probability, and the relationship of random numbers to the problems they are being used to solve.

When working on a simulation problem I like to start with a ranking of solutions to a particular question in order of desirability. Starting from the top I go down the list until I find a solution I can feasibly employ. To answer the question of how would I prefer to simulate a distribution I would suggest the following ordering:

1. sample from a theoretical distribution
2. sample from an empirical distribution created from sample data

Having made this choice the next question might be "now should the data be sampled?"

1. directly
2. inverse CDF
3. discrete simulation
   a. rejection sampling
   b. composition method
   c. other methods

The choice of any of these methods involves the generation of random numbers. We then need to choose a method of generating those numbers. If a manual procedure is suitable I would suggest a set of icosahedral, 20 face, dice sold by the Japanese Standards Association. They offer two sets. The A set, intended for research applications is precisely constructed of durable material, and sells for $10.00. The B set is intended for general and educational use. The Bset sells for $4.00. The address is 1-24 Akasaka 4, Minato-ku, Tokyo, 107 Japan. For most needs we invoke a computer random number algorithm. What if we were to try to define the "random" sequence produced by the computer?

In probability theory random variables are real functions on a sample space. However we may even question the definition of probability. In that probability seems to attempt to unite unknowns of the universe into theory we might attempt of develop experiments to explain theory rather than the reverse. If regular theory could explain the observed phenomena we should not require probability. In practice probability is sometimes defined to exist (1), but there may be alternative interpretations (2).

In trying to find a way to generate numbers to solve problems we can't define we happen upon the theory of randomness. It is hard to think of a mathematical sequence that is random in an intuitive sense because random implies a process that is not reproducable. Physical processes seem to be the only thing that are not reproducable. The people who produced the one million random digits for RAND (3), even after carefull tuning up of their generating machine, felt compelled to "rerandomize" the numbers.

Compromising our intuition, we agree that a random sequence can be deterministic, and we look at mathematical sequences of numbers. The idea is to define an infinite sequence that satisfies the established criteria for randomness. Then, one attempts to find a finite series with the same attributes. A sequence, $S_n$, in this context refers to a series of numbers, $r_i$, $0 \le i < n$ with $0 \le r_i < 1$. Examine an n element sequence, $S_n$, element by element, and count the $r_i$ that satisfy $a \le r_i < b$. Let this number be I(n). If

$$\lim_{n \to \infty} I(n)/n = b - a$$

the sequence is defined to be equidistributed. Let

$$Pr(S_n) = \lim_{n \to \infty} I(n)/n$$

Given k - r dimensional subsequences with bounds $0 \le a < b \le 1$, i = 0, r, the whole sequence is k-distributed if it is independently equidistributed in each dimension for k. The equivalent one dimensional model states that if

$$Pr(a_1 \le r_{m+1} < b_1,\ a_2 \le r_{m+2} < b_2,\ \ldots,\ a_k \le r_{m+k} < b_k) = (b_1 - a_1)(b_2 - a_2) \ldots (b_k - a_k)$$

the sequence is k-distributed. If a sequence is k-distributed for all positive integers k then, it is $\infty$-distributed, or completely equidistributed. While Franklin(4) does not say that an $\infty$-distributed sequence is random he notes that equidistribution is a property of a truly random sequence. Knuth (5) asks the question "does $\infty$-distributed = random?" He then states a definition that answers the question affirmatively. He also states that he has created an algorithm to compute a sequence of real numbers which is $\infty$-distributed. He carries out an interesting dialog about randomness, and refines the initial conjecture.

What does all of this mean to us who are interested in using computer generated random variables to solve problems? In one sense very little, for in an infinite random sequence it is entirely possible to have some very unacceptable finite strings, ten thousand consecutive sevens for instance. But it gives a hint where to look for random finite sequences. An excellent paper by Franklin (4) discusses theorems concerning the degree of equidistribution of some numeric sequences. Among the topics discussed are the following:

1. the series $x_n = \{n\alpha\}$ is equidistributed for irrational $\alpha$
2. multiply sequences, $x_{n+1} = \{Nx_n + \theta\}$, for N integer > 1, are equidistributed for almost all $x_0$, but not guaranteed even if $x_0$ irrational.
3. for almost all $\theta > 1$ the sequence $x_n = \{\theta^n\}$ is equidistributed. The derived r-dimensional sequence from $x_n = \{\theta^n\}$ is completely equidistributed.
4. Borel's proof that for almost all positive numbers $x_0 < 1$ the digits 0 - (N-1) appear with equal frequency (6,7). A book, (8), has been recently published on the subject.

A new approach toward the definition of theoretical randomness recently appeared in "Scientific American" (9).

Two pioneers in theoretical randomness, Weyl and Borel expounded their theorems prior to 1930. In 1934 Kendall and Babington Smith (10), in discussing the subject of randomness in samplint, note, "It appears, therefore that we cannot hope to define a random sample in terms of the properties of the sample itself, but only as a member of a class of samples." They also define a locally random set as a member of the possible sets of size N which approximately conform to expectations. They propose four tests, which, if passed, would qualify a sequence as locally random. The tests were the frequency test, serial test, poker test, and the gap test. In the 40 years since this article was published we have mushed together the concepts of pure randomness and local randomness. A better name for our computer random number algorithms would be locally randomized number selection techniques.

Purely random sequences may not demonstrate local randomness. Are locally random sequences purely random? Do we care? In the same paper that von Neumann made his much quoted statement concerning users of arithmetical methods to produce random digits being in a state of sin he also related, "We are dealing with mere "cooking recipes" for making digits; probably they cannot be justified, but should merely be judged by their results. Some statistical study of the digits generated by a given recipe should be made, but exhaustive tests are impractical. If digits work well with one problem, they seem usually to be successful with others of the same type."(19)

What constitutes a reasonable set of tests for locally random numbers? The frequency test, serial test, poker test, gap test, test of runs, coupon collector's test, and serial correlation are the common tests. Gruenberger (11) suggested the $d^2$ test for testing points generated in the unit square for monte carlo applications. Knuth (5) describes a permutation test. For all the literature devoted to the serial correlation test Knuth tells us that it and the frequency test are the weakest because they are so easily passed. He recommends the run test and the spectral test. The spectral test is based on the finite Fourier transform, and Knuth claims it is far superior to other tests used. All good random number generators pass it, but all linear congruential sequences which have been found to be bad fail it. The need for some testing is evident. Statistical justification for different tests would vary with the application. (5,12) discuss the subject.

We must be very carefull not to lump the sins of our simulation model onto the random number generator. Statistics are based on repeated measures of all levels under study. In simulation we sometimes load all of our sampling at the first level and forget higher levels. If we are simulating a queue we consider a range of subjects in that queue. We are less likely to consider that queue as one point in the range of queues. For an excellent example of the subject consult Kerrich (2) and Feller (13) under random walks, last visit and long leads. (14) is also a good reference for modeling

Suggestion for future work: think about randomness. Does it embody the concepts we need, or is there something better? Can we quantify randomness. Now it is dichotomous, either a series is random or it is not. Might it be possible to set confidence limits on a random sequence? On a more pragmatic level, we need to trace down and document all of the rumors about random number generators. The congruential method is simple to apply, but it has specific requirements which must be met. Knuth (5) points out an improvement that can be made by changing the modulus from $2^n$ to $2^n$ plus or minus 1. We should document generators for various seed values, multipliers, arithmetic processors and word lengths.

I would like to see some reports on sampling methods. I am partial to what is called perturbation theory, sensitivity analysis, or in statistics, robustness. I am looking for various algorithms allowing one to generate pairs of random variates with specified correlation, or moments. The compound or multiple sequence congruential generators need study. They seem to produce numbers which pass tests, but they throw unknowns into the method.

To conclude, I would like to give my answer to the title question, then explain it. Yes, the congruential method is best. It generates numbers which pass the requiared tests (16). The generator is well defined in the sense that we know the conditions for it to produce its maximum period, and we can even specify different full sequences based on the initial conditions (16, p 77). Barnett (17) relates the results of relaxing constraints on the generator.

Greenberger (18) documented second order correlation deficiencies in the congruential method. Whittlesey (20) notes poor results for the congruential method when tested for autocorrelation. He suggests the Tauseworth generator to solve that problem. There is very little evidence against the congruential method. This is probably indicative of the lack of quantitative methods in simulation. On the positive side there have been some useful studies published in Simuletter, the newsletter of the SIGSIM group of ACM (21, 22). They have been valuable in collecting available information about random number generators.

## References


1. Lindgren B.W., STATISTICAL THEORY, 2nd ed., Macmillan, 1968.
2. Kerrich J.E., "Random remarks (with references to Runs of Luck), American Statician, 15, 3, 16-20, 1961.
3. Brown G., "History of RAND's Random Digits -Summary," NBS Applied Math. Series, Number 12, 31-32, (1951).
4. Franklin J., "Deterministic simulation of random processes", Math. of Comp., 17, 81, 28-59, Jan., 1963.
5. Knuth D., SEMINUMERICAL ALGORITHMS , vol 2, Addison-Wesley, 1971.
6. Niven I. and Zukerman H., "On the definition of normal numbers", Pacific J. of Math., 1, 1, 103-109, March, 1951.
7. Zane B., "Uniform distribution modulo m of monimials", Am Math. Monthly, 71, 162 - 164, 1964.

8. Kuipers L. and Niederreiter H., UNIFORM DISTRIBUTION OF SEQUENCES, John Wiley, 1973.

9. Chaitin G., "Randomness and mathematical proof", Scientific Am., 232, 5, 47-52, May, 1975.

10. Kendall and Smith B., "Randomness and random sampling numbers", JRSS, Part I, 147-165, 1938.

11. Gruenberger F. and Mark A., "The $d^2$ test of random digits", MTAC, 5, 109-110, 1951.

12. Brownlee K.A., STATISTICAL THEORY AND METHODOLOGY IN SCIENCE AND ENGINEERING, 2nd ed., John Wiley, 1965 , 221-240.

13. Feller W., AN INTRODUCTION TO PROBABILITY THEORY AND ITS APPLICATIONS, vol I, 3rd ed., John Wiley, 1970, 73-83.

14. Hammersley J.M. and Handscomb D.C., MONTE CARLO METHODS, Meuthen and Co., London, 1975.

15. Mize J., "Multiple sequence random number generators on simulation results", Proc. of the Winter Simulation Conf., 67-76, 1972.

16. Jansson B. RANDOM NUMBER GENERATORS, Victor Pettersons Bokindustri Aktiebolag, Stockholm, 1966.

17. Barnett V.D., "The behavior of pseudo-random sequences generated on computers by the multiplicative congruential method", Math. Comp., 16, 63-69, 1962.

18. Greenberger M., "Method in randomness", CACM, 8, 3, 177-179, 1965.

19. von Neumann J., "Various techniques used in connection with random digits", NBS App. Math. No. 12, 36-38, 1951.

20. Whittlesey R.B., " A comparison of the correlational behavior of random number generators for the IBM 360", CACM, 11, 9, 641-644, Sept., 1968.

21. Babad J. and Stohr E., "The effect of different GPSS random number generators on simulation results", Simuletter, VI, 4, 55-61, 1975.

22. Overstreet C. "Evaluation of the Werner random number generator", Simuletter, VI, 4, 75, 1975.