JOURNAL OF RESEARCH of the National Bureau of Standards—B. Mathematical Sciences
Vol. 73B, No. 1, January–March 1969

Relations Within Sequences of Congruential Pseudo-Random Numbers

Peter H. Verdier

Institute for Materials Research, National Bureau of Standards, Washington, D.C. 20234

(August 20, 1968)

Conditions are exhibited under which simple, approximate linear relations may be found between sets of successive choices made by congruential pseudo-random number generators. These relations imply that the distributions in n-dimensional space produced by such generators can be very highly nonuniform. The results are illustrated with several examples. Restrictions on the parameters of the generator to minimize difficulties of this sort are discussed.

Key Words: Congruential generators; multidimensional distributions; random number generators.

1. Introduction

Consider a congruential pseudo-random number generator of the form:

$$X_{j+1} \equiv MX_j + C \pmod{P}, \tag{1}$$

where all the quantities involved are nonnegative integers. Generators of this type are commonly used to provide sequences of pseudo-random numbers [1].[1] For implementation on digital computers, P is frequently chosen to be the word size of the computer and M is chosen as $2^a + N$ or $10^a + N$, where N is a small integer, for binary and decimal machines, respectively. In practice, such a generator is used to make a choice between some number B of alternatives. For this purpose, we form the quantity $[BX_j/P]$, where the square brackets denote integer part. It is the purpose of this paper to point out that relatively short sequences of such quantities may be very highly nonrandom, in the sense that if such a sequence is taken to represent a unit hypercube in a hypercube of side B, many of the possible unit hypercubes can never be selected. That is to say, knowledge of the last k choices produced by the generator gives us considerable information about what the next choice will be, for distressingly small k. Nonuniform distributions of triplets, corresponding to $k=2$, have been observed by MacLaren and Marsaglia [2], for several congruential generators. As a special case, the results in this paper partly explain their findings.

2. Results

Consider a generator of the form (1), with

$$M = R + N, \tag{2}$$

$$R^k \equiv r \pmod{P}, \tag{3}$$

[1] Figures in brackets indicate the literature references at the end of this paper.

where N, k, and r are integers small in magnitude. By successive application of eq (1), we obtain:

$$X_{j+m} \equiv M^m X_j + (M^m - 1)C/(M-1) \qquad (\text{mod } P). \tag{4}$$

By using (3) and (4), we find:

$$\sum_{m=0}^{k} \binom{k}{m} (-N)^{k-m} X_m \equiv rX_0 + C\{R^k - (1-N)^k\}/(M-1) \qquad (\text{mod } P), \tag{5}$$

where we have set $j = 0$ for convenience.

Equation (5) states that given the last k values of X, the next one is determined. This fact of itself is neither novel nor alarming; we already know from eq (1) that precise knowledge of just the last X tells us what the next one will be. However, we now show, with the aid of eq (5), that *approximate* knowledge of the last k values of X can enable us to predict the approximate value of the next. To show this, we suppose that the generator is to be used repeatedly to pick one of B alternatives, where B is a positive integer less than P. We first form from each X_j a new quantity b_j, defined by:

$$b_j = [BX_j/P], \tag{6}$$

so that $0 \leq b_j \leq B-1$. In effect, we have divided the range of values of X into B intervals, labeled by the values b_j. Then we can write:

$$BX_m/P = b_m + \epsilon_m, \tag{7}$$

where $0 \leq \epsilon_m \leq (P-1)/P < 1$. We now transform the congruence (5) into a congruence with modulus B. We first subtract the right-hand side of (5) from both sides, obtaining on the left-hand side an integer congruent to zero (mod P). Division of the left-hand side by P results by definition in an integer, and subsequent multiplication by B results, again by definition, in an integer congruent to zero (mod B). Rearranging terms to restore the transformed right-hand side, we obtain:

$$\sum_{m=0}^{k} \binom{k}{m} (-N)^{k-m} b_m \equiv rb_0 + \zeta - \eta \qquad (\text{mod } B), \tag{8}$$

where we have set:

$$\zeta \equiv \frac{BC(R^k - (1-N)^k)}{P(M-1)} \qquad (\text{mod } B), \tag{9.1}$$

$$\eta \equiv \sum_{m=0}^{k} \binom{k}{m} (-N)^{k-m} \epsilon_m - r\epsilon_0 \qquad (\text{mod } B). \tag{9.2}$$

By considering the maximum possible sizes of the positive and negative terms in eq (9), using the fact that the ϵ's are all less than unity, we find from eq (8) that given the last k of the b's, the next one can only be one of S possibilities, where:

$$S = (1+|N|)^k + |(-N)^k - r| - |N|^k - 1, \text{ if } \zeta \text{ is an} \tag{10.1}$$

integer,

$$S = (1+|N|)^k + |(-N)^k - r| - |N|^k \text{ otherwise.} \tag{10.2}$$

(We ignore the trivial case $S=1$, obtained when the BX_m are all multiples of P.) If S is less than B, this means that there will be sequences of length $k+1$ which cannot occur at all. If we now choose:

$$R=d^a, \tag{11}$$

$$P=d^w, \tag{12}$$

where d is 2 or 10 for binary or decimal computers, respectively, we find that eq (3) is satisfied by $r=0$ and:

$$k=[(w-1)/a]+1, \tag{13.1}$$

the square brackets again denoting integer part. In the frequently employed special case of $C=0$ and M and all the X_j odd, eq (13.1) may be replaced by the stronger relation:

$$k=[(w-1-[\log_d 2])/a]+1. \tag{13.2}$$

3. Two Examples

In this section, we apply the results of the preceding section to two generators which have been discussed in the literature and found satisfactory by other tests. The first [3] is a multiplicative congruential generator with $M=2^{18}+3$, $C=0$, $P=2^{35}$, and the X_j all odd. From eq (13) we find $k=2$, so we expect a three-term recursion relation, corresponding to a set of parallel planes in three-space. From eq (8) we find that the equation obeyed is $b_{j+2}\equiv 6b_{j+1}-9b_j-\eta \pmod B$, and we can easily show from eq (9) that $-6<\eta<10$, leading to only 15 possibilities for b_{j+2}, given b_j and b_{j+1}. Thus, if this generator is used to select unit cubes at "random" in a cube of side greater than 15, some cubes will never be selected.

For the second example [4], we choose a mixed congruential generator with $M=2^{11}+1$, $C=1446722743$ octal, $P=2^{28}$. Then from eq (13), $k=3$, so we shall have a four-term recursion relation, given by eq (8). According to eq (10), if three successive b's are known, the next can only be one of 7 possibilities, if B is a multiple of 64, or 8 possibilities otherwise.

4. Discussion

In the preceding sections, we have shown that under certain circumstances, a generator of the form of eq (1), used with eq (6) to choose from among B possibilities, may have the property that knowing the last several choices made by the generator, we can predict that the next choice will be one of S possibilities, where S may be much smaller than B. This property constitutes a rather serious violation of the usual concept of a random choice. (Its disadvantages would be immediately obvious to, say, a casino operator who proposed to use such a generator in place of the conventional roulette wheel.) It should perhaps be noted that as one might expect from the form of eq (9), choice among the possibilities which *are* allowed is by no means made with equal frequency. Thus, to obtain the appearance of randomness, it is not sufficient merely to ensure that S is not less than B. Rather, one would expect that S needs to be considerably greater than B. Two examples may be found in the work of MacLaren and Marsaglia [2]. Among the generators for which they found poor triplet distributions were those with $M=2^{17}+3$, $C=0$, $P=2^{35}$, and $M=10^5+3$, $C=0$, $P=10^{10}$. For both generators, we find $k=2$, so that we expect triplets to be poorly distributed. From eq (10) we find $S=15$ in both cases. Since $B=10$ for the work in reference 2, no triplets are absolutely forbidden by eq (8). However, a chi-square test showed the distribution to be extremely nonuniform.

It is clear that in the cases of interest eq (10) will tend to be dominated by the first term on its right-hand side, i.e. $(1+|N|)^k$. The form of this term might suggest the use of a small multiplier, which from eq (3) will prohibit a small k. However, small multipliers give rise to other undesirable characteristics [5]. We need, therefore, to choose a nonsmall multiplier and a modulus such that for all possible decompositions of the multiplier according to eq (2) and (3), the value of S predicted by eq (10) is as large as possible (or, at any rate, considerably larger than some preselected number B of choices for which the generator is used). We do not see a way of going from eq (2), (3), and (10) to a formula for choosing M and P. However, these equations may be used to test a given trial M and P, in an obvious if somewhat laborious way.

As already remarked, generators of the forms implied by eqs (11) and (12) are especially common, because the implied division need not be performed explicitly. For generators of this form, eq (13) guarantees a recursion relation with $r=0$, for a value of k which will be small unless the multiplier is quite small. In this case, keeping S large amounts to requiring that the multiplier lie far from any multiple of all powers of d which are high enough to give rise to a small value of k. Again, it is not easy to see how this requirement can be turned into a prescription for choosing an optimum multiplier, but it is relatively easy to test trial multipliers.

If the modulus P is prime, eq (3) has no solutions with $r=0$. However, this does not make the choice of multiplier any easier. One can see from eq (10) that the effect of a small, nonzero value of r may be to decrease the value of S. Therefore, it is not clear that anything is gained by rejecting moduli of the form of eq (12).

Finally, we may note that whatever effect a nonzero constant C may have upon other statistical properties of a generator, it has virtually no effect upon the recursion properties discussed here. From eq (10), we see that a nonzero value of C can increase the number of possibilities by at most one, and that only for a suitable choice of B.

5. Conclusions

It seems clear that injudicious use of a congruential pseudo-random number generator can easily lead to highly nonuniform distributions in more than two dimensions, and that neither a nonzero additive constant nor a prime modulus will avoid this difficulty. Furthermore, routine testing of the multidimensional distributions produced by a generator may well not show up difficulties of the sort discussed here. Since B^k cells of storage are required to test a k-dimensional distribution, if k is greater than three, the storage capacities of present-day computers are likely to limit B to smaller values than those occurring in practice.

Note added in proof: A recent paper by G. Marsaglia [Proc. Nat. Acad. Sci. **61**, 25 (1968)] also deals with limitations on the randomness of n-tuples produced by congruential generators. The limitations established are weaker than those set forth here, but are independent of the choice of multiplier.

6. References

[1] For a review and bibliography, see Hull, T. E., and Dobell, A. R., Random number generators, SIAM Review **4**, 230–254 (1962).

[2] MacLaren, M. D., and Marsaglia, G., Uniform random number generator, J. Assoc. Comput. Mach. **12**, 83–89 (1965).

[3] Hull, T. E., and Dobell, A. R., Mixed congruential random number generators for binary machines, J. Assoc. Comput. Mach. **11**, 31–40 (1964).

[4] Peach, P., Bias in pseudo-random numbers, J. Am. Stat. Assoc. **56**, 610–618 (1961).

[5] Greenberger, M., An a priori determination of serial correlation in computer generated random numbers, Math. Comp. **15**, 383–389 (1961); Math Comp. corrigenda **16**, 126 (1962).

(Paper 73B1–285)