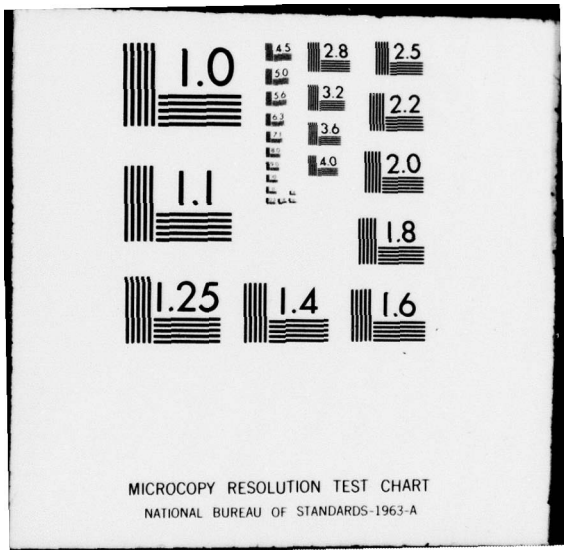
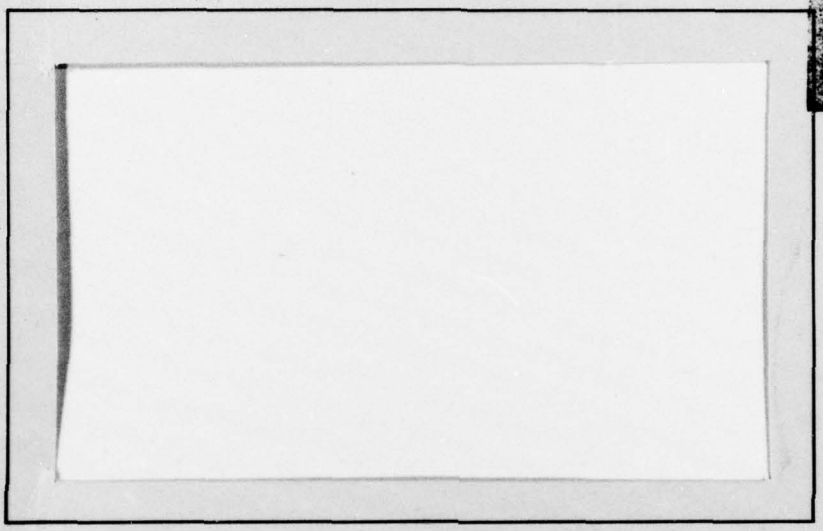
AD-A063 858 NORTH CAROLINA UNIV AT CHAPEL HILL F/G 12/1
A STATISTICAL EVALUATION OF MULTIPLICATIVE CONGRUENTIAL GENERAT--ETC(U)
DEC 78 G S FISHMAN, L R MOORE N00014-76-C-0302
UNCLASSIFIED TR-78-11 NL

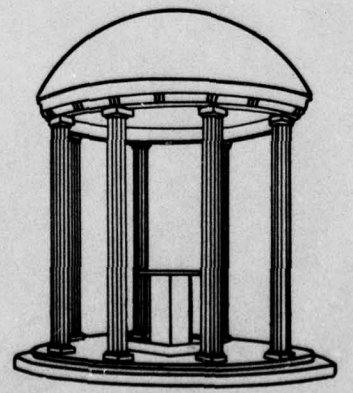
END
DATE FILMED
4 -79
DDC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
6.3
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

OPERATIONS RESEARCH AND SYSTEMS ANALYSIS



UNIVERSITY OF NORTH CAROLINA
AT CHAPEL HILL



(6) A STATISTICAL EVALUATION OF MULTIPLICATIVE CONGRUENTIAL GENERATORS WITH MODULUS $2^{31}-1$.

(10) George S. Fishman and Louis R. Moore

(9) Technical Report 78-11

(11) December 1978

(12) 34 pp

(14) TR-78-11

Curriculum in Operations Research
and Systems Analysis

University of North Carolina at Chapel Hill

(15) This research was supported by the Office of Naval Research under contract N00014-76-C-0302. Reproduction in whole or in part is permitted for any purpose of the United States Government.

ABSTRACT

This paper presents the results of empirically testing eight alternative multipliers for a multiplicative congruential generator with modulus $2^{31}-1$. The LLRANDOM number package (Learmonth and Lewis, 1973) uses one of the multipliers, the simulation programming language SIMSCRIPT II uses a second and the remaining six are the best of 50 candidate multipliers studied by Hoaglin (1976) using the theoretical spectral and lattice tests. The battery of tests raises serious doubts about three of the multipliers, including the one in LLRANDOM. The power of the tests is demonstrated by their rejection of RANDU, a notably poor random number generator. A comparison of the results for the eight multipliers with the eight worst multipliers (with regard to 2-tuples) in Hoaglin (1976) failed to show any apparent gross differences. Since examination of performance on the lattice test revealed that the 16 multipliers clustered in performance when compared to RANDU, one may conjecture that the poorer lattice test performance for the worst eight is too subtle for detection by our empirical tests. Since this failure may not be as serious as the lattice test implies, one may want to revise upward the criteria for acceptable performance on the lattice tests and similarly on the spectral test. In section 9 an analysis of algorithms for generating random numbers reveals that the choice of coding can significantly affect execution time for different multipliers.

1. Introduction

This paper presents the results of an empirical evaluation of eight suggested values for the multiplier A in the prime modulus multiplicative congruential random number generator

$$Z_i \equiv AZ_{i-1} \pmod{2^{31}-1} . \tag{1}$$

Lewis et al. (1969) suggested multiplier I and it is in common use in the LLRANDOM random number package (Learmonth and Lewis, 1973), in APL (Katzan, 1971), in the simulation program language SIMPL/1 (IBM, 1972) and in the IMSL Library (1977). Payne et al. (1969) suggested multiplier II, and its most common use is in the simulation programming language SIMSCRIPT II. The remaining six suggestions come from a study of 50 multipliers by Hoaglin (1976) who evaluated them as best using the spectral test (Coveyou and MacPherson, 1967) and the lattice test (Marsaglia, 1971). It is well known (Marsaglia, 1968) that all linear congruential generators are flawed. The spectral and lattice tests provide theoretical methods of assessing the seriousness of the flaw in a multiplier. These tests showed that the six multipliers considered here had less serious flaws than the remaining 44 multipliers studied.

In the present paper we examine the performance of these eight multipliers when subjected to a battery of tests designed to detect departures from randomness in a sequence of numbers. Let $U_i = Z_i/(2^{31}-1)$ for $i = 1,\ldots,n$. Hypotheses to be tested include:

H_0. $\{U_i\}$ is a sequence of i.i.d. random variables.

H_1. U_i has a uniform distribution on $(0,1)$.

H_2. (U_{2i-1},U_{2i}) has a uniform distribution on the unit square.

H_3. $(U_{3i-2},U_{3i-1},U_{3i})$ has a uniform distribution on the unit cube.

H_4. H_0, H_1, H_2 and H_3 hold simultaneously. (If H_0 and H_1 hold, H_2 and H_3 are redundant.)

For each multiplier, the results include a separate evaluation of the test of each hypothesis H_0, H_1, H_2 and H_3 and a collective evaluation via H_4 of the performance on the four hypotheses taken together. The collective evaluation reveals that three of the candidate multipliers (I, III and VII) are seriously suspect.

In any empirical study of this sort, one hopes that the collective power of the tests employed is sufficiently high to detect departures from randomness when they exist. We demonstrate this ability to detect departures in an application of these same tests to samples generated by RANDU, the random number generator used until recently in IBM's Scientific Subroutine Library. Marsaglia (1971) has demonstrated the notably poor performance of this generator on the lattice test.

In a further attempt to demonstrate the power to detect nonrandomness, we applied the tests to the eight multipliers in Hoaglin's study that performed poorest on the spectral and lattice tests for 2-tuples. An analysis of the results of the best and worst in the Hoaglin study showed no difference of note, leading to several conjectures. One is that the tests cannot consistently detect departures of the magnitude examined. (However, they can detect departures as substantial as RANDU's.) Another is that the criteria suggested in Knuth (1969) and Marsaglia (1971) for a multiplier to "pass" the spectral and lattice tests, respectively, may be considerably more conservative than first thought. We discuss these conjectures in more detail in Section 8.

2. Distribution Testing Procedures

For each multiplier, we collected 100 independent samples, each with $n = 200{,}000$ observations. For sample i and hypothesis j a statistic $T_{i,j}$ was computed. Then for hypothesis j, $T_{1,j}, T_{2,j}, \ldots, T_{100,j}$ were subjected to

a battery of tests. Let $T_{i,j}$ have cumulative distribution function (c.d.f.) G_j under hypothesis j. Then $P_{i,j} = 1-G_j(T_{i,j})$ has the uniform distribution on $(0,1)$ and

$$F_{n,j}(t) = \frac{1}{n} \sum_{i=1}^{n} I_{(0,t]}(P_{i,j}) \qquad 0 \le t \le 1, \tag{2}$$

where I denotes the indicator function, is an empirical c.d.f. If hypothesis j is true

$$D_{n,j} = \sup_t |F_{n,j}(t) - t| \tag{3}$$

has the Kolmogorov-Smirnov (K-S) distribution. Also

$$V_{n,j} = \int_0^1 I_{[0,t]}(F_{n,j}(t))dt \tag{4}$$

has the uniform distribution on $(0,1)$ (Dwass, 1958). Finally, for large n

$$A^2_{n,j} = n \int_0^1 \{[F_{n,j}(t)-t]^2/t(1-t)\}dt \tag{5}$$

has a distribution given by Anderson and Darling (1952).

The motivation for goodness-of-fit tests based on $D_{n,j}$, $V_{n,j}$ and $A^2_{n,j}$ arises from the distinct departures from behavior under hypothesis j that each is designed to detect. The quantity $D_{n,j}$ measures the maximal absolute deviation between the empirical and the hypothesized c.d.f.'s, $V_{n,j}$ measures the proportion of $F_{n,j}$ that lies below the hypothesized c.d.f. and $A^2_{n,j}$ measures the extent of deviation, principally in the tails of the empirical c.d.f.

3. Testing for Independence H_0

To test H_0 we relied on runs-up-and-down statistics. Let $R_{i,k}$ and $\tilde{R}_{i,k}$ denote the numbers of runs up and down, respectively, of length k on

replication i. Then the statistics for H_0 are

$$
\begin{aligned}
(6) \qquad T_{i,0} = \sum_{k,\ell=1}^{6} \{ & c_{k\ell}[R_{i,k} - E(R_{i,k})][R_{i,\ell} - E(R_{i,\ell})] \\
& + d_{k\ell}[R_{i,k} - E(R_{i,k})][\tilde{R}_{i,\ell} - E(\tilde{R}_{i,\ell})] \\
& + c_{k\ell}[\tilde{R}_{i,k} - E(\tilde{R}_{i,k})][\tilde{R}_{i,\ell} - E(\tilde{R}_{i,\ell})]\} \qquad i = 1,\ldots,100
\end{aligned}
$$

and each asymptotically has the chi-squared distribution with 12 degrees of freedom. The quantities $c_{k\ell}$ and $d_{k\ell}$ are computable from the covariance matrix of $R_{i,1},\ldots,R_{i,6},\tilde{R}_{i,1},\ldots,\tilde{R}_{i,6}$ (Levene, 1953). For 200,000 observations one has

$$
\begin{aligned}
E(R_{i,1}) &= E(\tilde{R}_{i,1}) = 41670 \\
E(R_{i,2}) &= E(\tilde{R}_{i,2}) = 18330 \\
E(R_{i,3}) &= E(\tilde{R}_{i,3}) = 5280 \\
E(R_{i,4}) &= E(\tilde{R}_{i,4}) = 1150 \\
E(R_{i,5}) &= E(\tilde{R}_{i,5}) = 200 \\
E(R_{i,6}) &= E(\tilde{R}_{i,6}) = 30 \\
E(R_{i,7}^{+}) &= E(\tilde{R}_{i,7}^{+}) = 4.4 \ .
\end{aligned} \qquad (7)
$$

Here $R_{i,7}^{+}$ and $\tilde{R}_{i,7}^{+}$ denote the numbers of runs up and down, respectively, of length 7 or more. Although a statistic similar to (6) that incorporates either $R_{i,7}^{+}$ or $\tilde{R}_{i,7}^{+}$ can be constructed and asymptotically has a chi-squared distribution with 13 degrees of freedom, the small value of $E(R_{i,7}^{+})$ in the present case encouraged us to work with (6) to avoid any discretization error that, say, $R_{i,7}^{+}$ might induce. Note that inclusion of $R_{i,7}^{+}$ and $\tilde{R}_{i,7}^{+}$ in (6) would produce a degenerate distribution for $T_{i,0}$ (Wolfowitz, 1944).

By way of interpretation, excessive short (long) runs imply more mixing

(clustering) than one would expect to find in a purely random sequence. The decision to examine runs up and runs down explicitly rather than study their sum was motivated by the finding in Tootill et al. (1971) that runs up and runs down can exhibit distinct behavioral patterns for certain Tausworthe random number generators (Tausworthe, 1965). Although the generator (1) is not of this type, we decided to allow for the possibility of analogous behavioral distinctions.

Column 3 of Table 1 lists the P-values for the test statistics (3), (4) and (5) for each multiplier. In particular, a P-value for the K-S (A-D) test is the probability that a random variable from the K-S (A-D) distribution exceeds $D_{100,0}$ ($A^2_{100,0}$) in value under H_0. For (4) the P-value is $2\min(V_{100,0}, 1-V_{100,0})$, since $V_{100,j}$ is a uniform deviate under H_j. Except for RANDU, the P-values in column 3 give little cause for concern.

4. Testing for Uniformity H_1

To test H_1 we chose a chi-squared goodness-of-fit statistic. Consider K cells on the unit interval each of length 1/K. Let $N_{i,k}$ denote the number of the N observations on replication i that fall into the interval ((k-1)/K, k/K]. Then under H_1

$$(8) \qquad T_{i,1} = \frac{K}{N} \sum_{k=1}^{K} (N_{i,k} - N/K)^2 = \frac{K}{N} \sum_{k=1}^{K} N_{i,k}^2 - N \qquad i = 1,\ldots,100$$

each asymptotically have a chi-squared distribution with K-1 degrees of freedom. Choosing $K = 2^{12} = 4096$ implied a cell width of $1/K = 0.000244140625$ and enabled us to test the first 12 bits of U_i.

Column 4 of Table 1 shows the P-values for $D_{100,1}$, $V_{100,1}$ and $A^2_{100,1}$ for each multiplier. The P-values again arouse little suspicion. The failure to detect nonuniformity in RANDU may be an indication of the poor power of this

Table 1

P-Values for Testing Hypotheses
(RANDU and Multipliers I through VIII)

Multiplier	A	Relative Execution Time[a]	Test	H_0	H_1	H_2	H_3	H_4 min	H_4 max
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
RANDU			K-S	.000***	.339	.000***	.000***	.000***	.245
			V	.001***	.850	.046**	.000***	.000***	.137
			A-D	.000***	.352	.002***	.000***	.000***	.126
I.	16807	1.00	K-S	.228	.245	.286	.929	.067*	.004***
			V	.141	.415	.016**	.462	.222	.000***
			A-D	.163	.196	.415	.757	.101	.000***
II.	630360016	3.03	K-S	.600	.888	.396	.154	.840	.830
			V	.071*	.612	.620	.039**	.623	.313
			A-D	.403	.912	.263	.120	.569	.639
III.	1078318381	4.42	K-S	.872	.842	.446	.011**	.006***	.297
			V	.736	.366	.045**	.159	.003***	.566
			A-D	.814	.512	.258	.025**	.011**	.451
IV.	1203248318	4.82	K-S	.743	.632	.500	.117	.880	.099*
			V	.409	.929	.139	.159	.663	.291
			A-D	.774	.890	.448	.193	.818	.067
V.	397204094	2.37	K-S	.281	.203	.604	.851	.648	.269
			V	.500	.255	.556	.363	.214	.412
			A-D	.389	.252	.420	.837	.616	.214
VI.	2027812808	7.28	K-S	.415	.188	.773	.871	.583	.085*
			V	.132	.603	.555	.652	.227	.257
			A-D	.473	.168	.943	.827	.585	.247

Table 1 (continued)

P-Values for Testing Hypotheses

(RANDU and Multipliers I through VIII)

Multiplier	A	Relative Execution Time	Test	H_0	H_1	H_2	H_3	H_4 min	H_4 max
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VII.	1323257245	5.31	K-S	.757	.338	.671	.141	.003***	.483
			V	.375	.955	.551	.431	.029**	.420
			A-D	.495	.245	.551	.186	.003***	.377
VIII.	764261123	3.46	K-S	.674	.229	.207	.806	.436	.117
			V	.969	.567	.474	.491	.229	.158
			A-D	.466	.507	.254	.854	.306	.366

* .05 < P-value ≤ .1 . ** .01 < P-value ≤ .05 . *** P-value ≤ .01 .

[a] Each run was made using the LLRANDOM package with the corresponding multiplier replacing A = 16807 . See Section 9 for discussion of relative execution times.

test or an indication that RANDU has the local property of one-dimensional uniformity.

5. Testing for Bivariate Uniformity H_2

Hypothesis H_2 is designed to detect nonuniformity when the U_i are taken in nonoverlapping pairs or 2-tuples. One motivation for this testing arises from the theoretical observation in Marsaglia (1968) that the randomness of k-tuples becomes more suspect as k increases. The spectral and lattice tests support this observation. In particular, see Hoaglin (1976) and Marsaglia (1971). Ideally one would like to test for the uniformity in distribution of k-tuples of the k-dimensional unit hypercube. In practice, such testing is excessively expensive, even for $k = 2$.

Let us divide the unit interval into K cells, each of width $1/K$. Let N_{ijk} denote the frequency with which nonoverlapping 2-tuples fall into the square $(((j-1)/K, j/K] \times ((k-1)/K, k/K])$ on replication i. For fixed K the quantities

$$T_{i,2} = \frac{K^2}{N} \sum_{j,k=1}^{K} (N_{ijk} - N/K^2)^2 \qquad i = 1,\ldots,100 \tag{9}$$

each asymptotically have the chi-squared distribution with $K^2 - 1$ degrees of freedom. Suppose we had chosen as before $K = 4096$. Then there would be $K^2 = 16777216$ cells. To guarantee a mean of 5 per cell under H_2 would require $N > 80$ million observations per replication or over 8 billion observations per multiplier. Since $2^{31}-1 < 4.3$ billion, such a sample size is not possible using this test procedure. Because of this demonstrated excessiveness, we chose $K = 128$, which required 16384 cells, implied a cell width of 0.0078125 and, for $N = 200{,}000$, a mean of $n/K^2 = 12.21$ per cell under H_2. This choice of K enabled us to study the first 7 bits of each coordinate of a 2-tuple.

Column 5 in Table 1 displays the P-values for $D_{100,2}$, $V_{100,2}$ and $A^2_{100,2}$. Although $V_{100,2}$ appears low for multipliers I and III, we do not regard this suspiciously at this point, since we are simultaneously examining $3\times8 = 24$ P-values. What is of concern to us is when $D_{100,j}$, $V_{100,j}$ and $A^2_{100,j}$ all give low P-values for a given multiplier. We demonstrate cases of this sort shortly when testing H_4. Note the extremely poor showing for RANDU.

6. Testing for Trivariate Uniformity H_3

Hypothesis H_3 tests for nonuniformity when the U_i are taken in nonoverlapping triplets or 3-tuples. The motivation is similar to that for testing H_2. Let $N_{ijk\ell}$ denote the frequency on replication i with which nonoverlapping triplets fall into the cube $(((j-1)/K, j/K] \times ((k-1)/K, k/K] \times ((\ell-1)/K, \ell/K])$ for $j,k,\ell = 1,\ldots,K$ with K specified. Then

$$T_{i,3} = \frac{K^3}{N} \sum_{j,k,\ell=1}^{K} (N_{ijk\ell} - N/K^3)^2 \qquad i = 1,\ldots,100 \tag{10}$$

each asymptotically have the chi-squared distribution with K^3-1 degrees of freedom. Here we choose $K = 2^4 = 16$ which gives $K^3 = 16\times16\times16 = 4096$ cells and an expectation of $N/K^3 = 48.83$ per cell. A perusal of the P-values in column 6 of Table 1 indicates a rejection of H_3 for RANDU and reveals some concern about multiplier III and a lesser concern about multiplier II.

7. An Omnibus Test H_4

In reviewing the P-values for H_0, H_1, H_2 and H_3 one finds it difficult to say that any multiplier except RANDU seriously errs in an omnibus sense. This finding is entirely in keeping with other empirical investigations using tests with vague alternative hypotheses. To overcome this inadequacy, we test H_4: H_0, H_1, H_2 and H_3 are true simultaneously.

Recall that $T_{i,0}, T_{i,1}, T_{i,2}, T_{i,3}$ are the statistics on replication i for hypotheses 0 through 3. Under H_4 $P_{i,j} = 1-G_j(T_{i,j})$ for $j = 0,\ldots,3$ each have the uniform distribution on $(0,1)$. Let

$$Z_{i,j} = \Phi^{-1}(P_{i,j})$$

where $\Phi^{-1}(\cdot)$ denotes the inverse function of the unit normal c.d.f. Let

$$Z_{i,min} = \min(Z_{i,0}, Z_{i,1}, Z_{i,2}, Z_{i,3})$$

(11)

$$Z_{i,max} = \max(Z_{i,0}, Z_{i,1}, Z_{i,2}, Z_{i,3}) .$$

Under H_4 the statistics

$$T_{i,4} = 1-\Phi(-Z_{i,min}, -Z_{i,min}, -Z_{i,min}, -Z_{i,min})$$

(12)

$$T^*_{i,4} = 1-\Phi(Z_{i,max}, Z_{i,max}, Z_{i,max}, Z_{i,max}) ,$$

where $\Phi(\cdot,\cdot,\cdot,\cdot)$ denotes a four-dimensional multivariate normal distribution each have the uniform distribution on $(0,1)$. Under H_4, $T_{i,4}$ is the probability of observing a minimum P-value smaller than $\min(P_{i,0}, P_{i,1}, P_{i,2}, P_{i,3})$ and $T^*_{i,4}$ is the probability of observing a maximum P-value greater than $\max(P_{i,0}, P_{i,1}, P_{i,2}, P_{i,3})$.

Figure 1

Possible Arrangements of $P_{i,0}, P_{i,1}, P_{i,2}$ and $P_{i,3}$

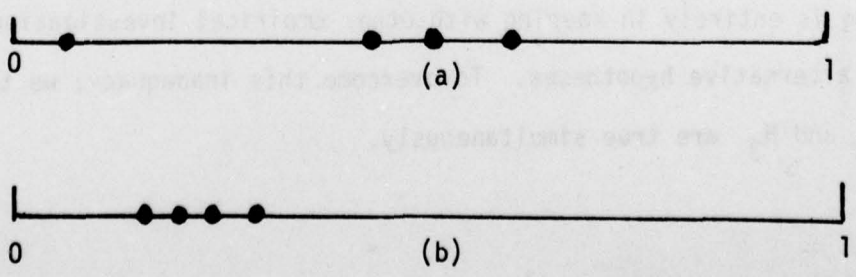


Figure 1 provides a rationale for studying $T_{i,4}$ and $T^*_{i,4}$. Figure 1a shows three "reasonable" $P_{i,j}$'s and one excessively small one. One way to evaluate the extent to which this behavior is consistent with H_4 is to test $T_{1,4},\ldots,T_{100,4}$ for uniformity on (0,1). Figure 1b shows a case in which no $P_{i,0}$, $P_{i,1}$, $P_{i,2}$, or $P_{i,3}$ is small nor is any moderately large. Here one tests $T^*_{1,4},\ldots,T^*_{100,4}$ for uniformity on (0,1) to determine the extent to which the behavior in Figure 1b is consistent with H_4. Although tests that explicitly include the other order statistics can be performed, we felt that the two selected tests represented sufficiently extreme situations that, if significant, would be regarded as unequivocal evidence for disqualifying a multiplier.

Because $P_{i,0}$, $P_{i,1}$, $P_{i,2}$ and $P_{i,3}$ are based on the same sample data of 200,000 observations, treating them as independent would be unreasonable. Instead we concentrate on $Z_{i,0}$, $Z_{i,1}$, $Z_{i,2}$, and $Z_{i,3}$ to acknowledge the dependence appropriately. Under H_4 these quantities have a four-dimensional multivariate normal distribution with mean vector $\underline{0}$ and unknown covariance matrix $\underline{\Sigma}$. Consider the Cholesky decomposition representation

$$\underline{c}\,\underline{c}' = \underline{\Sigma}^{-1}$$

$$\underline{c} = \begin{pmatrix} c_{00} & 0 & 0 & 0 \\ c_{10} & c_{11} & 0 & 0 \\ c_{20} & c_{21} & c_{22} & 0 \\ c_{30} & c_{31} & c_{32} & c_{32} \end{pmatrix}$$

$$c_j = \sum_{i=j}^{3} c_{ij} .$$

Then one can express (12) as

$$T_{i,4} = 1 - \prod_{j=0}^{3} \Phi(-c_j Z_{i,min})$$

(14)

$$T^*_{i,4} = 1 - \prod_{j=0}^{3} \Phi(c_j Z_{i,max}) \ .$$

Since $\underline{\Sigma}$ is unknown, the computation of (14) is not directly possible. Here we estimated $\underline{\Sigma}$ by

$$\hat{\underline{\Sigma}} = ||\hat{\sigma}_{ij}||$$

$$\hat{\sigma}_{ij} = \frac{1}{n-1} \sum_{k=1}^{n} (Z_{k,i} - \bar{Z}_i)(Z_{k,j} - \bar{Z}_j)$$

$$\bar{Z}_i = \frac{1}{n} \sum_{k=1}^{n} Z_{k,i} \qquad n = 200,000$$

from which we estimated c_0, c_1, c_2 and c_3 and (14) by substituting the estimates for c_0, c_1, c_2 and c_3. If $T_{i,4}$ and $T^*_{i,4}$ have the proper distribution under H_4 then estimating $\underline{c}$ will cause the data to appear to follow H_4 more closely than using the true $\underline{c}$. Thus the tests employed here are conservative.

Columns 7 and 8 of Table 1 show the P-values for H_4. Here the test based on the maximum provides strong evidence against multiplier I and the test based on the minimum provides strong evidence against multipliers III and VII. RANDU's failure on the minimum test is consistent with its poor showing on H_0, H_2 and H_3.

In interpreting the results for the minimum or maximum test in column 7 or 8, one needs to pay attention to the issue of multiplicity across the eight multipliers (ignoring RANDU). Under the hypothesis that the eight multipliers each satisfy H_4, the probability of observing at least two K-S test statistics there with P-values $\leq .006$ on the minimum test is .0010. Alternatively, under the same hypothesis, the probability of observing at least one K-S test statistic with P-value $\leq .004$ on the maximum test is $1-(.996)^8 = .0316$. We regard these

probabilities as sufficiently small to reject equality of behavior on the minimum and maximum tests across the eight multipliers. Therefore, we regard the poor behavior of multipliers I, III and VII as real and not simply due to sampling.

8. Other Prime Modulus Multipliers

As mentioned earlier, the six multipliers selected for study from Hoaglin's paper were evaluated there as the best performers on the spectral and lattice tests. A question naturally arises as to how sensitive the battery of empirical tests presented here is to the flaws present in the remaining 44 prime modulus multipliers there. As a first inquiry, we selected the worst eight multipliers there, based on their performance on the spectral and lattice tests for 2-tuples, and subjected them to the same battery of empirical tests. The results in Table 2 show no apparent gross departure from those in Table 1.

To understand why the tests detect the poor behavior of RANDU but do not reject all multipliers IX through XVI, we return to lattice test performance. The lattice test relies on the reduced basis vectors of the lattice of n-tuples. This is the basis that is as close as possible to being orthogonal, given the constraints of the lattice (Hoaglin, 1976). The ratio L_n of the lengths of the longest to the shortest basis vectors provides the figure of merit. A ratio of unity is ideal.

Table 3 lists this ratio for all multipliers examined in this study. As measured by L_n, RANDU performs considerably more poorly for $n = 3$, 4 and 5. Therefore, one possible explanation of the less than complete rejection of multipliers IX through XVI is that the flaws revealed by the lattice test are too subtle for detection. Since the rule of thumb for considering a multiplier acceptable with regard to n-tuples is $L_n < 2$, one may want to revise this bound upward in view of our results. A similar upward revision may apply to the acceptance criteria for the spectral test.

Table 2

P-values for Testing Hypotheses
(Multipliers IX through XIV)

Multiplier	A	Test	H_0	H_1	H_2	H_3	H_4 min	H_4 max
IX	1214170817	K-S	.020**	.071*	.514	.239	.149	.646
		V	.733	.370	.055*	.903	.303	.872
		A-D	.116	.082*	.344	.233	.206	.629
X	190576451	K-S	.668	.609	.069*	.913	.081*	.752
		V	.865	.608	.052*	.962	.045**	.511
		A-D	.533	.535	.082*	.773	.076*	.589
XI	1865416386	K-S	.845	.324	.175	.051*	.004***	.067*
		V	.805	.189	.272	.335	.006***	.482
		A-D	.927	.377	.099*	.087*	.010***	.198
XII	507435369	K-S	.670	.236	.894	.251	.874	.679
		V	.922	.481	.992	.585	.356	.964
		A-D	.812	.456	.989	.556	.852	.558
XIII	2139391393	K-S	.069*	.106	.957	.903	.001***	.001***
		V	.313	.156	.892	.543	.022**	.002***
		A-D	.247	.145	.985	.929	.001***	.000***
XIV	1277850838	K-S	.896	.264	.513	.851	.605	.540
		V	.694	.248	.721	.328	.373	.060*
		A-D	.977	.146	.564	.644	.670	.375
XV	163252367	K-S	.091*	.512	.238	.135	.886	.528
		V	.737	.446	.096*	.121	.686	.655
		A-D	.109	.331	.161	.205	.795	.643
XVI	55279518	K-S	.857	.151	.979	.597	.740	.844
		V	.644	.087*	.938	.265	.560	.949
		A-D	.914	.129	.995	.706	.929	.867

* .05 < P-value ≤ .10 . ** .01 < P-value ≤ .05 . *** P-value ≤ .01 .

Table 3
Lattice Test Performance[a]

Multiplier	L_2	L_3	L_4	L_5	L_6
RANDU	1.00	1819	1872	274	--
I	7.60	3.39	2.07	1.67	3.36
II	1.26	2.92	1.64	1.52	2.11
III	1.03	2.11	1.88	1.55	1.13
IV	1.07	1.34	1.17	1.62	1.89
V	2.82	2.63	1.50	1.20	1.86
VI	1.88	2.84	1.39	1.51	1.24
VII	1.64	2.36	1.71	1.24	1.24
VIII	1.69	1.62	1.88	1.38	1.41
IX	27.91	1.20	1.71	2.00	1.21
X	21.94	3.10	1.45	3.78	2.47
XI	18.1	1.17	3.14	1.41	2.37
XII	12.16	3.48	1.54	1.41	1.33
XIII	10.81	4.30	2.84	2.73	1.32
XIV	8.77	4.02	1.28	2.65	1.30
XV	8.33	1.19	2.32	2.10	2.09
XVI	7.05	2.53	1.95	3.18	1.59

[a]Source: For RANDU, (Marsaglia 1971). For Multipliers I through XVI, (Hoaglin 1976).

9. Execution Time

In Table 1 we listed the relative mean execution times observed for each multiplier (except RANDU). These samples were obtained by using the LLRANDOM package (Learmonth and Lewis, 1973) with a substitution of multipliers for 16807. The apparent increase in execution time as A increases in magnitude makes the desirability of large multipliers questionable. To put this issue in perspective, we examine an algorithm due to Payne et al. (1969) that is in common use for generating deviates from prime modulus generators.

Suppose one has storage areas A, X and Y each capable of holding p binary digits and a working area D capable of holding 2p binary digits. Payne et al. (1969) developed the following algorithm with A containing the multiplier and X containing the most recently generated random number.

LCG1

1. $D \leftarrow X \cdot A$.
2. $Y \leftarrow i.p.[D/2^{p-1}]$.
3. $X \leftarrow D - Y \cdot 2^{p-1}$.
4. $X \leftarrow X + Y$.
5. $Y \leftarrow i.p.[X/2^{p-1}]$.
6. If $Y \neq 0$, $X \leftarrow X - (2^{p-1}-1)$.
7. Return X .

Steps 2 and 3 may be performed using shift operations on D. Step 5 and the check in step 6 may be performed on p bit word machines by an overflow check.

It is instructive to examine how often the check in step 6 is satisfied. Of the possible $(2^{p-1}-2)$ different initial values of X, A^* of these cause

the check in step 6 to be satisfied where

$$A^* = \begin{cases} \text{i.p.}[(A-2)(A-1)/2A] & \text{if } A \le 2^{p-2} \\ \text{i.p.}\left[(A-2)(A-1)/2A + \dfrac{2^{p-1}+1-2A}{A}\right] & \text{if } A > 2^{p-2} \end{cases} .$$

Therefore, the average execution time for LCG1 tends to follow the relationship

$$\beta_0 + \beta_1 A^*/(2^{p-1}-2)$$

where β_0 is the time to perform steps 1 through 7 of LCG1 without the check in step 6 being satisfied and β_1 is the time to perform the operation in step 6 when the check is satisfied. For most values of A, A^* is very close to A/2 and the average execution time for LCG1 is $\beta_0 + \beta_1 A/2^p$.

In practice, execution time is code-dependent as well as algorithm-dependent. In particular, a code that produces a very small β_1 is most desirable. The code provided in Payne et al. (1969) and used with slight modification in SIMSCRIPT II should produce a very small β_1 . The code is self-contained. However, the code in LLRANDOM for step 6 will produce a relatively large value for β_1 . This is due to calls to a routine outside the generator code (when an overflow occurs). Regrettably, this observation did not come to our attention until all runs were completed. In summary, all implementations of the Payne et al. algorithm have execution times proportional to A . However, the choice of code determines the constant of proportionality.

10. References

1. Anderson, T. W. and D. A. Darling (1952). "Asymptotic Theory of Certain Goodness of Fit Criteria Based on Stochastic Processes," Ann. Math. Stat., Vol. 23, pp. 193-212.

2. Anderson, T. W. and D. A. Darling (1954). "A Test of Goodness of Fit," J.A.S.A., Vol. 49, pp. 765-769.

3. Coveyou, R. R. and R. D. MacPherson (1967). "Fourier Analysis of Uniform Random Number Generators," J. ACM, Vol. 14, pp. 100-119.

4. Dwass, M. (1958). "On Several Statistics Related to Empirical Distribution Functions," Ann. Math. Stat., Vol. 29, pp. 188-191.

5. Hoaglin, D. (1976). "Theoretical Properties of Congruential Random-Number Generators: An Empirical View," Memorandum NS-340, Department of Statistics, Harvard University.

6. IBM, SIMPL/1 Program Reference Manual (1972). SH19-5060-0.

7. IMSL Library (1977). International Mathematical and Statistical Libraries, Inc., Houston, Texas.

8. Katzan, H., Jr. (1971). APL User's Guide, Van Nostrand Reinhold, New York.

9. Knuth, D. L. (1969). The Art of Computer Programming, Vol. 2: Seminumerical Algorithms, Addison-Wesley, Reading, Mass.

10. Learmonth, G.P. (1975). "Empirical Tests of Multipliers for the Prime-Modulus Random Number Generator $X_{i+1} \equiv AX_i \mod 2^{31}-1$," Proceedings of the Ninth Interface Symposium on Computer Science and Statistics, D. C. Hoaglin and R. E. Welsch, eds.

11. Learmonth, G. and P. A. W. Lewis (1973). "Naval Postgraduate School Random Number Generator Package LLRANDOM," Monterey, California.

12. Levene, H. (1952). "On the Power Function of Tests of Randomness Based on Runs Up and Down," Ann. Math. Stat., Vol. 23, pp. 34-56.

13. Lewis, P. A. W., A. S. Goodman and J. M. Miller (1969). "A Pseudo-Random Number Generator for the System/360," IBM Systems J., Vol. 8, No. 2, pp. 136-145.

14. Marsaglia, G. (1968). "Random Numbers Fall Mainly in the Planes," Proc. Natl. Acad. Sci., Vol. 61, pp. 25-28.

15. Marsaglia, G. (1972). "The Structure of Linear Congruential Sequences," in Applications of Number Theory to Numerical Analysis, S. K. Zaremba, ed., Academic Press.

16. Payne, W. H., J. R. Rabung and T. P. Bogyo (1969). "Coding the Lehmer Pseudo-random Number Generator," Comm. ACM, Vol. 12, No. 2, pp. 85-86.

17. Tausworthe, R. C. (1965). "Random Number Generated by Linear Recurrence Modulo Two," Math. Comp., Vol. 19, pp. 201-209.

18. Tootill, J. P. R., W. D. Robinson and A. G. Adams (1971). "The Runs Up-and-Down Performance of Tausworthe Pseudo-Random Number Generators," J. ACM, Vol. 18, pp. 381-399.

19. Wolfowitz, J. (1944). "Asymptotic Distribution of Runs Up and Down," Ann. Math. Stat., Vol. 15, pp. 163-172.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: 78-11	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): A STATISTICAL EVALUATION OF MULTIPLICATIVE CONGRUENTIAL GENERATORS WITH MODULUS $2^{31}-1$		5. TYPE OF REPORT & PERIOD COVERED: Technical Report 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): George S. Fishman and Louis R. Moore		8. CONTRACT OR GRANT NUMBER(s): N00014-76-C-0302
9. PERFORMING ORGANIZATION NAME AND ADDRESS: University of North Carolina, Chapel Hill, N.C. 27514		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research, Arlington, Virginia 22217		12. REPORT DATE: December 1978 13. NUMBER OF PAGES: 19
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report): Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report)		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Anderson-Darling Test
Chi-Squared Test
Kolmogorov-Smirnov Test
Linear Congruential Generator
Random Number Generation
Runs Test
Serial Test
Uniformity

(2 to the 31st power) -1

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This paper presents the results of empirically testing eight alternative multipliers for a multiplicative congruential generator with modulus $2^{31}-1$. The LLRANDOM number package (Learmonth and Lewis 1973) uses one of the multipliers, the simulating programming language SIMSCRIPT II uses a second and the remaining six are the best of 50 candidate multipliers studied by Hoaglin (1976) using the theoretical spectral and lattice tests. The battery of tests raises serious doubts about three of the multipliers, including the one in LLRANDOM.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601 |

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

The power of the tests is demonstrated by their rejection of RANDU, a notably poor random number generator. A comparison of the results for the eight multipliers with the eight worst multipliers (with regard to 2-tuples) in Hoaglin (1976) failed to show any apparent gross differences. Since examination of performance on the lattice test revealed that the 16 multipliers clustered in performance when compared to RANDU, one may conjecture that the poorer lattice test performance for the worst eight is too subtle for detection by our empirical tests. Since this failure may not be as serious as the lattice test implies, one may want to revise upward the criteria for acceptable performance on the lattice tests and similarly on the spectral test. In section 9 an analysis of algorithms for generating random numbers reveals that the choice of coding can significantly affect execution time for different multipliers.