JOURNAL OF COMPUTATIONAL PHYSICS 67, 482–486 (1986)

Note

The Use of Cebysev Mixing to Generate Pseudo-random Numbers

1. Introduction

In their paper on Cebysev mixing [1], Erber, Everett, and Johnson suggest the use of Cebysev mixing to generate pseudo-random numbers, stating:

> In fact a central result of this work is that the numbers $(4/\pi)\cos^{-1}(\frac{1}{2}Z_{n+1})-2$, are to a very good approximation uniformly and "randomly" distributed between -2 and $+2$ when Z_{n+1} is calculated in double precision from the simple recurrence $Z_{n+1}=Z_n^2-2$.

This paper presents the results of applying several statistical tests for randomness to the proposed pseudo-random number generator.

2. Pseudo-random Numbers

There is no general agreement as to a definition of a random sequence. Definitions have ranged from the requirement that a sequence pass some statistical tests, the choice of tests depending on the use to which the sequence is to be put, to the requirement that a sequence pass all statistical tests. Knuth [2] gives a sequence of definitions intermediate between the above two views. Knuth's definitions are extensions of the criterion that the sequence should be equidistributed. A sequence $\{U_n\}$ in the interval $[0, 1)$ is *equidistributed* if

$$\Pr(a \leqslant U_n < b) = b - a$$

for $0 \leqslant a < b \leqslant 1$. Here $\Pr(S(n))$ is the (limiting) value of the proportion of the time that the statement $S(n)$ is true. If $h(n)$ is the number of values between 1 and n for which $S(n)$ holds, then

$$\Pr(S(n)) = \lim_{n\to\infty} \frac{h(n)}{n}.$$

A sequence is *k-distributed* if

$$\Pr(a_1 \leqslant U_n < b_1, \ldots, a_k \leqslant U_{n+k-1} < b_k) = (b_1 - a_1) \cdots (b_k - a_k)$$

0021-9991/86 $3.00

for any $0 \leqslant a_i < b_i \leqslant 1$, $i = 1,..., k$. For example, $\{U_n\}$ is 2-distributed if the pairs

$$(U_1, U_2), (U_2, U_3), (U_3, U_4),...$$

are equidistributed over the unit square.

A sequence is ∞-*distributed* if it is k-distributed for all positive integers k. Knuth's weakest definition of a random sequence requires that the sequence be ∞-distributed [2, p. 152]. If a sequence is ∞-distributed then it satisfies a variety of empirical tests of randomness, including those which are considered in the next section.

3. Empirical Tests

The proposed random number generator was subjected to ten empirical tests of randomness [2, Sect. 3.3.2]:

(1) the coupon collector's test,
(2) the distribution of the mean and variance,
(3) the frequency distribution test,
(4) the Kolmogorov–Smirnov test,
(5) the gap test,
(6) the maximum of t test,
(7) the poker test,
(8) the serial correlation test,
(9) the serial test for successive pairs, and
(10) the runs test.

A sequence of 10,000 Cebysev mixing values,

$$V_n = (4/\pi) \cos^{-1}(\tfrac{1}{2}Z_{n+1}) - 2,$$

where

$$Z_{n+1} = Z_n^2 - 2,$$

were generated from the seed $Z_0 = \pi - 3$ used in [1]. These values were transformed to the unit interval by

$$U_n = \tfrac{1}{4}V_n + 0.5.$$

The tests were performed using the RANDOM package for evaluating pseudo-random number generators [3]. The tests were all at the 0.05 level of significance.

The proposed generator passed tests (2), (3), (4), (6), and (8), but failed the other tests, showing strong non-random patterns. The tests were also performed for other seeds used in the Erber paper, with similar results.

4. The Results of the Tests

The Serial Test for Successive Pairs. Because of the continuous nature of the generating functions, correlated pairs can be expected. For example, suppose the Cebysev mixing sequence is $\{V_n\}$,

$$V_n = (4/\pi)\cos^{-1}(\tfrac{1}{2}Z_n) - 2.$$

If V_{n+k} is close to V_n then V_{n+k+1} will be close to V_{n+1}. Thus the pairs (V_n, V_{n+1}) will not be independently distributed and the Cebysev mixing sequence should fail the serial test for consecutive pairs [2, p. 60].

A sequence of initial digits Y_n was obtained by

$$Y_n = \text{trunc}(U_n * 10), \qquad \text{so that } 0 \leqslant Y_n \leqslant 9.$$

The sequence $\{Y_n\}$ was partitioned into 5000 pairs (Y_{2i-1}, Y_{2i}), for $1 \leqslant i \leqslant 5000$, and the frequency of each pair of initial digits was recorded. If the pairs were uniformly and independently distributed, then the expected frequency for each pair would be 50. The results obtained are shown in Table I.

The computed chi-square test statistic is 20,111.7227, and the critical value of chi-square test statistic (alpha of 0.05) is 123.2253 with 99 degrees of freedom.

The Gap Test. This test examines the lengths of "gaps" between occurrences of U_n. Let $a = 0.30$ and $b = 0.60$. A gap of length k is a consecutive subsequence

TABLE I

Frequency Table for Observed Pairs of Initial Digits

First digit	Second digit 0	1	2	3	4	5	6	7	8	9
0	209	251	0	0	0	0	0	0	0	0
1	0	0	227	244	0	0	0	0	0	0
2	0	0	0	0	262	265	0	0	0	0
3	0	0	0	0	0	0	247	260	0	0
4	0	0	0	0	0	0	0	0	273	217
5	0	0	0	0	0	0	0	0	253	243
6	0	0	0	0	0	0	255	262	0	0
7	0	0	0	0	258	244	0	0	0	0
8	0	0	237	273	0	0	0	0	0	0
9	263	257	0	0	0	0	0	0	0	0

TABLE II

Frequency Table of Gaps

Gap length	Observed frequency	Expected frequency
0	0	911.7000
1	1041	638.1900
2	1014	446.7330
3	489	312.7131
4	256	218.8992
5	114	153.2294
6	81	107.2606
7	24	75.0824
8	20	175.1923

Note. Last category includes gaps of length 8 or more.

$U_n,..., U_{n+k}$ in which U_{n+k} is between a and b, but he other U's are not. The results are shown in Table II.

The number of observed gaps in the sequence of 10,000 pseudo-uniform [0, 1] random numbers is 3039, the computed chi-square test statistic is 2180.6360, and the critical value of chi-square test statistic (alpha of 0.05) is 15.5073 with 8 degrees of freedom.

The Permutation Test. The sequence $\{U_n\}$ was divided into 3333 triples. The six possible orderings of the smallest value, A, the middle value, B, and the largest value, C, were tabulated in Table III.

The computed chi-square test statistic is 1552.5356 and the critical value of chi-square test statistic (alpha of 0.05) is 11.0705 with 5 degrees of freedom.

The Runs Test. A run up of length k occurs when

$$U_n > U_{n+1} < U_{n+2} < \cdots < U_{n+k} > U_{n+k+1}.$$

The number of runs up and runs down are tabulated in Table IV.

The critical value of the chi-square test statistic (alpha of 0.05) is 12.5916 with 6

TABLE III

Frequency Table of Permutations

Permutation type	Observed frequency	Expected frequency
(C, A, B)	633	555.5000
(B, C, A)	878	555.5000
(B, A, C)	475	555.5000
(C, B, A)	0	555.5000
(A, C, B)	221	555.5000
(A, B, C)	1126	555.5000

581/67/2-17

TABLE IV

Table of Runs

Run length	Expected number	Observed number of runs up	Observed number of runs down
1	1667.33	1	3256
2	2083.38	1710	3372
3	916.55	846	0
4	263.82	401	0
5	57.52	216	0
6	11.90	199	0

Note. Last category includes runs of length 6 or more.

degrees of freedom. The computed chi-square test statistic for runs up is 12,923.6875. The computed chi-square test statistic for runs down is 3749.3918.

5. Conclusion

The proposed random number generator derived from Cebysev mixing has the advantage that it is amenable to theoretical analysis. However, because of its strong correlations, it should be used with caution.

References

1. T. Erber, P. Everett, and P. W. Johnson, *J. Comput. Phys.* **32**, 168 (1979).
2. D. Knuth, *The Art of Computer Programming*, Vol. 2, *Seminumerical Algorithms* (Addison–Wesley, Reading, Mass., 1981).
3. J. R. Crigler, *et al.*, "Random: A Computer Program for Evaluating Pseudo-uniform Random Number Generators," National Technical Information Service Report Tr 82-93, August 1982.

Received: August 13, 1985; revised: February 7, 1986

John M. Hosack
Department of Mathematics
Colby College
Waterville, Maine 04901