ON IMPROVING PSEUDO-RANDOM NUMBER GENERATORS

Lih-Yuan Deng
E. Olusegun George

Department of Mathematical Sciences
Memphis State University
Memphis, TN 38152

Yu-Chao Chu

Department of Preventive Medicine
University of Tennessee – Memphis
Memphis, TN 38163

ABSTRACT

Some theoretical and empirical justifications for the combination generators are given. It is shown that adding enough random variates, whether or not they are independent, the fractional part of their sum will converge to a uniform distribution. Empirical study shows that combination generators can even transform some "bad" random generators into a much better one.

1 INTRODUCTION

The ideal goal in generating random numbers is to find an algorithm that will generate truly random numbers. It is well-known, however, that truly random and independent variates cannot be computer-generated using any algorithms. In fact, as observed by Park and Miller (1988), good uniform random number generators are hard to find and some of the popular generators display distinctly non-uniform characteristics. Unfortunately, most of the standard algorithms seem to have been proposed under the false assumption that they could produce truly uniform random variates, when in fact some of them generate pseudo-random numbers which are significantly non-random. Deng (1988) and Deng and Chhitkara (1991) showed that inaccuracies in generated random numbers are invariably carried over and sometimes magnified when these numbers are transformed to produce variates of interest.

Several improvements over the traditional congruential method have been proposed in the literature. Knuth (1981), Wichmann and Hill (1982), Marsaglia (1985), L'Ecuyer (1988), Collins (1987), and Anderson (1990) all suggested the use of the combination generator. Based on his empirical study on several popular generators, Marsaglia (1985) concluded that combination generators seemed to be the best generator. Some justifications are available for the combination generator, but all are based on some unrealistic assumptions. Horton (1948) and Horton and Smith (1949) showed that the sum of several integer pseudo numbers modulo a positive integer $M$ converges to a discrete uniform distribution over $0, 1, 2, ..., M - 1$. Deng and George (1990) showed that the sums, modulo 1, of nearly uniform continuous random variables were much more uniform than the individual variables. Brown and Solomon (1979), Marsaglia (1985) and L'Ecuyer (1988) also provided some theoretical support for combination generators under the unrealistic assumption that individual generators were independent of each other.

One of our major results in this paper is to remove the independence assumption of the generated sequence. In section 2, conditions are given for the convergence of sum, modulo 1, of several possibly dependent random variables to a $U(0, 1)$ variate. In section 3, we extend this result to a multidimensional case. We show that if each component of a sequence of continuous random vectors is "stretched out", then the fractional part of the components will converge to independent uniform random variables. This theorem provides a simple method for generating a sequence of asymptotically independent $U(0, 1)$ random variables. The result of an empirical study is presented in section 4. Simulation results show that the fractional part of a sum of dependent uniform random variables or non-uniform variates is quite close to $U(0, 1)$, even for a sample size as small as 4.

2 ASYMPTOTIC UNIFORMITY

Deng and George (1990) proves that the fractional part of a sum of two independent "nearly" uniform random variables produces a "nearly" uniform random variable whose distribution is closer to a $U(0, 1)$ than the parent distribution. Specifically, they proves the following theorem:

Theorem. Let $X_1, X_2, ..., X_n$ be $n$ independent random variables distributed over $[0, 1]$, with p.d.f. $f_k(x_k), k = 1, 2, ..., n$. Let $U_n = \sum_{k=1}^{n} X_k \mod 1$, and $f_{U_n}(u)$ be the p.d.f. of $U_n$. If $|f_k(x_k) - 1| \leq \epsilon_k, k = \ldots$
1, 2, ..., n, then
\[ |f_{U_n}(u) - 1| \leq \prod_{k=1}^{n} \epsilon_k \]
and
\[ U_n \overset{d}{\rightarrow} U(0, 1), \quad \text{as} \quad \prod_{k=1}^{n} \epsilon_k \rightarrow 0. \]

The implication of this result is that one can generate a more uniform variate by taking the fractional sum of several "nearly" uniformly distributed variates, and as will be seen in the empirical study reported in section 4, that the number of terms needed to achieve uniformity may be very small. However, as we noted before, no computer-generated sequences can be safely assumed to be independent of each other. We will show next that the asymptotic uniformity result can be proved without the independence assumption.

The following lemma gives a simple relationship between the p.d.f. of the fractional part of a random variable and the p.d.f. of the original random variable.

**Lemma 2.1.** Let \( Y \) be any continuous random variable with p.d.f. \( f_Y(y) \). Then the p.d.f. of \( U = Y \mod 1 \) is given as
\[ f_U(u) = \sum_m f_Y(u + m), \quad 0 < u < 1. \]

**Proof.** Denote the c.d.f. of \( U \) and \( Y \) as \( F_U(\cdot) \) and \( F_Y(\cdot) \), respectively. By the definition of \( U \), we have
\[ F_U(u) = \sum_m \Pr(m \leq Y \leq u + m) = \sum_m (F_Y(u + m) - F_Y(m)), \quad 0 < u < 1. \]

Lemma 2.1 follows easily by taking derivative w.r.t. \( u \) on both sides. □

The following lemma states that if a random sequence converges to a \( U(0,1) \) distribution, then adding any constant sequence \( a_n \) and then taking the fractional part will again converge to a \( U(0,1) \) distribution.

**Lemma 2.2.** Let \( \{U_n, n = 1, 2, \ldots\} \) be a sequence of continuous random variables such that the c.d.f. of \( U_n, F_{U_n}(t) \) converges uniformly to the c.d.f. of \( U(0,1) \). Then for any sequence of constants \( \{a_n, n = 1, 2, \ldots\} \), we have
\[ Y_n = (U_n + a_n) \mod 1 \overset{d}{\rightarrow} U \sim U(0,1). \]

**Proof.** Since
\[ Y_n = [(U_n \mod 1) + (a_n \mod 1)] \mod 1, \]
without loss of generality, we may assume
\[ 0 \leq a_n < 1 \quad \text{and} \quad 0 \leq U_n < 1. \]

From the definition of \( Y_n \), we have, for \( 0 \leq u < 1, \)
\[ \Pr(Y_n \leq u) = \Pr(0 \leq U_n + a_n \leq u) + \Pr(1 \leq U_n + a_n \leq 1 + u) \]
\[ = \left\{ \begin{array}{ll}
F_{U_n}(1 + u - a_n) - F_{U_n}(1 - a_n), & \text{if } u < a_n \\
F_{U_n}(u - a_n) + [1 - F_{U_n}(1 - a_n)], & \text{if } u > a_n,
\end{array} \right. \]

where \( F_{U_n}(\cdot) \) is the c.d.f. of \( U_n \). Since \( U_n \overset{d}{\rightarrow} U(0,1) \) and \( U_n \) is continuous random variable, we have \( F_{U_n}(t) = t + o(1) \), for all \( t \in (0, 1) \). This implies that
\[ \Pr(Y_n \leq u) = u + o(1) \rightarrow u \quad \text{as } n \rightarrow \infty. \]

Next, we will prove one of our key results.

**Theorem 2.1.** Let \( \{Z_n, n = 1, 2, \ldots\} \) be a sequence of continuous random variables with p.d.f. \( f_{Z_n}(z) \). Suppose that the p.d.f. of \( Z_n, f_{Z_n}(z) \) converges uniformly to \( f_Z(z) \), where \( f_Z(z) \) is the p.d.f. of a continuous random variable \( Z \). Then for any sequence of constants \( \{a_n, b_n, n = 1, 2, \ldots\} \) such that \( \lim_{n \rightarrow \infty} b_n = \infty \),
\[ U_n = (b_n Z_n + a_n) \mod 1 \overset{d}{\rightarrow} U(0,1) \quad \text{as } n \rightarrow \infty. \]

**Proof.** From Lemma 2.2, Theorem 2.1 follows immediately if we can show that
\[ b_n Z_n \mod 1 \overset{d}{\rightarrow} U(0,1) \quad \text{as } n \rightarrow \infty. \]

Without loss of generality, we may assume \( a_n = 0 \). From Lemma 2.1, the p.d.f. of \( U_n \) is given as
\[ f_{U_n}(u) = \sum_m f_{Z_n}(u + m) \]
\[ = \sum_m f_{Z_n}(u + m) \frac{1}{b_n} = \sum_m f_Z(u + m) \frac{1}{b_n} + o(1) \]
\[ = \int f_Z(z)dz + o(1) \]
\[ = 1 + o(1) \rightarrow 1 \text{ as } n \rightarrow \infty. \]

Hence by Scheffé's theorem, \( b_n Z_n + a_n \mod 1 \overset{d}{\rightarrow} U(0,1). \) □
Corollary 2.1. Let \( \{Y_n, n = 1, 2, \ldots\} \) be a sequence of continuous random variables with p.d.f. \( f_{Y_n}(y) \). Suppose there exists a sequence of constants \( \{a_n, b_n\}, n = 1, 2, \ldots \) such that \( \lim_{n \to \infty} b_n = \infty \) and the p.d.f. of \( Z_n = (Y_n - a_n)/b_n \) converges uniformly to a p.d.f. \( f_Z(z) \). Then

\[
U_n = Y_n \mod 1 \xrightarrow{d} U(0, 1), \quad \text{as } n \to \infty.
\]

Proof. From \( Z_n = (Y_n - a_n)/b_n \), we have \( U_n = b_n Z_n + a_n \mod 1 \). Corollary 2.1 follows directly from Theorem 2.1. \( \Box \)

Corollary 2.2. Let \( Z \) be any continuous random variable with p.d.f. \( f_Z(z) \). Suppose there exists a sequence of constants \( \{a_n, b_n\}, n = 1, 2, \ldots \) such that \( \lim_{n \to \infty} b_n = \infty \). Then

\[
U_n = (b_n Z + a_n) \mod 1 \xrightarrow{d} U(0, 1), \quad \text{as } n \to \infty.
\]

A simple application of Corollary 2.1 is to take \( Y_n = \sum_{k=1}^{n} X_k \). If one can find "normalizing" constants of \( Y_n \) as described in Corollary 2.1, then \( Y_n \mod 1 = \sum_{k=1}^{n} X_k \mod 1 \xrightarrow{d} U(0, 1) \). No independence assumption among \( X_k \) is required. In fact, even if all \( X_k \) are identical, i.e. \( X_k = X \), we have \( \sum_{k=1}^{n} X_k \mod 1 = nX \mod 1 \xrightarrow{d} U(0, 1) \), according to Corollary 2.2.

There is a very simple explanation for the above theorem and its corollaries. If a continuous random variable is "stretched far out" (either by scaling or adding several random variables), then the stretched variable, such as \( Y_n \) in Corollary 2.1 and \( b_n Z_n + a_n \) in Theorem 2.1, should be (roughly speaking) uniformly distributed, within each subinterval \( [m, m + 1) \) for each integer \( m \). Therefore \( U_n = Y_n \mod 1 \) and \( U_n = (b_n Z_n + a_n) \mod 1 \) should converge to \( U(0, 1) \).

An interesting and somewhat surprising interpretation of Corollary 2.2 is that the lower significant digits of any continuous random variable tends to be uniformly distributed. However, this observation is not directly applicable to pseudo-random generators because no generator is capable of generating a truly continuous variate.

3 ASYMPTOTIC INDEPENDENCE

In this section, we will propose a method to generate a sequence of uniform random variables which will be asymptotically uniform \( U(0, 1) \) distributed and asymptotically independent of each other.

Lemma 3.1. Let \( (Y_1, Y_2) \) be any continuous random vector with the joint p.d.f. \( f_{Y_1, Y_2}(y_1, y_2) \). Then the joint p.d.f. of \( (U_1, U_2) = (Y_1 \mod 1, Y_2 \mod 1) \) is given as

\[
f_{U_1, U_2}(u_1, u_2) = \sum_{m} \sum_{l} f_{Y_1, Y_2}(u_1 + m, u_2 + l),
\]

for \( 0 < u_1, u_2 < 1 \).

Proof. The proof is similar to Lemma 2.1. \( \Box \)

Theorem 3.1. Let \( \{(Z_{1n}, Z_{2n}), n = 1, 2, \ldots\} \) be a sequence of continuous random vectors with joint p.d.f. \( f_{Z_{1n}, Z_{2n}}(z_1, z_2) \). Suppose that \( f_{Z_{1n}, Z_{2n}}(z_1, z_2) \) converges uniformly to \( f_{Z_1, Z_2}(z_1, z_2) \), as \( n \to \infty \), where \( f_{Z_1, Z_2}(z_1, z_2) \) is the joint p.d.f. of a continuous random vector \( (Z_1, Z_2) \). Consider two sequences of constant pairs \( \{(a_{in}, b_{in}), n = 1, 2, \ldots\} \), \( i = 1, 2 \), such that \( \lim_{n \to \infty} b_{in} = \infty \). Let

\[
U_{in} = (b_{in} Z_{in} + a_{in}) \mod 1 \quad i = 1, 2.
\]

Then

1. \( U_{in} \xrightarrow{d} U(0, 1), \quad \text{as } n \to \infty, \quad \text{for } i = 1, 2, \) and
2. \( U_{1n} \) and \( U_{2n} \) are asymptotically independent.

Proof. Part (1) is proved in the Theorem 2.1. To prove Part (2), we use Lemma 3.1. Without loss of generality, we may assume \( a_{1n} = 0 \) and \( a_{2n} = 0 \). The joint c.d.f., \( f_{U_{1n}, U_{2n}}(u_1, u_2) \), of \( (U_{1n}, U_{2n}) \) is

\[
\sum_{m} \sum_{l} f_{Z_{1n}, Z_{2n}}(u_1 + m, u_2 + l) = \sum_{m} \sum_{l} f_{Z_{1n}, Z_{2n}} \left( \frac{u_1 + m}{b_{1n}}, \frac{u_2 + l}{b_{2n}} \right) \frac{1}{b_{1n} b_{2n}} + o(1)
\]

\[
= \int \int f_{Z_1, Z_2}(z_1, z_2) dz_1 dz_2 + o(1)
\]

\[
= 1 + o(1) \xrightarrow{} 1 \quad \text{as } n \to \infty.
\]

Hence, \( U_{1n} \) and \( U_{2n} \) are asymptotically independent. \( \Box \)

Corollary 3.1. Let \( \{(Y_{1n}, Y_{2n}), n = 1, 2, \ldots\} \) be a sequence of continuous random vectors with p.d.f. \( f_{Y_{1n}, Y_{2n}}(y_1, y_2) \). Suppose there exists a sequence of constants \( \{(a_{in}, b_{in}), n = 1, 2, \ldots\} \), \( i = 1, 2 \), such that \( \lim_{n \to \infty} b_{in} = \infty \). Let \( Z_{in} = (Y_{in} - a_{in})/b_{in} \), \( i = 1, 2 \). Suppose the joint p.d.f. of \( (Z_{1n}, Z_{2n}) \) converges uniformly to a joint p.d.f. \( f_{Z_1, Z_2}(z_1, z_2) \). Then

1. \( U_{in} = Y_{in} \mod 1 \xrightarrow{d} U(0, 1), \quad \text{as } n \to \infty, \quad \text{for } i = 1, 2, \) and
2. $U_{1n}$ and $U_{2n}$ are asymptotically independent.

Proof. From $Z_{in} = (Y_{in} - a_{in})/b_{in}$, we have $U_{in} = b_{in}Z_{in} + a_{in} \mod 1$. Corollary 3.1 follows directly from the above theorem.

Corollary 3.2. Let $(Z_1, Z_2)$ be any continuous random vector with joint p.d.f. $f_{Z_1, Z_2}(z_1, z_2)$. Suppose there exists a sequence of constants $\{(a_{in}, b_{in}), n = 1, 2, \ldots\}$, $i = 1, 2$, such that $\lim_{n \to \infty} b_{in} = \infty$. Let $U_{in}$ be

$$(b_{in}Z_{i} + a_{in}) \mod 1 \text{ for } i = 1, 2.$$ Then

1. $U_{in} \overset{d}{\to} U(0, 1)$, as $n \to \infty$, for $i = 1, 2$, and
2. $U_{1n}$ and $U_{2n}$ are asymptotically independent.

Remarks:

1. It is straightforward to generalize the above results (Theorem 3.1 and Corollary 3.1, 3.2) from two joint random variables to more than two random variables. The precise statements for higher dimensions will be therefore omitted.

2. Similar to section 2, one can give an intuitive explanation for the results proved in this section. If a continuous random vector is "stretched far out" (either by scaling or adding several random vectors) in all directions, then "locally" (within each dimension of the square of the partition) the stretched vector should be (roughly speaking) uniformly distributed in each coordinate and the components should be independent of each other.

Let $\{X_{ik}, i = 1, 2, \ldots\}$, $k = 1, 2, \ldots, n$ be $n$ sequences of random variables representing $n$ separate random number generators. Define

$$Y_{in} = \sum_{k=1}^{n} X_{ik}, \quad i = 1, 2, \ldots.$$ Let

$$U_{in} = Y_{in} \mod 1.$$ Under a very weak condition, it follows from Corollary 3.1 that each variate in the new sequence $\{U_{in}, i = 1, 2, \ldots\}$ will follow approximately a uniform distribution, $U(0, 1)$. Furthermore, any two variates in the sequence $U_{1n}$ and $U_{2n}$ are asymptotically independent, for any $i_1 \neq i_2$. The condition required is the existence of "normalizing" constants for $Y_{1n}$ and $Y_{2n}$ as described in Corollary 3.1. Note that this condition is certainly much weaker than the usual requirements on normalizing constants needed from central limit theorems. No independence or finite variance assumption is required. The removal of independence assumption of random sequence is of practical importance because no computer-generated sequence satisfies the independence assumption. As we have noted before, the previous discussion for random vector $(U_{1n}, U_{2n})$ can be easily extended to any higher dimensions. Hence, we have produced a sequence of random variates which is approximately i.i.d. $U(0,1)$ distributed.

To make the previous discussion clearer, let us consider the following diagram representing $n$ random variate generator:

| 1st RNG | $X_{11}$ | $X_{12}$ | $X_{13}$ | $X_{1n}$ | $X_{1n}$ | $X_{1n}$ | $X_{1n}$ |
| 2nd RNG | $X_{12}$ | $X_{13}$ | $X_{14}$ | $X_{1n}$ | $X_{1n}$ | $X_{1n}$ | $X_{1n}$ |
| nth RNG | $X_{1n}$ | $X_{1n}$ | $X_{1n}$ | $X_{1n}$ | $X_{1n}$ | $X_{1n}$ | $X_{1n}$ |

$\sum_{k=1}^{n} X_{ik} \mod 1 \quad U_{1n} \quad U_{1n} \quad U_{1n}$

For simplicity, take $i_1 = 1$ and $i_2 = 2$. According to Corollary 3.1, $U_{1n} \overset{d}{\to} U(0, 1)$ and $U_{2n} \overset{d}{\to} U(0, 1)$, as $n \to \infty$. Furthermore, $U_{1n}$ and $U_{2n}$ will be asymptotically independent as $n$ gets larger. Note that neither "between generators" nor "within generator" independence is assumed. The only required condition is the existence of normalizing constants for $(Y_{1n}, Y_{2n})$ as described in Corollary 3.1. Using similar argument as above and the higher dimension extension of Theorem 3.1, we can see that the random sequence $U_{1n}, U_{2n}, U_{3n}, \ldots$ will be approximately i.i.d. $U(0, 1)$, for large $n$.

Wichmann and Hill (1982) proposes a portable random number generator by combining three multiplicative congruential generators $\{X_i, i = 1, 2, \ldots\}, \{Y_i, i = 1, 2, \ldots\}$ and $\{Z_i, i = 1, 2, \ldots\}$ as follows

$$U_i = X_i + Y_i + Z_i \mod 1.$$ This generator has a much longer period of generating cycle than a single multiplicative method. However, Wichmann and Hill (1982) have not given any theoretical justification. The previous discussion (with $n = 3$) provides theoretical justifications for the uniformity as well as independence of the $U_i$'s. The theorems in this section also suggest the following modification:

$$U_i = b_1 X_i + b_2 Y_i + b_3 Z_i \mod 1,$$

where $b_1, b_2, b_3$ are some large numbers. The above generator will give good results even if (1) $\{X_i, i = 1, 2, \ldots\}, \{Y_i, i = 1, 2, \ldots\}$ and $\{Z_i, i = 1, 2, \ldots\}$ were
“bad” generators and/or (2) three generators are correlated. In practice, however, one should not choose b’s so large that we may lose some significant digits in a computer multiplication. The empirical study in section 4 also shows that b’s does not have to be large.

If we are given a random number generator which is not generating uniform and random sequence, then the following simple procedure may be used to generate nearly uniform pseudo random variables that are almost independent. Let \( \{X_i, i = 1, 2, \ldots\} \) be a random sequence generated from a generator. Define

\[
U_{in} = \sum_{j=1}^{n} X_{(i-1)n+j} \mod 1.
\]

Using a generalization of Corollary 3.1, and assuming the weak condition holds, we can now show that the random sequence \( \{U_{in}, i = 1, 2, \ldots\} \) is asymptotically independent and identically \( U(0, 1) \) distributed. In section 4, we will demonstrate through extensive empirical study that the proposed procedure will indeed transform a very bad generator into a much better one.

**4 EMPIRICAL STUDY**

In this section, we will present the result of an empirical study that demonstrate that the asymptotic results of the previous sections hold well even for a very small sample size (e.g. \( n = 4 \)). Suppose that \( U_j \) is generated from a “better” pseudo-random number generator. We use the pseudo-random number generator provided by the IMSL routine. We then “distort” \( X_i \) either as a convex combination of \( U_j \)

\[
X_i = \sum_{j=0}^{k} c_j \cdot U_{i+j}, \quad \sum_{j=0}^{k} c_j = 1, 0 < c_j < 1
\]

or generate a non-uniform variate, say

\[
X_i \sim \text{Beta}(\alpha, \beta).
\]

It is obvious that \( \{X_i, i = 1, 2, \ldots\} \) is a poor uniform random number generator. It is either correlated or non-uniformly distributed. Using the results in previous sections, we will transform \( \{X_i, i = 1, 2, \ldots\} \) into a “good” uniform random generator. We will show through our empirical study that either

\[
Y_i = \sum_{j=1}^{n} X_{n(i-1)+j} \mod 1
\]

or

\[
Z_i = b_1 X_{2i} + b_2 X_{2i-1} \mod 1
\]

will yield a much better generator. The empirical study also shows that our asymptotic theory works pretty well even for small values of \( n, b_1 \) and \( b_2 \). In this study, we choose \( n = 4 \), \( b_1 = 3 \) and \( b_2 = 5 \).

The empirical study procedure is as follows:

1. Generate \( U_1, U_2, \ldots \) from the IMSL routine RNUNF.
2. Define
   
   (a) \( X_i = (\sum_{j=0}^{k} c_j \cdot U_{i+j}) \), where \( 0 < c_j < 1 \), for \( 0 \leq j \leq k \), and \( \sum_{j=0}^{k} c_j = 1 \) [Note that if \( c_0 = 1 \), then \( X_i = U_i \)]
   
   (b) \( X_i \sim \text{Beta}(\alpha, \beta) \). [Note that if \( (\alpha, \beta) = (1, 1) \), then \( X_i = U_i \)]
3. Let \( Y_i = X_{4i} + X_{4i-1} + X_{4i-2} + X_{4i-3} \mod 1 \)
4. Let \( Z_i = 3X_{2i} + 5X_{2i-1} \mod 1 \)
5. For each random sample \( X, Y, Z \), perform the following tests for randomness from IMSL routines:
   
   (a) goodness-of-fit test,
   
   (b) Good’s serial pair test, with lag=1,
   
   (c) triplets test,
   
   (d) \( d^2 \) test(Grunenberger and Mark (1951)).
6. Chi-square statistics of each test for \( X, Y \) and \( Z \) are recorded.
7. Repeat steps (1) to (6) 10,000 times, and calculate the percentage of Chi-square statistics (for \( X, Y \) and \( Z \)) greater than the tabulated percentile of \( \chi^2 \) values, with appropriate degrees of freedom and \( \alpha = 0.10, 0.05, 0.01 \).

The following is a summary of how these tables are obtained:

1. Four empirical tests using IMSL routines are performed:

<table>
<thead>
<tr>
<th>Test</th>
<th>( \chi^2 )</th>
<th>pair</th>
<th>triplet</th>
<th>( d^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>routine</td>
<td>chigf</td>
<td>pairs</td>
<td>dcube</td>
<td>dsgar</td>
</tr>
<tr>
<td>1,000</td>
<td>2,000</td>
<td>3,000</td>
<td>2,000</td>
<td></td>
</tr>
</tbody>
</table>

2. Each entry in Tables A-1 to A-4 represents the percentage of the observed \( \chi^2 \) larger than its tabulated \( \chi^2 \) values in 10,000 samples chosen. The “distorted” generator is \( X_i = \sum_{j=0}^{4} c_j U_{i+j} \) and four different sets of \( c_j \) chosen as follows:

<table>
<thead>
<tr>
<th>Experiment</th>
<th>( (c_0, c_1, c_2, c_3, c_4) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>(0.2,0.2,0.2,0.2,0.2)</td>
</tr>
<tr>
<td>(2)</td>
<td>(0.1,0.1,0.1,0.1,0.1)</td>
</tr>
<tr>
<td>(3)</td>
<td>(0.3,0.3,0.1,0.1,0.1)</td>
</tr>
<tr>
<td>(4)</td>
<td>(0.4,0.2,0.2,0.2,0.1)</td>
</tr>
</tbody>
</table>
3. Each entry in Tables B-1 to B-4 represents the percentage of the observed $\chi^2$ larger than its tabled $\chi^2$ values in 10,000 samples chosen. The “distorted” generator is $X_i \sim \text{Beta}(\alpha, \beta)$ and four different sets of $\alpha, \beta$ chosen as follows:

<table>
<thead>
<tr>
<th>Experiment</th>
<th>$(\alpha, \beta)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>(0.6, 0.6)</td>
</tr>
<tr>
<td>(2)</td>
<td>(2.0, 1.0)</td>
</tr>
<tr>
<td>(3)</td>
<td>(1.0, 2.0)</td>
</tr>
<tr>
<td>(4)</td>
<td>(0.8, 1.2)</td>
</tr>
</tbody>
</table>

From the above empirical study, we can see that the random sequence $X_i$ (either $\sum_{j=0}^{k} c_j U_{i+j}$ or $\text{Beta}(\alpha, \beta)$) is far from being i.i.d. $U(0,1)$ distributed. The percentage of the computed chi-square statistics greater than the chi-square table value is 100 percent in every case generated. However, our proposed transformation (both $Y$ or $Z$) will yield random sequence whose distribution is very close to i.i.d. $U(0,1)$ distribution. The percentage of the computed chi-square statistics greater than the chi-square table value is very close to its nominal value.

5 CONCLUDING REMARKS

We have provided some theoretical justifications for the asymptotic uniformity and asymptotic independence of the combination generators without the usual assumption of independence of the generators. Theorems in this paper also give us a general method of transforming a possibly bad generator into a much better generator which will yield an asymptotically independent and uniformly distributed random sequence. Our empirical study demonstrates that only a small number of terms is required in our asymptotic theory to achieve a much more uniform random sequence. As pointed out in Park and Miller (1988), some generators provided by certain computer system may not be very “random”. Combining several generators into a single generator will provide some protection against the possibility of “bad” system-provided generators. Since uniform variate generation is the key to generating other commonly used probability distributions, these results should be useful in many applications.

REFERENCES


AUTHOR BIOGRAPHIES

LIH-YUAN DENG is a visiting associate professor in the Department of Mathematical Sciences at the University of Houston–Clear Lake. He is on leave from the Department of Mathematical Sciences at Memphis State University. He received the B.S. and M.S. degree in Mathematics in 1975 and 1977, both from the National Taiwan University. He then received the M.S. degree in Computer Science in 1982 and Ph. D. degree in Statistics in 1984, both from the University of Wisconsin–Madison. His research interests are in random number generation, survey sampling design and analysis, variance estimation,
Table A-1: Percentage of $\chi^2$ statistics > $\chi^2_\alpha$, goodness-of-fit test

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>$X_i = \sum_{i=0}^{4} c_j U_{i+j}$</th>
<th>$Y_i = \sum_{i=1}^{4} X_{4(i-1)+j} \mod 1$</th>
<th>$Z_i = 3X_{2i} + 5X_{2i-1} \mod 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>1.000</td>
<td>0.102 0.105 0.105 0.102</td>
<td>0.099 0.097 0.099 0.098</td>
</tr>
<tr>
<td>0.05</td>
<td>1.000</td>
<td>0.053 0.052 0.054 0.053</td>
<td>0.047 0.049 0.052 0.048</td>
</tr>
<tr>
<td>0.01</td>
<td>1.000</td>
<td>0.010 0.012 0.012 0.010</td>
<td>0.010 0.011 0.011 0.008</td>
</tr>
</tbody>
</table>

Table A-2: Percentage of $\chi^2$ statistics > $\chi^2_\alpha$, Good's serial pair test

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>$X_i = \sum_{i=0}^{4} c_j U_{i+j}$</th>
<th>$Y_i = \sum_{i=1}^{4} X_{4(i-1)+j} \mod 1$</th>
<th>$Z_i = 3X_{2i} + 5X_{2i-1} \mod 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>1.000</td>
<td>0.101 0.099 0.101 0.103</td>
<td>0.100 0.104 0.102 0.102</td>
</tr>
<tr>
<td>0.05</td>
<td>1.000</td>
<td>0.054 0.051 0.055 0.052</td>
<td>0.055 0.053 0.053 0.054</td>
</tr>
<tr>
<td>0.01</td>
<td>1.000</td>
<td>0.011 0.011 0.014 0.011</td>
<td>0.013 0.010 0.012 0.012</td>
</tr>
</tbody>
</table>

Table A-3: Percentage of $\chi^2$ statistics > $\chi^2_\alpha$, triplets test

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>$X_i = \sum_{i=0}^{4} c_j U_{i+j}$</th>
<th>$Y_i = \sum_{i=1}^{4} X_{4(i-1)+j} \mod 1$</th>
<th>$Z_i = 3X_{2i} + 5X_{2i-1} \mod 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>1.000</td>
<td>0.104 0.098 0.104 0.107</td>
<td>0.106 0.102 0.104 0.105</td>
</tr>
<tr>
<td>0.05</td>
<td>1.000</td>
<td>0.054 0.049 0.054 0.056</td>
<td>0.055 0.054 0.052 0.055</td>
</tr>
<tr>
<td>0.01</td>
<td>1.000</td>
<td>0.011 0.011 0.013 0.011</td>
<td>0.012 0.011 0.011 0.010</td>
</tr>
</tbody>
</table>

Table A-4: Percentage of $\chi^2$ statistics > $\chi^2_\alpha$, $d^2$ test

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>$X_i = \sum_{i=0}^{4} c_j U_{i+j}$</th>
<th>$Y_i = \sum_{i=1}^{4} X_{4(i-1)+j} \mod 1$</th>
<th>$Z_i = 3X_{2i} + 5X_{2i-1} \mod 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>1.000</td>
<td>0.108 0.097 0.101 0.104</td>
<td>0.098 0.096 0.103 0.103</td>
</tr>
<tr>
<td>0.05</td>
<td>1.000</td>
<td>0.056 0.050 0.051 0.051</td>
<td>0.049 0.047 0.053 0.049</td>
</tr>
<tr>
<td>0.01</td>
<td>1.000</td>
<td>0.013 0.009 0.011 0.010</td>
<td>0.010 0.010 0.010 0.009</td>
</tr>
</tbody>
</table>

Table B-1: Percentage of $\chi^2$ statistics > $\chi^2_\alpha$, goodness-of-fit test

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>$X_i = \sum_{i=0}^{4} c_j U_{i+j}$</th>
<th>$Y_i = \sum_{i=1}^{4} X_{4(i-1)+j} \mod 1$</th>
<th>$Z_i = 3X_{2i} + 5X_{2i-1} \mod 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>1.000</td>
<td>0.102 0.104 0.104 0.101</td>
<td>0.116 0.100 0.100 0.108</td>
</tr>
<tr>
<td>0.05</td>
<td>1.000</td>
<td>0.051 0.054 0.054 0.052</td>
<td>0.060 0.050 0.050 0.053</td>
</tr>
<tr>
<td>0.01</td>
<td>1.000</td>
<td>0.011 0.009 0.009 0.010</td>
<td>0.013 0.012 0.012 0.010</td>
</tr>
</tbody>
</table>

Table B-2: Percentage of $\chi^2$ statistics > $\chi^2_\alpha$, Good's serial pair test

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>$X_i = \sum_{i=0}^{4} c_j U_{i+j}$</th>
<th>$Y_i = \sum_{i=1}^{4} X_{4(i-1)+j} \mod 1$</th>
<th>$Z_i = 3X_{2i} + 5X_{2i-1} \mod 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>1.000</td>
<td>0.101 0.101 0.101 0.096</td>
<td>0.106 0.097 0.097 0.099</td>
</tr>
<tr>
<td>0.05</td>
<td>1.000</td>
<td>0.050 0.051 0.051 0.050</td>
<td>0.055 0.049 0.049 0.050</td>
</tr>
<tr>
<td>0.01</td>
<td>1.000</td>
<td>0.010 0.011 0.011 0.012</td>
<td>0.010 0.010 0.010 0.011</td>
</tr>
</tbody>
</table>
Table B-3: Percentage of $\chi^2$ statistics $> \chi^2_\alpha$, triplets test

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>$X_i = \sum_{j=0}^4 c_j U_{i+j}$</th>
<th>$Y_i = \sum_{j=1}^4 X_4(i-1)+j \mod 1$</th>
<th>$Z_i = 3X_{2i} + 5X_{2i-1} \mod 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>1.000</td>
<td>0.096 0.108 0.108 0.100</td>
<td>0.104 0.101 0.101 0.106</td>
</tr>
<tr>
<td>0.05</td>
<td>1.000</td>
<td>0.048 0.055 0.055 0.049</td>
<td>0.054 0.053 0.053 0.053</td>
</tr>
<tr>
<td>0.01</td>
<td>1.000</td>
<td>0.009 0.010 0.010 0.010</td>
<td>0.010 0.011 0.011 0.010</td>
</tr>
</tbody>
</table>

Table B-4: Percentage of $\chi^2$ statistics $> \chi^2_\alpha$, $d^2$ test

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>$X_i = \sum_{j=0}^4 c_j U_{i+j}$</th>
<th>$Y_i = \sum_{j=1}^4 X_4(i-1)+j \mod 1$</th>
<th>$Z_i = 3X_{2i} + 5X_{2i-1} \mod 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>1.000</td>
<td>0.099 0.108 0.108 0.098</td>
<td>0.113 0.098 0.098 0.100</td>
</tr>
<tr>
<td>0.05</td>
<td>1.000</td>
<td>0.049 0.055 0.055 0.049</td>
<td>0.058 0.048 0.048 0.051</td>
</tr>
<tr>
<td>0.01</td>
<td>1.000</td>
<td>0.011 0.011 0.011 0.011</td>
<td>0.011 0.009 0.009 0.011</td>
</tr>
</tbody>
</table>

and statistical computing. He is a member of ACM and ASA.

**E. OLUSEGUN GEORGE** is an associate professor in the Department of Mathematical Sciences at Memphis State University. His research interest includes generalized linear models, logistic regression, distribution theory and random number generation. He is a chapter representative of the American Statistical Association.

**YU-CHAO CHU** is a research associate in the Department of Preventive Medicine at the University of Tennessee - Memphis. She received the M.S. and Ph. D. degree in Mathematics (with concentration in Statistics) in 1985 and 1990, both from Memphis State University. Her research interest include random number generation, statistical computing, and biostatistics.