Statistica Sinica **7**(1997), 993-1003

STATISTICAL JUSTIFICATION OF COMBINATION GENERATORS

Lih-Yuan Deng, Dennis K. J. Lin, Jiannong Wang and Yilian Yuan

University of Memphis, Pennsylvania State University, Smithkline Beecham and IMS America

Abstract: The combination generator, first proposed by Wichmann and Hill (1982), is constructed by taking the fractional part of the sum of several random number generators. It is probably one of the most popular random number generators used. Its empirical performance is superior to the classical Lehmer congruential generator. However, its theoretical justification is somewhat primitive. In this paper, we give some theoretical support for such an important generator, from a statistical theory viewpoint. Specifically, we prove that the combination generator method is superior to each component random number generator method, in terms of (1) uniformity and (2) independence.

Key words and phrases: Asymptotic independence, asymptotic uniformity, combination generator, Lehmer's congruential generator.

1. Introduction

Consider the following n multiplicative linear congruential generators (MLCGs), proposed by Lehmer (1951):

$$X_{j,i+1} = B_j X_{j,i} \bmod m_j, \quad i \geq 0, j = 1, \ldots, n,$$

where $X_{j,0}$(initial seed), B_j(multiplier) are positive integers and m_j(modulus) are different prime numbers.

Wichmann and Hill (1982) suggested to add three MLCGs and take the fractional part:

$$U_{W,i} = \sum_{j=1}^{3} \frac{X_{j,i}}{m_j} \bmod 1.$$

Through a simple example, they claimed that this procedure "ironed out" the imperfections in the component variates. Zeisel (1986) observed that a linear combination of several MLCGs with different modulus is equivalent to another MLCG with a large multiplier and a large modulus.

L'Ecuyer (1988) considered a variation of the Wichmann and Hill (1982) method:

$$U_{L,i} = \sum_{j=1}^{n} \frac{\delta_j X_{j,i}}{m_1} \bmod 1,$$

where $\delta_j = (-1)^{j-1}$. He proved that if generators are independent of each other and if one of the generators is uniformly distributed, then the combined generator will also be uniformly distributed. L'Ecuyer and Tezuka (1991) studied the structural properties of these two classes of combined random number generators (RNGs) and they extended the observation by Zeisel (1986).

Several authors have performed empirical studies of the combination generator. Marsaglia (1985) empirically compared several popular generators and concluded that the combination generator is superior to others. Collings (1987), L'Ecuyer (1988) and Anderson (1990) also found good empirical performance of the combination generators.

Some theoretical justifications were given in Horton (1948), Horton and Smith (1949), Brown and Solomon (1979), Marsaglia (1985) and L'Ecuyer (1988). For additional justification of the combined generators, see Deng and George (1990), Deng, George and Chu (1991) and Deng and Chu (1991). Formal definition of the combination generator method and some of its properties are given in Section 2. The main results which include all theoretical results mentioned above as special cases are presented in Section 3. Specifically, we show that the combination generator method is superior to individual generators in terms of uniformity and independence.

2. Combination Generators

Suppose that $\{(U_{j0}, U_{j1}, U_{j2}, \ldots), j = 1, \ldots, n\}$ are n sequences of random variates generated by any RNG. No assumption is necessary regarding how each RNG is generated. Our goal is to study, from a statistical viewpoint, the property of the combined sequence $\{Y_1, Y_2, \ldots\}$ where $Y_i = U_{1i} + U_{2i} + \cdots + U_{ni} \bmod 1$. In particular, for any positive integer k, we investigate the joint probability distribution of the k-dimensional random vector $\boldsymbol{Y} = \boldsymbol{X}_1 + \boldsymbol{X}_2 + \cdots + \boldsymbol{X}_n \bmod 1$, where $\boldsymbol{X}_j = (U_{ji_1}, \ldots, U_{ji_k})'$ with $i_1 < i_2 < \cdots < i_k$. For a vector $\boldsymbol{x} = (x_1, \ldots, x_k)'$, let $\boldsymbol{x} \bmod 1 = (x_1 \bmod 1, x_2 \bmod 1, \ldots, x_k \bmod 1)'$.

Throughout this paper, we use a k-dimensional random vector $\boldsymbol{X}$ to represent a specific k components realization of a RNG. Furthermore, for simplicity, we assume the existence of the p.d.f. $f_{\boldsymbol{X}}(\boldsymbol{x})$ for $\boldsymbol{X}$. Although this assumption may not be realistic since any RNG can only generate finitely many points in $[0, 1]$, and any theory based on the assumption of the existence of an RNG with a p.d.f. is only an approximation, nevertheless this greatly reduces the need for exact computations with discrete values.

Let $\boldsymbol{\delta}_j$ be the k-vector vertex in $[0, 1]^k$ corresponding to the binary representation of j for $j = 0, 1, 2, \ldots, 2^k - 1$. For example, for $k = 2$, $\boldsymbol{\delta}_0 = (0, 0)'$, $\boldsymbol{\delta}_1 = (0, 1)'$, $\boldsymbol{\delta}_2 = (1, 0)'$ and $\boldsymbol{\delta}_3 = (1, 1)'$. For each $\boldsymbol{y} \in (0, 1)^k$, there are exactly 2^k partitions of $[0, 1]^k$, $\{A_{j,\boldsymbol{y}}, j = 0, 1, 2, \ldots, 2^k - 1\}$, where $A_{j,\boldsymbol{y}}$ corresponds to

the sub-cube of the partition containing the jth vertex $\boldsymbol{\delta}_j$. For example, if $k = 2$, $\boldsymbol{y} = (y_1, y_2)'$, $A_{0,\boldsymbol{y}} = [0, y_1] \times [0, y_2]$, $A_{1,\boldsymbol{y}} = [0, y_1] \times [y_2, 1]$, $A_{2,\boldsymbol{y}} = [y_1, 1] \times [0, y_2]$, $A_{3,\boldsymbol{y}} = [y_1, 1] \times [y_2, 1]$.

The following lemma establishes some properties of $A_{j,\boldsymbol{y}}$.

Lemma 2.1. *Let $A_{j,\boldsymbol{y}}, j = 0, 1, 2, \ldots, 2^k - 1$ be defined as above and $\boldsymbol{y} \in (0, 1)^k$. Let $\boldsymbol{I}_{A_{j,\boldsymbol{y}}}(\boldsymbol{x})$ be the indicator function for $A_{j,\boldsymbol{y}}$.*

(1) *For any integrable function $g(\boldsymbol{x})$ defined over $A_{j,\boldsymbol{y}}$,*

$$\int_{x \in A_{j,\boldsymbol{y}}} g(\boldsymbol{\delta}_j + \boldsymbol{y} - \boldsymbol{x}) d\boldsymbol{x} = \int_{x \in A_{j,\boldsymbol{y}}} g(\boldsymbol{x}) d\boldsymbol{x}.$$

(2) *For any function $h(\boldsymbol{x})$ defined over $[0, 1]^k$ and any number r,*

$$\Big| \sum_{j=0}^{2^k-1} \boldsymbol{I}_{A_{j,\boldsymbol{y}}}(\boldsymbol{x}) h(\boldsymbol{x})|^r = \sum_{j=0}^{2^k-1} \boldsymbol{I}_{A_{j,\boldsymbol{y}}}(\boldsymbol{x}) |h(\boldsymbol{x})\Big|^r.$$

Proof. Note that $A_{j,\boldsymbol{y}}$ is symmetric around $(\boldsymbol{\delta}_j + \boldsymbol{y})/2$ which is the center of the subcube $A_{j,\boldsymbol{y}}$. Therefore, $\boldsymbol{x} \in A_{j,\boldsymbol{y}}$ if and only if $\boldsymbol{\delta}_j + \boldsymbol{y} - \boldsymbol{x} \in A_{j,\boldsymbol{y}}$. To prove Part (1), we use this fact and make a change of variable $\boldsymbol{u} = \boldsymbol{\delta}_j + \boldsymbol{y} - \boldsymbol{x}$. Part (2) follows from the fact that the indicator functions $\boldsymbol{I}_{A_{j,\boldsymbol{y}}}(\boldsymbol{x})$'s are mutually exclusive: for each $\boldsymbol{x} \in [0, 1]^k$, exactly one term in the summation has value with $\boldsymbol{I}_{A_{j,\boldsymbol{y}}}(\boldsymbol{x}) = 1$ while all the remaining terms are zero.

We first find the p.d.f. of the fractional part of the sum of two independent random vectors. The following Lemma from Deng and Chu (1991) can be considered as a k-dimensional extension of Deng and George (1990).

Lemma 2.2. (Deng and Chu (1991), Theorem 1) *Let $\boldsymbol{X}_1, \boldsymbol{X}_2$ be any two independent random vectors over $[0, 1]^k$, with the p.d.f.'s $f_{\boldsymbol{X}_1}(\boldsymbol{x})$ and $f_{\boldsymbol{X}_2}(\boldsymbol{x})$. Let $\boldsymbol{Y} = \boldsymbol{X}_1 + \boldsymbol{X}_2 \, mod 1$, and $f_{\boldsymbol{Y}}(\boldsymbol{y})$ be the p.d.f. of $\boldsymbol{Y}$. In addition,*

$$f_{\boldsymbol{X}_i}(\boldsymbol{x}) = 1 + g_{\boldsymbol{X}_i}(\boldsymbol{x}), \quad \text{for } i = 1, 2, \tag{2.1}$$

where $g_{\boldsymbol{X}_i}(\boldsymbol{x})$ is the "deviation" of the p.d.f. of $\boldsymbol{X}_i$ from the uniform p.d.f. Then

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = 1 + \sum_{j=0}^{2^k-1} \int_{x \in A_{j,\boldsymbol{y}}} g_{\boldsymbol{X}_2}(\boldsymbol{\delta}_j + \boldsymbol{y} - \boldsymbol{x}) \cdot g_{\boldsymbol{X}_1}(\boldsymbol{x}) d\boldsymbol{x}. \tag{2.2}$$

3. Main Results

For $0 < \epsilon < 1$, let $L_r(\epsilon)$ be the class of p.d.f. $f_{\boldsymbol{X}}(\boldsymbol{x})$ over $[0, 1]^k$ which is in the neighborhood of the uniform p.d.f. such that

$$||f_{\boldsymbol{X}}(\boldsymbol{x}) - 1||_r = \Big(\int_{[0,1]^k} |f_{\boldsymbol{X}}(\boldsymbol{x}) - 1|^r d\boldsymbol{x} \Big)^{1/r} \leq \epsilon.$$

The value of 1 is so special here because it is the p.d.f. of $U[0,1]^k$. Furthermore, we set $0 < \epsilon < 1$, because it is desirable for a vector RNG to generate a distribution that is close to $U[0,1]^k$.

The following theorem provides a theoretical justification of the goodness of the combination generators. Specifically, the theorem shows that the fractional part of the sum of two independent (nearly) uniform random vectors will produce a distribution whose p.d.f. is closer to a uniform distribution.

Theorem 3.1. *Let $\boldsymbol{X}_1, \boldsymbol{X}_2$ be any two independent random vectors over $[0,1]^k$, with the p.d.f.'s $f_{\boldsymbol{X}_1}(\boldsymbol{x})$ and $f_{\boldsymbol{X}_2}(\boldsymbol{x})$. Let $\boldsymbol{Y} = \boldsymbol{X}_1 + \boldsymbol{X}_2 \, mod 1$, and $f_{\boldsymbol{Y}}(\boldsymbol{y})$ be the p.d.f. of $\boldsymbol{Y}$. For any $r_1, r_2 \geq 1$ such that $\frac{1}{r_1} + \frac{1}{r_2} = 1$, we have*

$$|f_{\boldsymbol{Y}}(\boldsymbol{y}) - 1| \leq ||f_{\boldsymbol{X}_1}(\boldsymbol{x}) - 1||_{r_1} ||f_{\boldsymbol{X}_2}(\boldsymbol{x}) - 1||_{r_2}.$$

That is, if $f_{\boldsymbol{X}_1}(\boldsymbol{x}) \in L_{r_1}(\epsilon_1)$ and $f_{\boldsymbol{X}_2}(\boldsymbol{x}) \in L_{r_2}(\epsilon_2)$ then $f_{\boldsymbol{Y}}(\boldsymbol{y}) \in L_\infty(\epsilon_1 \cdot \epsilon_2)$.

Proof. Let $g_{\boldsymbol{X}_i}(\boldsymbol{x}) = f_{\boldsymbol{X}_i}(\boldsymbol{x}) - 1$ be defined as in (2.1). From Lemma 2.2, we know that the p.d.f. of $\boldsymbol{Y} = \boldsymbol{X}_1 + \boldsymbol{X}_2 \bmod 1$ is

$$\begin{aligned} f_{\boldsymbol{Y}}(\boldsymbol{y}) &= 1 + \sum_{j=0}^{2^k-1} \int_{x \in A_{j,\boldsymbol{y}}} g_{\boldsymbol{X}_2}(\boldsymbol{\delta}_j + \boldsymbol{y} - \boldsymbol{x}) \cdot g_{\boldsymbol{X}_1}(\boldsymbol{x}) d\boldsymbol{x} \\ &= 1 + \int_{\boldsymbol{x} \in [0,1]^k} \Big(\sum_{j=0}^{2^k-1} \boldsymbol{I}_{A_{j,\boldsymbol{y}}}(\boldsymbol{x}) g_{\boldsymbol{X}_2}(\boldsymbol{\delta}_j + \boldsymbol{y} - \boldsymbol{x}) \Big) \cdot g_{\boldsymbol{X}_1}(\boldsymbol{x}) d\boldsymbol{x} \\ &= 1 + \int_{\boldsymbol{x} \in [0,1]^k} h_{\boldsymbol{X}_2, \boldsymbol{y}}(\boldsymbol{x}) \cdot g_{\boldsymbol{X}_1}(\boldsymbol{x}) d\boldsymbol{x}, \end{aligned} \tag{3.1}$$

where

$$h_{\boldsymbol{X}_2, \boldsymbol{y}}(\boldsymbol{x}) = \sum_{j=0}^{2^k-1} \boldsymbol{I}_{A_{j,\boldsymbol{y}}}(\boldsymbol{x}) g_{\boldsymbol{X}_2}(\boldsymbol{\delta}_j + \boldsymbol{y} - \boldsymbol{x}),$$

and $\boldsymbol{I}_{A_{j,\boldsymbol{y}}}(\boldsymbol{x})$, the indicator function of $\boldsymbol{x}$ for $\boldsymbol{x} \in A_{j,\boldsymbol{y}}$, is defined in Lemma 2.1.

From Lemma 2.1,

$$\begin{aligned} \int_{\boldsymbol{x} \in [0,1]^k} |h_{\boldsymbol{X}_2, \boldsymbol{y}}(\boldsymbol{x})|^{r_1} d\boldsymbol{x} &= \int_{\boldsymbol{x} \in [0,1]^k} \Big(\Big| \sum_{j=0}^{2^k-1} \boldsymbol{I}_{A_{j,\boldsymbol{y}}}(\boldsymbol{x}) g_{\boldsymbol{X}_2}(\boldsymbol{\delta}_j + \boldsymbol{y} - \boldsymbol{x}) \Big| \Big)^{r_1} d\boldsymbol{x} \\ &= \sum_{j=0}^{2^k-1} \int_{\boldsymbol{x} \in [0,1]^k} \boldsymbol{I}_{A_{j,\boldsymbol{y}}}(\boldsymbol{x}) |g_{\boldsymbol{X}_2}(\boldsymbol{\delta}_j + \boldsymbol{y} - \boldsymbol{x})|^{r_1} d\boldsymbol{x} \\ &= \sum_{j=0}^{2^k-1} \int_{x \in A_{j,\boldsymbol{y}}} |g_{\boldsymbol{X}_2}(\boldsymbol{\delta}_j + \boldsymbol{y} - \boldsymbol{x})|^{r_1} d\boldsymbol{x} \end{aligned} \tag{3.2}$$

$$= \sum_{j=0}^{2^k-1} \int_{\boldsymbol{z}\in A_{j,\boldsymbol{y}}} |g_{\boldsymbol{X}_2}(\boldsymbol{z})|^{r_1} d\boldsymbol{z} \tag{3.3}$$
$$= \int_{\boldsymbol{z}\in[0,1]^k} |g_{\boldsymbol{X}_2}(\boldsymbol{z})|^{r_1} d\boldsymbol{z}.$$

To derive equations (3.2) and (3.3), we have used Lemma 2.1. Applying Hölder's inequality (Hardy, Littlewood and Pólya (1952), page 156, Theorem 210) for (3.1), we have

$$|f_{\boldsymbol{Y}}(\boldsymbol{y}) - 1| \le \Big(\int_{\boldsymbol{x}\in[0,1]^k} |h_{\boldsymbol{X}_2,\boldsymbol{y}}(\boldsymbol{x})|^{r_1} d\boldsymbol{x} \Big)^{1/r_1} \Big(\int_{\boldsymbol{x}\in[0,1]^k} |g_{\boldsymbol{X}_1}(\boldsymbol{x})|^{r_2} d\boldsymbol{x} \Big)^{1/r_2}$$
$$= \Big(\int_{\boldsymbol{z}\in[0,1]^k} |g_{\boldsymbol{X}_2}(\boldsymbol{z})|^{r_1} d\boldsymbol{z} \Big)^{1/r_1} \Big(\int_{\boldsymbol{x}\in[0,1]^k} |g_{\boldsymbol{X}_1}(\boldsymbol{x})|^{r_2} d\boldsymbol{x} \Big)^{1/r_2}.$$

To see that Theorem 3.1 is a much stronger result than that in Deng and George (1990) and Deng and Chu (1991), we note that by the Liapounov inequality (Hardy, Littlewood and Pólya (1952), page 157, Theorem 211), we have

$$\Big(\int_{[0,1]^k} |f_{\boldsymbol{X}}(\boldsymbol{x}) - 1|^r d\boldsymbol{x} \Big)^{1/r} \le \Big(\int_{[0,1]^k} |f_{\boldsymbol{X}}(\boldsymbol{x}) - 1|^s d\boldsymbol{x} \Big)^{1/s}$$

for $r < s < \infty$. Therefore, $L_s(\epsilon) \subset L_r(\epsilon),\ r < s < \infty$. Using the fact $f_{\boldsymbol{X}_i}(\boldsymbol{x}) \in L_\infty(\epsilon_i) \subset L_2(\epsilon_i)$, and Theorem 3.1 with $r_1 = r_2 = 2$, we have $f_{\boldsymbol{Y}}(\boldsymbol{y}) \in L_\infty(\epsilon_1 \cdot \epsilon_2)$. That is, if $|f_{\boldsymbol{X}_i}(\boldsymbol{x}) - 1| \le \epsilon_i$, for $i = 1, 2$, then $|f_{\boldsymbol{Y}}(\boldsymbol{y}) - 1| \le \epsilon_1 \cdot \epsilon_2$. Essentially, we have proved here that the combination generator will not only improve the "uniformity" of the generator but also its "independence" when considering the joint distribution of any k-dimensional random vectors.

According to Theorem 3.1, one can produce a random vector whose distribution is closer to $U[0,1]^k$ by taking the fractional part of the sum of several random vectors: Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be n independent random vectors over $[0,1]^k$, with the p.d.f. $f_{\boldsymbol{X}_i}(\boldsymbol{x})$, for $i = 1, \ldots, n$. Let $\boldsymbol{Y} = \sum_{i=1}^n \boldsymbol{X}_i \bmod 1$, and $f_{\boldsymbol{Y}}(\boldsymbol{y})$ be the p.d.f. of $\boldsymbol{Y}$.

(1) If $|f_{\boldsymbol{X}_i}(\boldsymbol{x}) - 1| \le \epsilon_i$, for $i = 1, \ldots, n$, then $|f_{\boldsymbol{Y}}(\boldsymbol{y}) - 1| \le \prod_{i=1}^n \epsilon_i$.
(2) If $\prod_{i=1}^n \epsilon_i \longrightarrow 0$, then $\boldsymbol{Y}$ converges in distribution to $U[0,1]^k$.
(3) In particular, if one of the $\boldsymbol{X}_i$ is uniformly distributed over $[0,1]^k$, then $\boldsymbol{Y}$ is uniformly distributed over $[0,1]^k$.

The major weakness of the above results is that we have assumed the $\boldsymbol{X}_i$'s are *independent* of each other. In practice, since a sequence generated by any RNG is deterministic, this independence assumption seems to be unrealistic. To show that the combination generator is indeed superior even when $\boldsymbol{X}_i$'s are dependent, we next consider the extreme case that all $\boldsymbol{X}_i$ are equal to $\boldsymbol{X}$. In this case, we

will show that the distribution of $\boldsymbol{Y} = n\boldsymbol{X} \bmod 1$ will be closer to the uniform distribution than that of each $\boldsymbol{X}$. In fact, this statement holds for a more general case

$$\boldsymbol{Y} = \boldsymbol{D}\boldsymbol{X} \bmod 1, \quad \boldsymbol{D} = \operatorname{diag}(n_1, \ldots, n_k), \tag{3.4}$$

where n_j's are positive integers. Furthermore, we show that for large n_j's, the distribution of $\boldsymbol{Y}$ is nearly distributed as $U[0,1]^k$.

Lemma 3.1. *Let $\boldsymbol{D}$ be defined as in* (3.4) *and $\boldsymbol{S_D} = \{\boldsymbol{i} = (i_1, \ldots, i_k) | 0 \le i_j \le n_j - 1\}$.*

(1) Let $\boldsymbol{X}$ be any random vector over $[0,1]^k$, with the p.d.f $f_{\boldsymbol{X}}(\boldsymbol{x})$. Then the p.d.f. of $\boldsymbol{Y} = \boldsymbol{D}\boldsymbol{X} \bmod 1$ is

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = \frac{1}{\prod_{j=1}^k n_j} \sum_{\boldsymbol{i} \in \boldsymbol{S_D}} f_{\boldsymbol{X}}(\boldsymbol{D}^{-1}(\boldsymbol{y} + \boldsymbol{i})).$$

(2) For any integrable function $e(\boldsymbol{x})$ defined over $[0,1]^k$, we have

$$\int_{\boldsymbol{x} \in [0,1]^k} e(\boldsymbol{x}) d\boldsymbol{x} = \frac{1}{\prod_{j=1}^k n_j} \sum_{\boldsymbol{i} \in \boldsymbol{S_D}} \int_{\boldsymbol{y} \in [0,1]^k} e(\boldsymbol{D}^{-1}(\boldsymbol{y} + \boldsymbol{i})) d\boldsymbol{y}.$$

Proof. Note that $\boldsymbol{Y} = \boldsymbol{D}\boldsymbol{X} \bmod 1$ is not a one-to-one transformation of $\boldsymbol{X}$. To prove Part (1), we partition $[0,1]^k$ into $\prod_{j=1}^k n_j$ subcubes, $B_{\boldsymbol{i}}, \boldsymbol{i} \in \boldsymbol{S_D}$, where $B_{\boldsymbol{i}}$ is a shift of the subcube $B_{\boldsymbol{0}} = [0, 1/n_1] \times [0, 1/n_2] \times \cdots \times [0, 1/n_k]$: $B_{\boldsymbol{i}} = B_{\boldsymbol{0}} + \boldsymbol{D}^{-1}\boldsymbol{i}$. The transformation $\boldsymbol{Y} = \boldsymbol{D}\boldsymbol{X} \bmod 1$ is now one to one over $\boldsymbol{x} \in B_{\boldsymbol{i}}$,

$$\boldsymbol{Y} = \boldsymbol{y} \quad \text{if and only if} \quad \boldsymbol{X} = \boldsymbol{D}^{-1}(\boldsymbol{y} + \boldsymbol{i}) \in B_{\boldsymbol{i}}, \quad \boldsymbol{i} \in \boldsymbol{S_D}.$$

Then the p.d.f. of $\boldsymbol{Y} = \boldsymbol{D}\boldsymbol{X} \bmod 1$ is

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = \sum_{\boldsymbol{i} \in \boldsymbol{S_D}} f_{\boldsymbol{X}}(\boldsymbol{D}^{-1}(\boldsymbol{y} + \boldsymbol{i})) J_i = \frac{1}{\prod_{j=1}^k n_j} \sum_{\boldsymbol{i} \in \boldsymbol{S_D}} f_{\boldsymbol{X}}(\boldsymbol{D}^{-1}(\boldsymbol{y} + \boldsymbol{i})),$$

where $J_{\boldsymbol{i}} = |\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{y}}| = \frac{1}{\prod_{j=1}^k n_j}$ is the Jacobian for the subcube $B_{\boldsymbol{i}}$. This proves Part (1).

To prove Part (2), we again use the same partition of $[0,1]^k$, $\{B_{\boldsymbol{i}}, \boldsymbol{i} \in \boldsymbol{S_D}\}$.

$$\int_{\boldsymbol{x} \in [0,1]^k} e(\boldsymbol{x}) d\boldsymbol{x} = \sum_{\boldsymbol{i} \in \boldsymbol{S_D}} \int_{\boldsymbol{x} \in B_{\boldsymbol{i}}} e(\boldsymbol{x}) d\boldsymbol{x} = \sum_{\boldsymbol{i} \in \boldsymbol{S_D}} \int_{\boldsymbol{x} \in B_{\boldsymbol{0}}} e(\boldsymbol{x} + \boldsymbol{D}^{-1}\boldsymbol{i}) d\boldsymbol{x}.$$

Letting $\boldsymbol{x} = \boldsymbol{D}^{-1}\boldsymbol{y}$ in the above integrals with the Jacobian $\frac{1}{\prod_{j=1}^k n_j}$, we have

$$\int_{\boldsymbol{x} \in [0,1]^k} e(\boldsymbol{x}) d\boldsymbol{x} = \frac{1}{\prod_{j=1}^k n_j} \sum_{\boldsymbol{i} \in \boldsymbol{S_D}} \int_{\boldsymbol{y} \in [0,1]^k} e(\boldsymbol{D}^{-1}(\boldsymbol{y} + \boldsymbol{i})) d\boldsymbol{y}.$$

Theorem 3.2. *Let $\boldsymbol{X}$ be any random vector over $[0,1]^k$, with the p.d.f. $f_{\boldsymbol{X}}(\boldsymbol{x})$. Let $\boldsymbol{Y}=\boldsymbol{DX} \bmod 1$, where $\boldsymbol{D}$ is defined as in (3.4), and $f_{\boldsymbol{Y}}(\boldsymbol{y})$ is the p.d.f. of $\boldsymbol{Y}$.*

(1) *For any $r \geq 1$, we have $||f_{\boldsymbol{Y}}(\boldsymbol{y})-1||_r \leq ||f_{\boldsymbol{X}}(\boldsymbol{x})-1||_r$. Therefore, if $f_{\boldsymbol{X}}(\boldsymbol{x}) \in L_r(\epsilon)$ then $f_{\boldsymbol{Y}}(\boldsymbol{y}) \in L_r(\epsilon)$.*

(2) *$|f_{\boldsymbol{Y}}(\boldsymbol{y})-1| \to 0$, as $\min(n_1,\ldots,n_k) \to \infty$. That is, the components of $\boldsymbol{Y}$ will not only be "more uniform" but also "more independent" of each other.*

(3) *Moreover, $\boldsymbol{Y}$ and $\boldsymbol{X}$ will be asymptotically independent of each other.*

Proof. From Part (1) of Lemma 3.1, it is easy to see that

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) - 1 = \frac{1}{\prod_{j=1}^k n_j} \sum_{\boldsymbol{i} \in \boldsymbol{S_D}} \Big(f_{\boldsymbol{X}}(\boldsymbol{D}^{-1}(\boldsymbol{y}+\boldsymbol{i})) - 1 \Big)$$

and

$$\begin{aligned} |f_{\boldsymbol{Y}}(\boldsymbol{y}) - 1|^r &\leq \Big[\frac{1}{\prod_{j=1}^k n_j} \sum_{\boldsymbol{i} \in \boldsymbol{S_D}} |f_{\boldsymbol{X}}(\boldsymbol{D}^{-1}(\boldsymbol{y}+\boldsymbol{i})) - 1| \Big]^r \\ &\leq \frac{1}{\prod_{j=1}^k n_j} \sum_{\boldsymbol{i} \in \boldsymbol{S_D}} |f_{\boldsymbol{X}}(\boldsymbol{D}^{-1}(\boldsymbol{y}+\boldsymbol{i})) - 1|^r. \end{aligned}$$

Here we used the following inequality, with $\phi(x) = x^r, r \geq 1$ and $m = \prod_{j=1}^k n_j$ (see Hardy, Littlewood and Pólya (1952), page 72, (3.6.1))

$$\phi\Big(\frac{1}{m}\sum_{i=1}^m x_i\Big) \leq \frac{1}{m}\sum_{i=1}^m \phi(x_i),$$

for any convex function $\phi(x)$.

Integrating $\boldsymbol{y}$ over $[0,1]^k$, we get

$$\begin{aligned} \int_{\boldsymbol{y}\in[0,1]^k} |f_{\boldsymbol{Y}}(\boldsymbol{y}) - 1|^r d\boldsymbol{y} &\leq \frac{1}{\prod_{j=1}^k n_j} \sum_{\boldsymbol{i} \in \boldsymbol{S_D}} \int_{\boldsymbol{y}\in[0,1]^k} |f_{\boldsymbol{X}}(\boldsymbol{D}^{-1}(\boldsymbol{y}+\boldsymbol{i})) - 1|^r d\boldsymbol{y} \\ &= \int_{\boldsymbol{x}\in[0,1]^k} |f_{\boldsymbol{X}}(\boldsymbol{x}) - 1|^r d\boldsymbol{x}. \end{aligned}$$

In the last equality, we have used Part (2) of Lemma 3.1. This proves Part (1). Part (2) follows easily from Part (1) of Lemma 3.1, because

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = \frac{1}{\prod_{j=1}^k n_j} \sum_{\boldsymbol{i} \in \boldsymbol{S_D}} f_{\boldsymbol{X}}(\boldsymbol{D}^{-1}(\boldsymbol{y}+\boldsymbol{i}))$$

is a Riemann sum of the k-fold integral of the p.d.f. $f_{\boldsymbol{X}}(\boldsymbol{x})$

$$\int_{\boldsymbol{x}\in[0,1]^k} f_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x} = 1$$

and n_j is the number of partitions in each coordinate.

To prove Part (3), it is sufficient to show that the jth component of $\boldsymbol{Y}$, say Y, is asymptotically independent of the jth component of $\boldsymbol{X}$, say X. For simplicity, we let $n = n_j$. For $0 < x, y < 1$, the joint c.d.f. of (X, Y) is

$$\begin{aligned} F_{X,Y}(x,y) &= \sum_{r=0}^{n-1} Pr(X \le x,\ r < nX \le r+y) \\ &= \sum_{r=0}^{n-1} Pr(X \le x,\ \frac{r}{n} < X \le \frac{r+y}{n}) \\ &= \sum_{r \le xn} Pr(X \le x,\ \frac{r}{n} < X \le \frac{r+y}{n}) \qquad (3.5) \\ &= \sum_{r \le xn} Pr(\frac{r}{n} < X \le \frac{r+y}{n}) + o(1) \qquad (3.6) \\ &= \sum_{r \le xn} \Big[F_X\Big(\frac{r+y}{n}\Big) - F_X\Big(\frac{r}{n}\Big)\Big] + o(1), \end{aligned}$$

where $F_X(\cdot)$ is the c.d.f. of X and $o(1)$ is a small order term converging to zero as $n \to \infty$. Note that (3.5) and (3.6) hold because

$$\{X \le x\} \cap \{\frac{r}{n} < X \le \frac{r+y}{n}\} = \begin{cases} \emptyset, & \text{if} \quad x < r/n, \\ \{\frac{r}{n} < X \le x\}, & \text{if} \quad r/n \le x < (r+y)/n, \\ \{\frac{r}{n} < X \le \frac{r+y}{n}\}, & \text{if} \quad x \ge (r+y)/n. \end{cases}$$

We then apply the Mean Value Theorem

$$F_X(b) - F_X(a) = (b-a) f_X(a + \theta(b-a)), \quad 0 \le \theta \le 1,$$

to each term in the summation. Thus, the joint c.d.f. of (X, Y) is

$$\begin{aligned} F_{X,Y}(x,y) &= y \sum_{r \le xn} f_X\,(\frac{\theta_r y + r}{n})\frac{1}{n} + o(1), \quad 0 \le \theta_r \le 1 \qquad (3.7) \\ &= y \int_0^x f_X(t)dt + o(1) \qquad (3.8) \\ &= yF_X(x) + o(1) \to yF_X(x) \text{ as } n \to \infty. \qquad (3.9) \end{aligned}$$

In (3.8), we used the fact that (3.7) is a Riemann sum of its integral.

From (3.9), we can see that X and Y are asymptotically independent of each other. This completes the proof of Part (3).

Part (1) of Theorem 3.2 shows that "stretching out" any continuous random vector $\boldsymbol{X}$ will be as good as the $\boldsymbol{X}$ itself. Part (2) shows that by stretching $\boldsymbol{X}$ in each direction and taking its fractional part we can obtain a distribution closer

to $U(0,1)$ while all components will be more independent of each other. Part (3) gives some justification for a MLCG with a large multiplier. It shows that the successive variates generated by a MLCG should be asymptotically independently and uniformly distributed.

4. Concluding Remarks

Some intuitive explanations for the results in Section 3 are possible. If we stretch out the p.d.f. of a continuous random vector, then it will become "flat" within each unit subcube. Taking the fractional part of the stretched vector will move it back to $[0,1]^k$ with its p.d.f. equal to the sum of the p.d.f. over each subcube. Consequently, we get a random vector that is more uniformly distributed over $[0,1]^k$ and meanwhile (roughly) independent of the original random vector(s).

Under the assumption of the existence of the p.d.f. of a random vector, we can do the stretching by either adding several random vectors (not necessarily independent of each other) (see Theorem 3.1)

$$\boldsymbol{Y} = \boldsymbol{X}_1 + \boldsymbol{X}_2 + \cdots + \boldsymbol{X}_n \bmod 1$$

or simply multiplying each component of $\boldsymbol{X}$ by a large number (see Theorem 3.2)

$$\boldsymbol{Y} = \boldsymbol{D}\boldsymbol{X} \bmod 1.$$

Since there are only a finite number of digits representable by a computer, the above assumption is not realistic for a discrete RNG. However, according to an extensive empirical study by Deng, George and Chu (1991), both methods can indeed be used to improve the uniformity of an RNG. Note that adding several variates together will increase the variability of the combined variate even for a discrete distribution. Therefore, in practice, the result in Theorem 3.1 should be more applicable than that in Theorem 3.2.

It follows immediately that the combination generator is a preferred method of improving one or several RNGs. There are several additional justifications: (1) multiplying a number in a MLCG will neither change its recurrence relationship nor will it have any effect on its lattice structure (2) combining several MLCG's with different modulus will result in an RNG with much larger period (e.g., see Wichmann and Hill (1982) and L'Ecuyer (1988)).

As noted in Section 1, L'Ecuyer and Tezuka (1991) showed that a linear combination of several MLCGs with different modulus is equivalent to another MLCG with a large multiplier and a large modulus. Their result is also consistent with Theorem 3.2 in that a large multiplier and large modulus will, in general, improve the statistical properties. Clearly, Theorem 3.1 also provides additional justification for the combination generator from a statistical theory viewpoint.

Our results can also be used to justify the statistical properties of the multiple recurrence generator (MRG). (For a review of the MRG, combination generator and other RNGs, see L'Ecuyer (1989, 1990)). The MRG is generated from (e.g. Knuth (1981), pages 28-29) a degree k primitive polynomial $f(x) = x^k - \alpha_1 x^{k-1} - \cdots - \alpha_k$ with period $p^k - 1$ by

$$X_m = (\alpha_1 X_{m-1} + \cdots + \alpha_k X_{m-k}) \bmod p, \quad m \geq k$$

for any initial non-zero vector $(X_0, \ldots, X_{k-1})$, where p is a large prime number. L'Ecuyer and Blouin (1988) has implemented Knuth's algorithm in a computer program to find some MRGs. Deng, Rousseau and Yuan (1992) gave an efficient search algorithm for an MRG and discussed its statistical properties. They gave theoretical justifications of the MRG using the results obtained in Deng and George (1990), Deng, George and Chu (1991) and Deng and Chu (1991). This paper is a further extension of the above theoretical results.

References

Anderson, S. L. (1990). Random number generators on vector supercomputers and other advanced architectures. *SIAM Review* **32**, 221-251.

Brown, M. and Solomon, H. (1979). On combining pseudorandom number generators. *Ann. Statist.* **3**, 691-695.

Collings, B. J. (1987). Compound random number generators. *J. Amer. Statist. Assoc.* **82**, 525-527.

Deng, L. Y. and Chu, Y. C. (1991). Combining random number generators. In *Proceedings of the* 1991 *Winter Simulation Conference*, 1043-1046.

Deng, L. Y. and George, E. O. (1990). Generation of uniform variates from several nearly uniformly distributed variables. *Comm. Statist.* **19**, 145-154.

Deng, L. Y., George, E. O. and Chu, Y. C. (1991). On improving pseudo-random number generators. In *Proceedings of the* 1991 *Winter Simulation Conference*, 1035-1042

Deng, L. Y., Rousseau, C. and Yuan, Y. (1992). Generalized Lehmer-Tausworthe random number generators. In *Proceedings of the* 30*th Annual ACM Southeast Regional Conference*, Raleigh, North Carolina, April 8-10, 108-115.

Hardy, G. H., Littlewood, J. E. and Pólya, G. (1952). *Inequalities*, Second Edition, Cambridge University Press, Cambridge.

Horton, H. B. (1948). A method for obtaining random numbers. *Ann. Math. Statist.* **19**, 81-85.

Horton, H. B. and Smith III, R. T. (1949). A direct method for producing random digits in any number system. *Ann. Math. Statist.* **20**, 82-90.

Knuth, D. E. (1981). *The Art of Computer Programming*, Vol 2: Seminumerical Algorithms, Second Edition, Addison-Wesley, Reading, Mass.

L'Ecuyer, P. (1988). Efficient and portable combined random number generators. *Comm. ACM* **31**, 742-748, 774.

L'Ecuyer, P. (1989). A tutorial on uniform variate generation. In *Proceedings of the* 1989 *Winter Simulation Conference*, IEEE Press, 40-49.

L'Ecuyer, P. (1990). Random numbers for simulation. *Comm. ACM* **33**, 85-97.

L'Ecuyer, P. and Blouin, F. (1988). Linear congruential generators of order $k > 1$. In *Proceedings of the* 1988 *Winter Simulation Conference*, IEEE Press, 432-439.

L'Ecuyer, P. and Tezuka, S. (1991). Structural properties for two classes of combined random number generators. *Math. Comp.* **57**, 735-746.

Lehmer, D. H. (1951). Mathematical methods in large-scale computing units. In *Proceedings of the Second Symposium on Large Scale Digital Computing Machinery*, Harvard University Press, Cambridge, 141-146.

Marsaglia, G. (1985). A current view of random number generators. In *Proceedings of the* 16*th Symposium on the Interface* (Editor by L. Billard), 3-10. Elsevier Science Publishers, North-Holland.

Wichmann, B. A. and Hill, I. D. (1982). An efficient and portable pseudo-random number generator. *Appl. Statist.* **31**, 188-190.

Zeisel, H. (1986). A remark on algorithm AS 183. *Appl. Statist.* **35**, p. 89.

Department of Mathematics Sciences, University of Memphis, Memphis, TN 38152, U.S.A.

E-mail: dengl@hermes.msci.memphis.edu

Department of Management Science and Information System, Pennsylvania State University, University Park, PA 16802-1913, U.S.A.

E-mail: dkl5@psuvm.psu.edu

Department of Biometrics, Smithkline Beecham 1250 S. Collegeville Road, Collegeville, PA 19426, U.S.A.

E-mail: jim-wang-1@sbphrd.com

Department of Statistics Services, IMS America 660 W. Germantown Pike Plymouth Meeting, PA 19462, U.S.A.

E-mail: yuanyϕ1@imsint.com

(Received December 1995; accepted February 1997)