The Annals of Statistics
1979, Vol. 7, No. 3, 691–695

ON COMBINING PSEUDORANDOM NUMBER GENERATORS[1]

BY MARK BROWN AND HERBERT SOLOMON

City College, CUNY and Memorial Sloan-Kettering Cancer Center; and Stanford University

A technique used in pseudorandom number generation is to combine two or more different generators with the goal of producing a new generator with improved randomness properties. We study such a class of generators and show that in a strong sense the combined generator does offer improvement. Our approach applies results from majorization theory.

1. Introduction. Many methods have been proposed, tested and employed for generating pseudorandom numbers ([2], [3], [4], [5], [8], [9], [11], [12], [14], [16], [18], [19]). The goal is to produce strings of numbers which behave like independent uniform [0, 1] random variables. The generators yield integers in the set $\{0, 1, \cdots, m-1\}$, which are then transformed to [0, 1] by division by m. Suppose that $X_1, X_2, \cdots$ and $Y_1, Y_2, \cdots$ are strings of numbers generated by two separate generators. Various suggestions have been made for combining the two strings to produce a new string $Z_1, Z_2, \cdots$ which hopefully improves upon X and Y. One method (discussed in Knuth [8], pages 26–27) is to set $Z_i = X_i + Y_i$ (mod m). Another, due to Maclaren and Marsaglia [11], which Knuth reports to be excellent ([8], page 31), uses the Y string to randomly permute the X string.

For the additive generator $Z_i = X_i + Y_i$ (mod m) we obtain the following result (Remark 1). For any k and corresponding choice of indices $i_1 < i_2 < \cdots < i_k$ consider the vectors $X_A = (X_{i_1}, \cdots, X_{i_k})$, $Y_A = (Y_{i_1}, \cdots, Y_{i_k})$ and $Z_A = (Z_{i_1}, \cdots, Z_{i_k})$. Let p_A, q_A and s_A denote the respective distributions of X_A, Y_A and Z_A; p_A, q_A and s_A are probability distributions on $\mathfrak{M}^k$ where $\mathfrak{M} = \{0, 1, \cdots, m-1\}$. Define r_k to be the uniform distribution over $\mathfrak{M}^k$; r_k is a vector of m^k components each equal to m^{-k}. Let $\|\cdot\|$ be an arbitrary symmetric norm on R^{m^k} ($\|x\| = \|\Pi x\|$ where Πx is any permutation of x). Then $\|s_A - r_k\| \leqslant \min(\|p_A - r_k\|, \|q_A - r_k\|)$.

For the generator suggested by Maclaren and Marsaglia a similar but weaker result is obtained. Using Y to shuffle $(X_1, \cdots, X_m)$ results in improvement for the joint distribution of $X_1, \cdots, X_m$ but not necessarily for the marginal distributions of subsets.

Received July 1976; revised January 1978.

[1]Research partially supported by ONR Contract N00014-76-C-0475 and ARO Grant DAAG29-77-G-0031

AMS 1970 *subject classifications.* Primary 65C10; secondary 68A55.

Key words and phrases. Pseudorandom number generators, Monte Carlo simulation, majorization, uniform distribution, Markov chains.

The potential value of our approach is that it can provide additional justification for some generators currently in use, and perhaps suggest new generators which would then be analyzed by traditional methods.

In our analysis we treat the strings X and Y as independent random vectors. In practice X and Y are deterministic strings of numbers. This creates a problem in the strict application of our results to pseudorandom number generation.

2. Majorization. By definition ([6], page 45) an n-vector a is said to be majorized by an n-vector b if upon reordering to achieve $a_1 \geqslant a_2 \geqslant \cdots \geqslant a_n$ and $b_1 \geqslant b_2 \geqslant \cdots \geqslant b_n$ it follows that $\Sigma_1^k a_i \leqslant \Sigma_1^k b_i$ for $k = 1, \cdots, n-1$ and $\Sigma_1^n a_i = \Sigma_1^n b_i$. A function ψ, $R^n \to R$, is defined to be Schur convex ([13], page 1189) if, whenever a is majorized by b, $\psi(a) \leqslant \psi(b)$. Schur convex functions include symmetric convex functions which in turn include symmetric gauge functions and symmetric norms ([1], page 229). By a symmetric norm on R^n we mean a function $\|\cdot\|$, $R^n \to R$, satisfying: $\|x\| \geqslant 0$ for all $x \in R^n$ with equality if and only if $x = 0$, $\|\alpha x\| = |\alpha|\,\|x\|$ for all $\alpha \in R$, $x \in R^n$, $\|x + y\| \leqslant \|x\| + \|y\|$ for all $x, y \in R^n$, and $\|x\| = \|\Pi x\|$ for all $x \in R^n$ and for all permutations Πx of x. We note that if r is the uniform distribution over $\{1, 2, \cdots, n\}$ $(r(i) = 1/n, i = 1, \cdots, n)$ and $\|\cdot\|$ is a symmetric norm on R^n, then $g(x) = \|x - r\|$ is a symmetric convex function and is, thus, Schur convex. Some references for majorization are [1], [6], [13] and [17].

Lemma 1 below contains four equivalent statements relating to majorization. The equivalence between (i) and (ii) is due to Hardy, Littlewood and Polya ([6], page 49); the fact that (ii) implies (iii) is found in [1], page 183 and (iii) $\Rightarrow$ (ii) in [1], page 181; the fact that (i) $\Rightarrow$ (iv) is the definition of Schur convexity and (iv) $\Rightarrow$ (i) because $\psi(x_1, \cdots, x_n) = \Sigma_1^j x_{(i)}$, where $x_{(i)}$ is the ith largest component of x, is symmetric and convex, and, therefore, Schur convex.

LEMMA 1. *The following statements are equivalent:*

(i) *a is majorized by b;*

(ii) *$a = Pb$ where P is doubly stochastic;*

(iii) *a is a mixture of permutations of b,* i.e., *$a = \Sigma p_i(\Pi_i b)$ where $(p_1, \cdots, p_n)$ is a probability vector and each $\Pi_i b$ is a permutation of b;*

(iv) *$\psi(a) \leqslant \psi(b)$ for all Schur convex functions ψ.*

THEOREM 1. *Suppose that X is a discrete random variable taking values in the set $\mathcal{X} = \{x_1, \cdots, x_n\}$ with probability distribution $p = (p_1, \cdots, p_n)$, where $p_i = P(x_i)$, and Y is a random variable, independent of X, taking values in the set $\mathcal{Y}$. For each $y \in \mathcal{Y}$ let T_y be a* 1-1 *transformation of $\mathcal{X}$ onto itself. Define $Z = T_Y X$ and let s be the distribution of Z. Then s is majorized by p.*

PROOF. Since T_y is 1-1 and onto the distribution of $T_y X$ is a permutation of p. Thus s is a mixture of permutations of p. By Lemma 1 s is majorized by p. $\square$

3. Applications to pseudorandom number generation. Suppose that $X = (X_1, \cdots, X_N)$ and Y are independent random vectors, with each X_i assuming values in $\mathfrak{M} = \{0, 1, \cdots, m-1\}$. Consider a subset of k indices $A = \{1 \leqslant i_1 < i_2 < \cdots < i_k \leqslant N\}$. Define p_A to be the distribution of $X_A = \{X_{i_1}, \cdots, X_{i_k}\}$; p_A is a probability distribution on $\mathfrak{M}^k$. For each y in the support of Y let T_y be a 1-1 transformation of $\mathfrak{M}^k$ onto $\mathfrak{M}^k$, and let s_A denote the distribution of $T_Y X$. Define r_k to be the uniform distribution over $\mathfrak{M}^k$ ($r_k(x) = m^{-k}$ for each $x \in \mathfrak{M}^k$).

COROLLARY 1. *Let s_A and p_A be as defined above. Then s_A is majorized by p_A. Thus $\psi(s_A) \leqslant \psi(p_A)$ for all Schur convex functions ψ, and, in particular, $\|s_A - r_k\| \leqslant \|p_A - r_k\|$ for any symmetric norm, $\|\cdot\|$, on $\mathfrak{R}^{m^k}$.*

PROOF. The majorization of s_A by p_A follows from Lemma 2 with $n = m^k$ and $\mathfrak{X} = \mathfrak{M}^k$. The other statements are consequences of majorization. (See Lemma 1 and our remarks on Schur convex functions).

REMARK 1. Consider $Z_i = X_i + Y_i \pmod m$ $i = 1, 2, \cdots, N$, where X_i and Y_i both assume values in $\mathfrak{M} = \{0, 1, \cdots, m-1\}$. In this case X and Y play symmetric roles. It follows from Corollary 1 that if q_A denotes the distribution of $Y_A = (Y_{i_1}, \cdots, Y_{i_k})$ then s_A is majorized by q_A. Thus $\psi(s_A) \leqslant \min(\psi(p_A), \psi(q_A))$ for all Schur convex functions. Also note that this conclusion applies to *any* subset A of the index set. Thus, for all $k \leqslant N$, all k dimensional marginal distributions of Z are at least as uniform, in the sense we described, as are the corresponding distributions of X and Y.

REMARK 2. If $T_{\mathbf{Y}}\mathbf{X}$ is of the form $(T_{Y_1}X_1, \cdots, T_{Y_N}X_N)$ where each $T_{Y_i}X_i$ is a mixture of 1-1 onto transformations and X and Y are independent, then the conclusion of Corollary 1 will hold for all A. In addition, if we have an $m \times m$ matrix B with rows labeled $0, \cdots, m-1$ and columns $0, \cdots, m-1$, with each row and column containing each of the numbers $0, \cdots, m-1$, exactly once, (an $m \times m$ Latin square), then defining $T_{Y_i}X_i = B(X_i, Y_i)$ leads to $\psi(s_A) \leqslant \min(\psi(p_A), \psi(q_A))$ for all A. The additive generator, $Z_i = X_i + Y_i \pmod m$, is of this form.

REMARK 3. We briefly consider a generator proposed by Maclarin and Marsaglia [11], and discussed in Knuth [8], page 30–31. Knuth remarks that the method produces sequences with excellent randomness properties and is quite efficient in terms of computer time usage. Under this method the first k elements of $\mathbf{X}$ are used to form a table. We observe Y_1 which tells us which element of the table to choose as Z_1. We replace this element by X_{k+1}. The process is then repeatedly applied to generate the string. Suppose that a string of n numbers $Z_1, \cdots, Z_n$, is generated by this method. We artificially enlarge this set to size $n + k$ by setting Z_{n+i} equal to the entry which sits in the ith place in the table after the string of n numbers has been generated. The new string $(Z_1, \cdots, Z_{n+k})$ is thus a random permutation of $(X_1, \cdots, X_{n+k})$, induced by $\mathbf{Y}$. Since a permutation of coordinates is a 1-1 onto transformation, $\mathfrak{M}^{n+k} \to \mathfrak{M}^{n+k}$, Theorem 1 applies. Thus s, the distribution of $(Z_1, \cdots, Z_{n+k})$, is at least as uniform in our sense as is that of $(X_1, \cdots, X_{n+k})$.

In general, improving the uniformity of a joint distribution does not necessarily improve the uniformity of marginals. For example, let $p(0, 0) = p(1, 0) = .1$ and $p(0, 1) = p(1, 1) = .4$, $\Pr(Y = 0) = \Pr(Y = 1) = .5$, $T_0(i, j) = (i, j)$, $T_1(i, j) = (j, i)$, $(Z_1, Z_2) = T_Y(X_1, X_2)$. Then $s(0, 0) = .1$, $s(1, 1) = .4$ and $s(1, 0) = s(0, 1) = .25$. Then s is majorized by p and the joint distribution of (Z_1, Z_2) is more uniform on $\{0, 1\} \times \{0, 1\}$ than that of (X_1, X_2). Nevertheless X_1 is perfectly uniformly distributed while Z_1 is not.

REMARK 4. In Theorem 1 we show that $\psi(s_A) \leqslant \psi(p_A)$ for all Schur convex ψ. The Schur convex functions of greatest interest to us are distances from r_k under symmetric norms. There are other relevant Schur functions which arise from information theory considerations. If a is a probability distribution over $\mathfrak{M}^k$ then $g(a, r_k) = \Sigma_{\alpha \in \mathfrak{M}^k} a(\alpha)\log(m^k a(\alpha))$, the Kullback-Leibler information number for discriminating between a and r_k when a is true, is Schur convex; $g(a, r_k) \geqslant 0$ with equality if and only if $a = r_k$, and, in interesting ways, can be interpreted as a measure of discrepancy between a and r_k (Kullback [10]). Similarly $g(r_k, a) = \Sigma_{\alpha \in \mathfrak{M}^k} m^{-k} \log(m^{-k}/a(\alpha))$, the Kullback-Leibler information number for discriminating between a and r_k when r_k is true, is Schur convex, as is $g(a, r_k) + g(r_k, a)$, the divergence between a and r_k. Substituting these Schur convex functions into the inequality $\psi(s_A) \leqslant \psi(p_A)$, derived in Corollary 1, strengthens the assertion that s is at least as uniform as p.

4. Combining several generators. Suppose we have a sequence of independent random vectors $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n, \cdots$. We combine $\mathbf{X}_{1, A}$ and $\mathbf{X}_2$ to form a vector $\mathbf{Z}_{2, A}$, then combine $\mathbf{Z}_{2, A}$ and $\mathbf{X}_3$ to form $\mathbf{Z}_{3, A}$, etc. Assume that at each stage the transformation is of the form $\mathbf{Z}_{n, A} = T_{n, \mathbf{X}_n}(\mathbf{Z}_{n-1, A})$, a mixture of 1-1 transformations of $\mathfrak{M}^k$ onto $\mathfrak{M}^k$. Represent the transition from stage $n - 1$ to stage n by the matrix P_n, where $P_n(\alpha, \beta) = \Pr(\mathbf{Z}_{n, A} = \beta | \mathbf{Z}_{n-1, A} = \alpha)$ for $\alpha, \beta \in \mathfrak{M}^k$. Define $s_{n, A}$ to be the distribution of $Z_{n, \alpha}$. Then $s_{n-1, A} P_n = s_{n, A}$ and $s_{n, A}$ is majorized by $s_{n-1, A}$ by Theorem 1; thus, by Lemma 1, P_n is doubly stochastic. The process $\{Z_{n, A}, n = 1, 2, \cdots\}$ is thus a nontime homogeneous doubly stochastic Markov chain on the state space $\mathfrak{M}^k$. Also assume that $\min_{\alpha, \beta} P_{n, \alpha, \beta} = \Delta_n \geqslant \Delta > 0$ for all n. Define $M_n = \max_\alpha s_n(\alpha)$, and $m_n = \min_\alpha s_n(\alpha)$. We will show that $M_n - m_n \leqslant (1 - m^k \Delta)^n$ which implies that $\max_\alpha |s_n(\alpha) - m^{-k}|$ goes to zero at a geometric rate. The method employed below is well known in the theory of Markov chains. Now:

$$M_n \leqslant M_{n-1}\big(1 - (m^k - 1)\Delta_n\big) + \Delta_n(1 - M_{n-1}) = M_{n-1}\big(1 - m^k \Delta_n\big) + \Delta_n; \tag{1}$$

$$m_n \geqslant m_{n-1}\big(1 - (m^k - 1)\Delta_n\big) + \Delta_n(1 - m_{n-1}) = m_{n-1}\big(1 - m^k \Delta_n\big) + \Delta_n. \tag{2}$$

Thus, by (1) and (2), $M_n - m_n \leqslant (M_{n-1} - m_{n-1})(1 - m^k\Delta)$ and thus, by iteration, $M_n - m_n \leqslant (1 - m^k\Delta)^n$, which proves the result.

Under the weaker condition $\Sigma_1^\infty \Delta_i = \infty$ we get $\lim_{n\to\infty}(M_n - m_n) = 0$ but the convergence need not be geometric. The condition $\Sigma\Delta_i = \infty$ is not necessary for convergence of $M_n - m_n$ to zero (and thus of $s_{n,A}$ to r_k). For example, if $\Delta_i = m^{-k}$ for any i then $s_{n,A} = r_k$ for all $n \geqslant i$.

REFERENCES

[1] Berge, C. (1963). *Topological Spaces*. MacMillan, New York.

[2] Beyer, W. A., Roof, R. B. and Williamson, D. (1971). The lattice structure of multiplicative congruential pseudorandom vectors. *Math. Comp.* **25** 345–363.

[3] Coveyou, R. R. (1960). Serial correlation in the generator of pseudorandom numbers. *J. Assoc. Comput. Mach.* 72–74.

[4] Dieter, U. (1972). Statistical interdependence of pseudo-random numbers generated by the linear congruential method. In *Applications of Number Theory to Numerical Analysis*. (ed. S. K. Zaremba). Academic Press.

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

[6] Hardy, G. H., Littlewood, J. E. and Polya, G. (1952). *Inequalities, (2nd ed.)*. Cambridge Univ. Press.

[7] Hull, T. E. and Dobell, A. R. (1962). Random number generators. *SIAM Rev.* **4** 230–254.

[8] Knuth, Donald E. (1969). *The Art of Computer Programming, Volume II; Seminumerical Algorithms*. Addison-Wesley, Reading, Massachusetts.

[9] Kuipers, L. and Niederreiter, H. (1974). *Uniform Distribution of Sequences*. John Wiley and Sons, New York.

[10] Kullback, S. (1959). *Information Theory and Statistics*. John Wiley and Sons, New York.

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

[12] Marsaglia, G. (1972). The structure of linear congruential sequences. In *Applications of Number Theory to Numerical Analysis*. (S. K. Zaremba, Ed.). Academic Press.

[13] Marshall, A. W. and Olkin, I. (1977). Majorization in multivariate distributions. *Ann. Statist.* **2** 1189–1200.

[14] Meyer, H. A. (Ed.). (1956). *Symposium on Monte Carlo Methods*. John Wiley and Sons, New York.

[15] Monte Carlo Methods. (1951). N.B.S. Applied Mathematics Series No. 12, U.S. Government Printing Office.

[16] Niederreiter, H. (1976). On the distribution of pseudo-random numbers generated by the linear congruential method III. *Math. Comp.* **30** 571–597.

[17] Proschan, F. and Sethuraman, J. (1977). Schur functions in statistics. I. The preservation theorem. *Ann. Statist.* **5** 256–262.

[18] Rotenberg, A. (1960). A new pseudo-random number generator. *J. Assoc. Comput. Mach.* **7** 75–77.

[19] Strawderman, W. E. (1971). Generation and testing of pseudo-random numbers. Technical Report No. 171, Depart. Statist., Stanford Univ.

67 Tintern Lane
Scarsdale, New York 10583

Office of U.S. Naval Research
223 Old Marylebone Road
London NW1 5TH
England