QUASI-MONTE CARLO VIA LINEAR SHIFT-REGISTER SEQUENCES

Pierre L'Ecuyer and Christiane Lemieux

Département d'Informatique et de Recherche Opérationnelle
Université de Montréal, C.P. 6128, Succ. Centre-Ville
Montréal, H3C 3J7, CANADA

ABSTRACT

Linear recurrences modulo 2 with long periods have been widely used for contructing (pseudo)random number generators. Here, we use them for quasi-Monte Carlo integration over the unit hypercube. Any stochastic simulation fits this framework. The idea is to choose a recurrence with a short period length and to estimate the integral by the average value of the integrand over all vectors of successive output values produced by the small generator. We examine randomizations of this scheme, discuss criteria for selecting the parameters, and provide examples. This approach can be viewed as a polynomial version of lattice rules.

1 MONTE CARLO VS QUASI-MONTE CARLO

1.1 The Monte Carlo Method

The aim of most stochastic simulations is to estimate a mathematical expectation, and this can be put into the framework of estimating the integral of a function f over the t-dimensional unit hypercube $[0,1)^t$, namely

$$\mu = \int_{[0,1)^t} f(\mathbf{u})d\mathbf{u}. \tag{1}$$

Randomness in simulations is indeed generated from a sequence of i.i.d. $U(0,1)$ (pseudo)random variables, i.e., a random point in $[0,1)^t$ if t uniforms are generated. When t is random, one can view the number of dimensions as infinite, with only a finite subset of the random numbers being used.

The usual estimator of μ is the average value of f over a point set $P_n = \{\mathbf{u}_0, \ldots, \mathbf{u}_{n-1}\} \subset [0,1)^t$,

$$Q_n = \frac{1}{n}\sum_{i=0}^{n-1} f(\mathbf{u}_i). \tag{2}$$

The integration error is $E_n = Q_n - \mu$. In the traditional Monte Carlo (MC) method, P_n is a set of n i.i.d. uniform random points over $[0,1)^t$. Then, $E[Q_n] = \mu$ and Var $[Q_n] = \sigma^2/n$, provided that $\sigma^2 = \int_{[0,1)^s} f^2(\mathbf{u})d\mathbf{u} - \mu^2 < \infty$, in which case one has the central limit theorem: $\sqrt{n}(Q_n - \mu)/\sigma \Rightarrow N(0,1)$, so $|E_n| = O_p(\sigma/\sqrt{n})$ (regardless of t) and this error can be estimated via either the central limit theorem, or large deviations theory, or some other probabilistic method (Bratley, Fox, and Schrage 1987; Fishman 1996; Law and Kelton 1991).

Generating the points P_n requires nt random numbers (assuming that t is a finite constant) and common wisdom says that the period length of the random number generator used for that purpose should be several orders of magnitude larger than nt (e.g., L'Ecuyer 1998).

What we suggest here is the opposite: Take a small random number generator which has only n states, and let P_n be the set of all vectors of t successive output values produced by the generator, from all its initial states (i.e., over all of its cycles). If the generator is designed so that P_n covers the unit hypercube more evenly than random points, it appears plausible that Q_n could become a better approximation of μ than the Q_n obtained by random points. The idea is not new: The lattice rules proposed by Korobov (1959) are in fact a special case. The idea was also discussed by Niederreiter (1986).

1.2 Quasi-Monte Carlo

Placing the points P_n *more evenly* than at random is the basic idea of so-called *quasi-Monte Carlo* methods. A precise meaning can be given to "more evenly" by defining a measure of *discrepancy* between the discrete distribution determined by the points of P_n and the uniform distribution over $[0,1)^t$. The point set P_n is said to have *low-discrepancy* if its discrepancy measure is significantly smaller than that of a typical random point set.

There are several ways of defining a discrepancy, many of them leading to an error bound of the form

$$|E_n| \le V(f)D(P_n) \qquad \text{for all } f \in \mathcal{F}, \tag{3}$$

where $\mathcal{F}$ is a Banach space of functions f with norm $\|\cdot\|$, $V(f) = \|f-\mu\|$ measures the variability of f, and $D(P_n)$ is the discrepancy of P_n (see, e.g., Hellekalek 1998; Hickernell 1998b; Niederreiter 1992). When $D(P_n)$ is the widely-used rectangular star discrepancy $D_n^*(P_n)$, defined in terms of rectangular boxes with one corner at the origin (e.g., Niederreiter 1992), (3) is the well-known Koksma-Hlawka inequality. A popular way of constructing point sets with low discrepancy $D_n^*(P_n)$ is by constructing so-called (t,m,s)-nets, for which $D_n^*(P_n) = O(n^{-1}(\ln n)^{t-1})$ (Larcher 1998; Niederreiter 1992; Niederreiter and Xing 1998). Then, if $V(f) < \infty$, the error bound converges at the deterministic rate $O(n^{-1}(\ln n)^{t-1})$, which is asymptotically better than the probabilistic rate $O_p(n^{-1/2})$ of the MC method.

This is nice in principle, but the worst-case bounds given by the Koksma-Hlawka inequality are (almost always) practically useless, because $D_n(P_n)$ and (especially) $V(f)$ are too hard to compute and, more importantly, the error bound is typically several orders of magnitude larger than the true error and (especially for large t) much too large to be of any use. This does not mean that QMC does not work, only that the error should be estimated by other tools than the Koksma-Hlawka inequality. An alternative is to randomize P_n, say m times, independently, so that its discrepancy remains low while the m corresponding replicates of Q_n are i.i.d. unbiased estimators of μ.

1.3 Outline

In Section 2, we overview one way of constructing a point set P_n by taking all vectors of successive values produced by a linear congruential generator (LCG) and shifting all these points by a common uniform random point, modulo 1. Such a P_n is a Korobov lattice rule (Sloan and Joe 1994). In Section 3, we look at what happens if we replace the LCG by a linear feedback shift register (LFSR) (or Tausworthe) generator. This gives lattice rules in a polynomial space. Explicit expressions for the error and for the variance of the randomized estimator are given in terms of the coefficients of a Walsh series expansion of f. Based on a functional ANOVA decomposition of $\mathrm{Var}[E_n]$, we introduce, in Section 4, selection criteria for the LFSR parameters which take into account the quality of certain low-dimensional projections. These criteria are somewhat related to (but different from) those defining a (t,m,s)-net. These same criteria could also be used for selecting (pseudo)random number generators. We give specific examples of small LFSR generators that satisfy these criteria. Larcher (see Larcher 1998 and the references cited there) has also studied polynomial lattice rules over $\mathbb{F}_2$ using Walsh expansions, but from a different viewpoint: His interest was mainly in (t,m,s)-net properties and Koksma-Hlawka error bound. In Section 5, we use our LFSR point sets for one simulation example.

2 RANDOMLY SHIFTED LATTICE RULES

Consider an LCG defined by the linear recurrence

$$x_i = (ax_{i-1}) \bmod n, \qquad u_i = x_i/n,$$

for some integers $0 < a < n$. Let $P_n = \{\mathbf{u} = (u_0, \ldots, u_{t-1}) : x_0 \in \mathbb{Z}_n\}$, where $\mathbb{Z}_n = \{0, \ldots, n-1\}$, the set of all t-dimensional vectors of successive output values produced by the LCG over all of its cycles. This P_n is the intersection of a lattice L_t with the unit hypercube $[0,1)^t$. In the context of QMC, such a P_n is called a *Korobov rule*. If n is a prime and a is primitive modulo n, the LCG has one cycle of length $n-1$ and one cycle of length 1 (the absorbing state 0), so it is easy to enumerate P_n by going through the nontrivial cycle and adding the point $\mathbf{u} = (0, \ldots, 0)$.

Write the Fourier expansion of f as

$$f(\mathbf{u}) = \sum_{\mathbf{h} \in \mathbf{Z}^t} \hat{f}(\mathbf{h}) \exp(2\pi\sqrt{-1}\,\mathbf{h}\cdot\mathbf{u}),$$

with *Fourier coefficients*

$$\hat{f}(\mathbf{h}) = \int_{[0,1)^t} f(\mathbf{u}) \exp(-2\pi\sqrt{-1}\,\mathbf{h}\cdot\mathbf{u})d\mathbf{u}.$$

The integration error with the lattice rule is then (Hickernell 1996; Sloan and Joe 1994)

$$E_n = \sum_{\mathbf{0} \ne \mathbf{h} \in L_t^*} \hat{f}(\mathbf{h}) \tag{4}$$

(assuming that this series converges absolutely) where $L_t^* = \{\mathbf{h} \in \mathbb{Z}^t : \mathbf{k}\cdot\mathbf{h} \in \mathbb{Z} \text{ for all } \mathbf{k} \in L_t\}$ is the *dual* lattice to L_t.

This E_n is hard to compute in practice, but its mean square can be estimated by the following technique, called a Cranley-Patterson rotation (Cranley and Patterson 1976). Generate $\mathbf{U}$ uniformly over $[0,1)^t$ and replace each $\mathbf{u}_i$ in P_n by $\tilde{\mathbf{u}}_i = (\mathbf{u}_i + \mathbf{U}) \bmod 1$ (where the "modulo 1" reduction is coordinate-wise). The set P_n is thus replaced by $\tilde{P}_n = \{\tilde{\mathbf{u}}_0, \ldots, \tilde{\mathbf{u}}_{n-1}\}$, and Q_n and E_n by $\tilde{Q}_n$ and

$\tilde{E}_n$. One can show (Lemieux and L'Ecuyer 2000b) that $E[\tilde{E}_n] = 0$ and

$$\text{Var}\,[\tilde{E}_n] = \sum_{\mathbf{0}\neq\mathbf{h}\in L_t^*} |\hat{f}(\mathbf{h})|^2. \tag{5}$$

Equations (4) and (5) suggest a discrepancy measure of the form

$$D(P_n) = \sum_{\mathbf{0}\neq\mathbf{h}\in L_t^*} w(\mathbf{h}) \quad \text{or} \quad D(P_n) = \sup_{\mathbf{0}\neq\mathbf{h}\in L_t^*} w(\mathbf{h}), \tag{6}$$

where the (arbitrary) weights $w(\mathbf{h})$ decrease with the size of $\mathbf{h}$ in a way that corresponds to how we think the Fourier coefficients $\hat{f}(\mathbf{h})$ decrease (see, e.g., Entacher, Hellekalek, and L'Ecuyer 2000; Hickernell 1998a; Lemieux and L'Ecuyer 2000a for examples).

To estimate the error, compute m i.i.d. copies of $\tilde{Q}_n$ with the same P_n (using m independent uniform shifts $\mathbf{U}$) and compute their sample variance, which is an unbiased estimator of Var $[\tilde{Q}_n] =$ Var $[\tilde{E}_n]$.

3 LATTICE RULES IN A RING OF POLYNOMIALS OVER $\mathbb{F}_2$

3.1 LFSR Generators

We consider the linear recurrence

$$x_n = (a_1 x_{n-1} + \cdots + a_k x_{n-k}) \bmod 2 \tag{7}$$

of *order* $k > 1$, where $a_k = 1$ and $a_j \in \{0,1\}$ for each j. This sequence is purely periodic and the period length of its longest cycle is $2^k - 1$ if and only if its *characteristic polynomial*

$$P(z) = -\sum_{i=0}^{k} a_i z^{k-i} \tag{8}$$

(where $a_0 = -1$) is a primitive polynomial over $\mathbb{F}_2$, the Galois field with 2 elements (Lidl and Niederreiter 1986). Tausworthe-type linear feedback shift register (LFSR) generators evolve according to (7) and produce the output

$$u_n = \sum_{i=1}^{L} x_{ns+i-1} 2^{-i} \tag{9}$$

at step n, where the parameters s and L are positive integers. Tezuka and L'Ecuyer (1991) and L'Ecuyer (1996) give an efficient algorithm for implementing this generator when $P(z)$ is a trinomial, $P(z) = z^k - z^q - 1$, and the parameters satisfy the conditions $0 < 2q < k \le L$ and $0 < s < k - q$.

Since trinomial-based generators of this type are unsatisfactory from the theoretical viewpoint (Lindholm 1968), Tezuka and L'Ecuyer (1991) proposed composite LFSR generators defined as follows. Take J LFSR generators that satisfy the above conditions, the jth one having the characteristic polynomial $P_j(z) = z^{k_j} - z^{q_j} - 1$, so it obeys

$$\begin{aligned} x_{j,i} &= (x_{j,i-r_j} + x_{j,i-k_j}) \bmod 2, \\ u_{j,n} &= \sum_{i=1}^{L} x_{ns_j+i-1} 2^{-i}, \end{aligned}$$

where $r_j = k_j - q_j$. Let

$$\begin{aligned} u_n &= u_{1,n} \oplus \cdots \oplus u_{J,n} \\ &= \sum_{i=1}^{L} ((x_{1,ns+i-1} + \cdots + x_{J,ns+i-1}) \bmod 2)\, 2^{-i}. \end{aligned}$$

If each $P_j(z)$ is a primitive trinomial and if the k_j's are relatively prime, $\{u_n\}$ is also an LFSR generator with period length $\rho = (2^{k_1} - 1) \cdots (2^{k_J} - 1)$, and (reducible) characteristic polynomial $P(z) = P_1(z) \cdots P_J(z)$ of degree $k = k_1 + \cdots + k_J$. Specific sets of parameters, as well as implementations in the C language, are provided by L'Ecuyer (1996, 1999). The parameters given there are for $k \ge 88$ and are for MC (the cardinality of P_n is 2^k). For QMC, we need smaller values of k, ranging (say) from about 10 to 25.

3.2 Equidistribution

For a point set P_n in $[0,1)^t$ and an arbitrary set of dimensions $I = \{i_1, \ldots, i_d\} \subseteq \{1, \ldots, t\}$, let $P_n(I)$ be the projection of P_n over the d-dimensional subspace determined by I. If we partition the interval $[0,1)$ into 2^ℓ segments of length $2^{-\ell}$, this partitions the d-dimensional unit hypercube into $2^{d\ell}$ cubic boxes of equal size. If P_n has cardinality 2^k, we say that $P_n(I)$ is *d-distributed to ℓ bits of accuracy*, or (d,ℓ)-*equidistributed*, if each box of the partition contains exactly $2^{k-d\ell}$ points of $P_n(I)$. This means that if we look at the first ℓ bits of each coordinate of the points of $P_n(I)$, each of the $2^{d\ell}$ possible $d\ell$-bit strings appears exactly the same number of times. Of course, this can happen only if $d\ell \le k$. To verify the equidistribution, one can write a system of linear equations that express these $d\ell$ bits as a function of the k bits of the initial state of the recurrence, $(x_0, \ldots, x_{k-1})$: One has d-distribution to ℓ bits of accuracy if and only if the matrix of this linear transformation has full rank, $d\ell$.

L'Ecuyer (1996, 1999) computed tables of combined LFSR generators for which $P_n(I)$ is d-distributed to ℓ bits of accuracy for each I of the form $\{1, \ldots, d\}$ and for each (d,ℓ) such that $d\ell \le k$, and

$\ell \le L$, where $L = 32$ or 64 (the word size). He called such generators *maximally equidistributed* (ME).

A related property is that of a "(t, m, s)-net" (a (q, k, t)-net, in our notation), where one considers all the partitions of $[0, 1)^t$ into rectangular boxes of dimensions $2^{-\ell_1}, \ldots, 2^{-\ell_t}$ (not only cubic boxes), such that $\ell_1 + \cdots + \ell_t = k - q$ for some integer q. In our notation, P_n is a (q, k, t)-net in base 2 if for each of these partitions, each box of the partition contains exactly 2^q points. See Niederreiter (1992) or Owen (1998) for further details. If $q = 0$, this implies the ME property. The (q, k, t)-net property is much harder to check than the ME property, especially when k is large and q is small, because it involves a much larger number of partitions, i.e., building and computing the rank of a much larger number of matrices.

We propose as a compromise, in Section 4, criteria based on enriched versions of the ME property, and motivated by a variance decomposition given in Section 3.5.

The point sets P_n that correspond to LFSR generators are *dimension-stationary* (Lemieux and L'Ecuyer 2000b), in the sense that $P_n(\{i_1, \ldots, i_v\}) = P_n(\{i_1 + j, \ldots, i_v + j\})$ for all $i_1, \ldots, i_v$ and j such that $1 \le i_1 < \cdots < i_v \le t$ and $1 \le j \le t - i_v$. This property is conveniently exploited to reduce the number of sets I for which the quality of the projection $P_n(I)$ must be examined: It suffices to consider those for which $i_1 = 1$. This property does not hold in general, e.g., for common (q, k, t)-net constructions with $q > 0$, the projections $P_n(\{i_1, \ldots, i_v\})$ and $P_n(\{i_1 + j, \ldots, i_v + j\})$ often differ in quality.

3.3 Polynomial Representation and General LFSR Implementation

The LFSR generators can be interpreted as linear congruential generators in a space of polynomials. To see this, we define a one-to-one mapping between the state space $\mathbb{F}_2^k$ of the recurrence (7) and the space $\mathbb{F}_2[z]/(P)$ of polynomials of degree less than k with coefficients in $\mathbb{F}_2$: To the state $s_n = (x_n, \ldots, x_{n+k-1})$, we associate the polynomial

$$p_n(z) = \sum_{j=1}^{k} c_{n,j} z^{k-j} \tag{10}$$

where

$$\begin{pmatrix} c_{n,1} \\ c_{n,2} \\ \vdots \\ c_{n,k} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ a_1 & 1 & \ldots & 0 \\ \vdots & & \ddots & \vdots \\ a_{k-1} & \ldots & a_1 & 1 \end{pmatrix} \begin{pmatrix} x_n \\ x_{n+1} \\ \vdots \\ x_{n+k-1} \end{pmatrix} \bmod 2. \tag{11}$$

We then have (see, e.g., L'Ecuyer 1994)

$$p_n(z) = z p_{n-1}(z) \bmod (P(z), 2), \tag{12}$$

where "mod $(P(z), 2)$" means the remainder of the polynomial division by $P(z)$, with the operations on the coefficients performed in $\mathbb{F}_2$. In other words, we have an LCG in $\mathbb{F}_2[z]/(P)$, with modulus $P(z)$ and multiplier z.

In the remainder of the paper, we restrict our attention to the implementation (9) and consider the point set $P_n = \{\mathbf{u} = (u_0, \ldots, u_{t-1}) : s_0 \in \mathbb{F}_2^k\}$. The polynomial LCG (12) has a lattice structure similar to that of the usual LCG (Couture, L'Ecuyer, and Tezuka 1993; Tezuka 1995; Couture and L'Ecuyer 2000). In the case of (9), the *dual lattice* is the space $\mathcal{L}_t^*$ of multivariate polynomials $\mathbf{h}(z) = (h_1(z), \ldots, h_t(z))$, where $h_i(z) = \sum_{j=0}^{\ell-1} h_{i,j} z^j$, $h_{i,j} \in \mathbb{F}_2$, $\ell \in \mathbb{N}$, and such that $\sum_{i=1}^{t} h_i(z) z^{(i-1)s} \bmod (P(z), 2) = 0$. In a deliberate abuse of notation, we identify each polynomial $\mathbf{h}(z)$ with the integer vector $\mathbf{h} = (h_1, \ldots, h_t)$, where $h_i = \sum_{j=0}^{\ell-1} h_{i,j} 2^j \in \mathbb{N}$, so $\mathcal{L}_t^*$ can also be viewed as a space of integer vectors $\mathbf{h}$. This dual lattice plays a role in providing error and variance expressions similar to (4) and (5), as we soon explain.

3.4 Walsh Expansion

For any multivariate polynomial $\mathbf{h} = \mathbf{h}(z)$ defined as above, and for $\mathbf{u} = (u_1, \ldots, u_t)$ where $u_i = \sum_{j \ge 1} u_{i,j} 2^{-j} \in [0, 1)$ and $u_{i,j} \ne 1$ for infinitely many j, define

$$\mathbf{h} \otimes \mathbf{u} = \sum_{i=1}^{t} \sum_{j=1}^{\infty} h_{i,j-1} u_{i,j} \bmod 2.$$

The *Walsh expansion in base 2* of $f : [0, 1)^t \to \mathbb{R}$ is then (e.g., Beauchamp 1984):

$$f(\mathbf{u}) = \sum_{\mathbf{h} \in \mathbb{N}^t} \tilde{f}(\mathbf{h}) (-1)^{\mathbf{h} \otimes \mathbf{u}}, \tag{13}$$

with coefficients

$$\tilde{f}(\mathbf{h}) = \int_{[0,1)^t} f(\mathbf{u}) (-1)^{\mathbf{h} \otimes \mathbf{u}} d\mathbf{u}. \tag{14}$$

Each term in (13) represents a piecewise-constant periodic function of $\mathbf{u}$ with frequency h_i along the ith axis and amplitude $\tilde{f}(\mathbf{h})$. Each vector $\mathbf{h}$ is a *bit selector*, which picks a finite number of bits from the binary expansion of $(u_1, \ldots, u_t)$. Intuitively, the $\mathbf{h}$'s for which $\|\mathbf{h}\|_\infty = \max_{1 \le i \le t} h_i$ is small are more important because they test the most significant bits of the $\mathbf{u}_j$. The following results are not hard to prove, and they also apply to the projections $P_n(I)$ (with obvious adaptations).

Proposition 1 *One has*

$$\sum_{j=0}^{n-1}(-1)^{\mathbf{h}\otimes\mathbf{u}_j} = \begin{cases} n & \text{if } \mathbf{h}\in\mathcal{L}_t^*, \\ 0 & \text{otherwise.}\end{cases}$$

Proposition 2 (Couture, L'Ecuyer, and Tezuka 1993.) *The point set P_n is t-distributed to ℓ bits of accuracy if and only if $\mathcal{L}_t^*$ contains no vector $\mathbf{h} = (h_1, \ldots, h_t) \neq \mathbf{0}$ such that $0 \le h_i < 2^\ell$ for each i, i.e., if and only if the shortest nonzero vector $\mathbf{h}$ in $\mathcal{L}_t^*$ has length $\|\mathbf{h}\|_\infty = \sup_{1\le i\le t}|h_i| \ge 2^\ell$ (with the sup norm).*

As pointed out to us by R. Couture, the counterpart of the Cranley-Patterson rotation for polynomial lattice rules over $\mathbb{F}_2$ (i.e., LFSR point sets) is to generate a single uniform random variable $\mathbf{U}$ in $[0,1)^t$ and replace P_n by $\tilde{P}_n = \{\tilde{\mathbf{u}}_0, \ldots, \tilde{\mathbf{u}}_{n-1}\}$, where $\tilde{\mathbf{u}}_i = \mathbf{u}_i \oplus \mathbf{U}$, the bitwise exclusive-or of the binary expansions of the coordinates of $\mathbf{u}_i$ and $\mathbf{U}$. We define the random variables $\tilde{Q}_n$ and $\tilde{E}_n$ as in Section 2, but with this new $\tilde{P}_n$. Note that this randomization of P_n is much simpler than the scrambling proposed by Owen (1997b) for nets, and possesses essentially the same variance properties (the details will appear in a forthcoming paper by Couture, L'Ecuyer, and Lemieux).

Proposition 3 *One has $E[\tilde{E}_n] = 0$ and, similar to (4) and (5), the integration error with P_n and the variance with $\tilde{P}_n$ can be written as*

$$E_n = \sum_{\mathbf{0}\neq\mathbf{h}\in\mathcal{L}_t^*} \tilde{f}(\mathbf{h}) \tag{15}$$

if this series converges absolutely, and

$$\mathrm{Var}\,[\tilde{E}_n] = \sum_{\mathbf{0}\neq\mathbf{h}\in\mathcal{L}_t^*} |\tilde{f}(\mathbf{h})|^2. \tag{16}$$

if f is square-integrable.

This suggests discrepancy measures of the form (6), with L_t^* replaced by $\mathcal{L}_t^*$. The weight should be chosen in accordance with our knowledge (or intuition) of how the Walsh coefficients are likely to behave as a function of $\mathbf{h}$. Again, we can make m independent shifts and compute a confidence interval for μ from the m i.i.d. copies of $\tilde{Q}_n$.

3.5 Functional ANOVA Decomposition

We now decompose the variance of $\tilde{E}_n$ in terms of the projections determined by the subsets I of $\{1, \ldots, t\}$. This will motivate discrepancy measures based on the quality of these projections. The ANOVA decomposition of Hoeffding (e.g., Owen 1998) is

$$f(\mathbf{u}) = \sum_{I\subseteq\{1,\ldots,t\}} f_I(\mathbf{u}),$$

where $f_I(\mathbf{u}) = f_I(u_1, \ldots, u_t)$ depends only on $\{u_i,\ i \in I\}$, $\int_{[0,1)^{2t}} f_I(\mathbf{u})f_J(\mathbf{v})d\mathbf{u}d\mathbf{v} = 0$ for all $I \neq J$, $f_\phi(\mathbf{u}) \equiv \mu$, and $\int_{[0,1)^t} f_I(\mathbf{u})d\mathbf{u} = 0$ for $I \neq \phi$, where ϕ denotes the empty set. For $v > 0$, $\sum_{|I|\le v} f_I(\cdot)$ is the least mean square approximation of $f(\cdot)$ by a sum of v-dimensional (or less) functions. The variance decomposes as

$$\begin{aligned}\sigma^2 &= \mathrm{Var}\,[\tilde{E}_n] \\ &= \sum_{I\subseteq\{1,\ldots,t\}} \sigma_I^2 \\ &= \sum_{I\subseteq\{1,\ldots,t\}} \sum_{\mathbf{0}\neq\mathbf{h}\in\mathcal{L}_t^*} |\tilde{f}_I(\mathbf{h})|^2\end{aligned}$$

where (for $I \neq \phi$) σ_I^2 is the variance of $f_I(\mathbf{U})$, and the coefficient $\tilde{f}_I(\mathbf{h})$ of the Walsh expansion of f_I is 0 unless $\mathbf{h}$ satisfies: $h_j \neq 0$ if and only if $j \in I$.

For typical simulation models, a large fraction of the variance is accounted for by a relatively small number of sets I, in the sense that $\sum_{I\subseteq\mathcal{J}} \sigma_I^2$ is near σ^2 for some class $\mathcal{J}$ of cardinality much less than 2^t. The most important sets I are often those that contain successive indices, or a small number of indices that are not too far apart. This suggests discrepancy measures of the form (6), where the sum (or sup) is over a class of vectors $\mathbf{h}$ that correspond to these types of sets I. We propose such measures in the next section.

4 SPECIFIC CRITERIA AND PARAMETER SETS

Let $\mathcal{L}_t^*(I)$ denote the projection of $\mathcal{L}_t^*$ over the subspace determined by I, and let $2^{\ell^*(I)}$ be the length of the shortest nonzero vector $\mathbf{h}$ in $\mathcal{L}_t^*(I)$. We want $\ell^*(I)$ to be large. If $|I| = j$ then $\ell^*(I) \le \lfloor k/j \rfloor$. We then define

$$\Delta(d,s) = \max_{I\in S(d,s)} \left[\lfloor k/j\rfloor - \ell^*(I)\right]. \tag{17}$$

where $S(d,s) = \{I = \{i_1, \ldots, i_j\} : i_1 = 1$, and either each $i_j \le s$ and $|I| \le d$, or I contains only consecutive indices$\}$. We say that the point set P_n is ME(d,s) if $\Delta(d,s) = 0$, i.e., if it is ME and if for each $I \subseteq \{1, \ldots, s\}$ of cardinality no more than d, the projection $P_n(I)$ is also ME. Note that ME$(1,k)$ is the same as ME.

In recent papers (Owen 1997a; Larcher 1998; Hickernell 2000), it has been pointed out that the quality criterion q for (q,k,t)-nets should be generalized to a vector of parameters $(q_I)_{I\subseteq\{1,2,\ldots,t\}}$ that would measure the quality of each projection $P_n(I)$ of the net, or at least a certain number of these projections. These q_I are defined in a similar way to q, but with the restriction that each l_j defining the rectangular boxes for which the equidistribution property is checked must be at least 1 when $j \in I$ and we have $q = \max_I q_I$. Since $k-|I|+1-\ell^*(I)$ is an upper bound on q_I for our LFSR point sets, the criterion we propose can be seen as a way to construct (q,k,t)-nets for which q_I can be bounded individually whenever $I \in S(d,t)$, because $\ell^*(I)$ is known in this case.

We performed exhaustive searches over all combined LFSR generators with either two or three components whose characteristic polynomials are primitive trinomials with relatively prime degrees, and which satisfy the implementation conditions mentioned in Section 3.1, to find the best ones with respect to $\Delta(3,10)$, which also turned out to be the best ones with respect to $\Delta(4,10)$. We give the search results in Table 1, in which $\delta_{v,u}$ is such that $\Delta(d,u) = \max_{1\le v\le d}\delta_{v,u}$, and $\delta_{v,u} = \max_{I\in S'(v,u)}[\lfloor k/j\rfloor - \ell^*(I)]$, where $j = |I|$ and $S^{'}(v,u) = \{I = \{i_1,\ldots,i_j\} : i_1 = 1$, and either $i_j \le u$ and $|I| = v > 1$ or I contains only consecutive indices, if $v = 1\}$. Most of the generators in the table are ME(2,10) and the smallest value of k for which we could find an ME(3,10) generator was $k = 19$.

Table 1: Best Combined LFSRs with their $\delta_{v,10}$

k	(k,q,s)	$\delta_{1,10}$	$\delta_{2,10}$	$\delta_{3,10}$	$\delta_{4,10}$	$\Delta(4,10)$
10	(7,1,3)	0	1	2	1	2
	(3,1,2)					
12	(5,2,3)	0	1	2	2	2
	(4,1,2)					
	(3,1,1)					
14	(9,4,3)	0	2	1	1	2
	(5,2,2)					
16	(11,2,7)	2	1	2	2	2
	(5,2,2)					
19	(10,3,4)	0	1	1	2	2
	(9,4,2)					

5 A NUMERICAL EXAMPLE

For a numerical illustration, we consider the pricing of an asian option on the arithmetic average, for a single asset. We assume the Black-Scholes model for the evolution of the asset value, with risk-free appreciation rate r, volatility σ, strike price K, and expiration time T. The average is over the values at the t observation points $T-t+1,\ldots,T$. To simulate each observation of the selling price, one needs t normal random variables. To reduce the variance, one can use the selling price of the option on the *geometric* average as a control variable, as well as antithetic variates. Details about this model can be found in Lemieux and L'Ecuyer (1998).

In Table 2, we give the estimated variance reduction factors (with respect to MC) obtained by the randomly-scrambled LFSR point sets (as in Section 3.4) given in Table 1. The parameters of the option are $S(0) = 100$, $r = \ln 1.09$, $\sigma = 0.2$ and $T = 120$. We use 100 randomizations $\mathbf{U}$ to estimate the variance. When the control variable and antithetic variates are used, we call this the ACV estimator. Otherwise, we have the *naive* estimator. For Monte Carlo, we used the same total sample size $100n$ (for a fair comparison).

Table 2: Estimated Variance Reduction Factors

s	n	$K=90$	$K=100$	$K=110$
		naive estimator		
10	1024	420	210	62
	4096	3200	1600	730
	16384	22000	11000	1800
	65536	55000	13000	2300
60	1024	78	55	9.3
	4096	200	88	7.4
	16384	1100	180	41
	65536	1000	200	41
		ACV estimator		
10	1024	17	17	4.5
	4096	64	22	7.4
	16384	122	22	12
	65536	74	29	18
60	1024	16	8.0	2.2
	4096	16	8.4	2.7
	16384	14	11	1.6
	65536	30	9.5	3.1

For this problem, the LFSR point sets from Table 1 reduce the variance by factors ranging approximately between 2 and 50000 compared to MC. As expected, the reduction factors usually increase with n and decrease with t. The improvement over MC is more important with the naive estimators than with the ACV ones: This had been noted previously by Lemieux and L'Ecuyer (1998) and Lemieux and L'Ecuyer (2000a). Also, the reduction factors decrease with K: The explaination is that when K is large, the function f is zero on most of the domain $[0, 1)^t$ and thus, the good equidistribution of LFSR point sets is not very useful. In this situation, *importance sampling* is an appropriate variance reduction technique, as discussed by Glasserman, Heidelberger, and Shahabuddin (1999).

Notice that the generator used for $k = 16$ is not ME: Among ME generators for this value of k, the best value of $\Delta(4, 10)$ that could be obtained was 3 and was given by a bad projection in dimension 3 (i.e., $\delta_{3,10} = 3$). This generator turned out to be quite bad for the asian option problem, giving sometimes estimators with more variance than MC. The one from Table 1 definitely gives better estimators than the ME one and this shows that looking at projections over non-consecutive indices is important for this type of application.

The results obtained in this example are quite promising given the simplicity of the method and the fact that it is faster than MC. They also compare favorably with results obtained by randomly-shifted LCGs chosen with an equivalent criterion.

ACKNOWLEDGMENTS

This work has been supported by NSERC-Canada grant No. ODGP0110050 and FCAR Grant No. 00ER3218 to the first author, and via an FCAR-Québec scholarship to the second author. We thank Raymond Couture, who suggested the exclusive-or randomization techniques leading to Proposition 3, and François Panneton, who helped computing the parameter table of Section 4.

REFERENCES

Beauchamp, K. G. 1984. *Applications of Walsh and related Functions*. London: Academic Press.

Bratley, P, B. L. Fox, and L. E. Schrage. 1987. *A Guide to Simulation*. Second ed. New York: Springer-Verlag.

Couture, R. and P. L'Ecuyer. 2000. Lattice computations for random numbers. *Mathematics of Computation*, 69(230):757–765.

Couture, R., P. L'Ecuyer, and S. Tezuka. 1993. On the distribution of k-dimensional vectors for simple and combined Tausworthe sequences. *Mathematics of Computation*, 60(202):749–761, S11–S16.

Cranley, R. and T. N. L. Patterson. 1976. Randomization of number theoretic methods for multiple integration. *SIAM Journal on Numerical Analysis*, 13(6):904–914.

Entacher, K., P. Hellekalek, and P. L'Ecuyer. 2000. Quasi-Monte Carlo node sets from linear congruential generators. In *Monte Carlo and Quasi-Monte Carlo Methods 1998*, ed. H. Niederreiter and J. Spanier, 188–198, Berlin. Springer.

Fishman, G. S. 1996. *Monte Carlo: Concepts, Algorithms, and Applications*. Springer Series in Operations Research, New York: Springer-Verlag.

Glasserman, P, P. Heidelberger, and P. Shahabuddin. 1999. Asymptotically optimal importance sampling and stratification for pricing path dependent options. *Mathematical Finance*, 9(2):117–152.

Hellekalek, P. 1998. On the assessment of random and quasi-random point sets. In *Random and Quasi-Random Point Sets*, ed. P. Hellekalek and G. Larcher, volume 138 of *Lecture Notes in Statistics*, 49–108. New York: Springer.

Hickernell, F. J. 1996. Quadrature error bounds with applications to lattice rules. *SIAM Journal on Numerical Analysis*, 33:1995–2016.

Hickernell, F. J. 1998a. A generalized discrepancy and quadrature error bound. *Mathematics of Computation*, 67:299–322.

Hickernell, F. J. 1998b. Lattice rules: How well do they measure up? In *Random and Quasi-Random Point Sets*, ed. P. Hellekalek and G. Larcher, volume 138 of *Lecture Notes in Statistics*, 109–166. New York: Springer.

Hickernell, F. J. 2000. What affects accuracy of quasi-Monte Carlo quadrature? In *Monte Carlo and Quasi-Monte Carlo Methods 1998*, ed. H. Niederreiter and J. Spanier, 16–55, Berlin. Springer.

Korobov, N. M. 1959. The approximate computation of multiple integrals. *Dokl. Akad. Nauk SSSR*, 124:1207–1210. in Russian.

Larcher, G. 1998. Digital point sets: Analysis and applications. In *Random and Quasi-Random Point Sets*, ed. P. Hellekalek and G. Larcher, volume 138 of *Lecture Notes in Statistics*, 167–222. New York: Springer.

Law, A. M. and W. D. Kelton. 1991. *Simulation Modeling and Analysis*. Second ed. New York: McGraw-Hill.

L'Ecuyer, P. 1994. Uniform random number generation. *Annals of Operations Research*, 53:77–120.

L'Ecuyer, P. 1996. Maximally equidistributed combined Tausworthe generators. *Mathematics of*

Computation, 65(213):203–213.

L'Ecuyer, P. 1998. Random number generation. In *Handbook of Simulation*, ed. J. Banks, 93–137. Wiley. chapter 4.

L'Ecuyer, P. 1999. Tables of maximally equidistributed combined LFSR generators. *Mathematics of Computation*, 68(225):261–269.

Lemieux, C and P. L'Ecuyer. 1998. Efficiency improvement by lattice rules for pricing asian options. In *Proceedings of the 1998 Winter Simulation Conference*, ed. D. J. Medeiros, E. F. Watson, J. S. Carson, and M. S. Manivannan, 579–586, Piscataway, NJ. IEEE Press.

Lemieux, C. and P. L'Ecuyer. 2000a. A comparison of Monte Carlo, lattice rules and other low-discrepancy point sets. In *Monte Carlo and Quasi-Monte Carlo Methods 1998*, ed. H. Niederreiter and J. Spanier, 326–340, Berlin. Springer.

Lemieux, C. and P. L'Ecuyer. 2000b. Selection criteria for lattice rules and other low-discrepancy point sets. *Mathematics and Computers in Simulation*. To appear.

Lidl, R. and H. Niederreiter. 1986. *Introduction to Finite Fields and Their Applications*. Cambridge: Cambridge University Press.

Lindholm, J. H. 1968. An analysis of the pseudo-randomness properties of subsequences of long m-sequences. *IEEE Transactions on Information Theory*, IT-14(4):569–576.

Niederreiter, H. 1986. Multidimensional numerical integration using pseudorandom numbers. *Mathematical Programming Study*, 27:17–38.

Niederreiter, H. 1992. *Random Number Generation and Quasi-Monte Carlo Methods*. volume 63 of *SIAM CBMS-NSF Regional Conference Series in Applied Mathematics*. Philadelphia: SIAM.

Niederreiter, H and C. Xing. 1998. Nets, (t, s)-sequences, and algebraic geometry. In *Random and Quasi-Random Point Sets*, ed. P. Hellekalek and G. Larcher, volume 138 of *Lecture Notes in Statistics*, 267–302. New York: Springer.

Owen, A. 1997a. Scrambling Sobol and Niederreiter-Xing points. Technical report, Department of Statistics, Stanford University, Palo Alto, California, U.S.A.

Owen, A. B. 1997b. Monte Carlo variance of scrambled equidistribution quadrature. *SIAM Journal on Numerical Analysis*, 34(5):1884–1910.

Owen, A. B. 1998. Latin supercube sampling for very high-dimensional simulations. *ACM Transactions of Modeling and Computer Simulation*, 8(1):71–102.

Sloan, I. H and S. Joe. 1994. *Lattice Methods for Multiple Integration*. Oxford: Clarendon Press.

Tezuka, S. 1995. *Uniform Random Numbers: Theory and Practice*. Norwell, Mass.: Kluwer Academic Publishers.

Tezuka, S and P. L'Ecuyer. 1991. Efficient and portable combined Tausworthe random number generators. *ACM Transactions on Modeling and Computer Simulation*, 1(2):99–112.

AUTHOR BIOGRAPHIES

PIERRE L'ECUYER is a professor in the "Département d'Informatique et de Recherche Opérationnelle", at the University of Montreal. He received a Ph.D. in operations research in 1983, from the University of Montréal. He obtained the *E. W. R. Steacie* grant from the Natural Sciences and Engineering Research Council of Canada for the period 1995–97. His main research interests are random number generation, efficiency improvement via variance reduction, sensitivity analysis and optimization of discrete-event stochastic systems, and discrete-event simulation in general. He is an Area Editor for the *ACM Transactions on Modeling and Computer Simulation*. More details at: `http://www.iro.umontreal.ca/~lecuyer`, where several of his recent research articles are available on-line.

CHRISTIANE LEMIEUX is currently a Ph.D. student in the "Département d'Informatique et de Recherche Opérationnelle", at the University of Montréal. She works on efficiency improvement by lattice rules. She completed her M.Sc. in mathematics (actuarial science) at the same university in 1996.

Published in the *Proceedings of the 1999 Winter Simulation Conference*, Dec. 1999, 632–639.

Corrections to Table 1 and update of the references: May 2000.