Pierre L'Ecuyer
Département d'informatique et de recherche opérationnelle, Université de Montréal, Canada

Christiane Lemieux
Department of Mathematics and Statistics, University of Calgary, Canada

ABSTRACT: Linear recurrences in a finite field or ring (i.e., modulo some positive integer m) are widely used for constructing (pseudo)random number generators with huge period lengths. Special cases include the linear congruential, multiple recursive, and Tausworthe random number generators. It is well known that the set of all vectors of t successive values produced by such a generator, from all possible seeds, has a lattice structure whose geometrical properties can be analyzed explicitly. If one takes a small period length, this set can also be used as a point set P_n for quasi-Monte Carlo integration in t dimensions, perhaps with a proper randomization (e.g., a random shift modulo 1 of the entire point set) for unbiasedness and error estimation purposes.

The aim of this paper is to review some recent ideas and results on these randomized quasi-Monte Carlo methods with lattice rules. We provide an explicit expression for the variance of the corresponding estimator in terms of the coefficients of the Fourier expansion of the integrand. We discuss certain aspects of selection criteria for such point sets (i.e., for selecting the parameters of the linear recurrence) from a practical perspective. We favor criteria that take into account the quality of the low-dimensional projections of P_n, and we explain why. These ideas apply to the construction of both quasi-Monte Carlo point sets and random number generation. We give examples of search results for good parameters, based on our proposed criteria. We also give concrete numerical examples that illustrate the effectiveness of the randomized quasi-Monte Carlo approach.

1 INTRODUCTION

Monte Carlo simulation aims at estimating an integral of the form

$$\mu = E[f(\mathbf{U})] = \int_{[0,1)^t} f(\mathbf{u}) d\mathbf{u} \qquad (1)$$

where $f : [0,1)^t \to \mathbb{R}$ can be a very complicated function, $\mathbf{u} = (u_1, \ldots, u_t) \in [0,1)^t$, and $\mathbf{U}$ is a t-dimensional vector of i.i.d. $U(0,1)$ random variables. Here, we assume that the number of dimensions t is deterministic and finite for ease of exposition, but the development applies more generally as well.

For the classical *Monte Carlo method* (MC), $\mathbf{U}$ is *simulated* by a vector $\mathbf{u}$ of t output values from a (pseudo)random number generator (RNG). This is repeated n times, yielding the point set $P_n = \{\mathbf{u}_0, \ldots, \mathbf{u}_{n-1}\} \subset [0,1)^t$, and μ is estimated by the average

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

If we assume that the $\mathbf{u}_i$ are i.i.d. uniform over $[0,1)^t$, and that

$$\sigma^2 = \int_{[0,1)^t} f^2(\mathbf{u}) d\mathbf{u} - \mu^2 < \infty, \qquad (3)$$

then we have $E[Q_n] = \mu$, $\mathrm{Var}[Q_n] = \sigma^2/n$, and $\sqrt{n}(Q_n - \mu)/\sigma \to N(0,1)$ in distribution as $n \to \infty$ by the central limit theorem, so the error converges as $|Q_n - \mu| = O_p(\sigma/\sqrt{n})$ regardless of t [Law & Kelton 2000, Fishman 1996].

The idea of the *Quasi-Monte Carlo* (QMC) method is to construct the point set P_n *more evenly* distributed over $[0,1)^t$ than typical random points, in the hope of reducing the estimation error $|Q_n - \mu|$. The non-evenness is assessed by measures of *discrepancy* $D(P_n)$ between the distribution of P_n and the uniform distribution over $[0,1)^t$. Several *low-discrepancy sequences* of points $P_\infty = \{\mathbf{u}_0, \mathbf{u}_1, \ldots\}$ have been constructed so that $D^*(P_n) = O(n^{-1}(\ln n)^t)$ when t is fixed and $n \to \infty$, where $P_n = \{\mathbf{u}_0, \ldots, \mathbf{u}_{n-1}\}$ and $D^*(P_n)$ is the rectangular star discrepancy of P_n. Then, the Koksma-Hlawka inequality yields the worst-

where $V(f)$ is the total variation of f in the sense of Hardy and Krause. Similar results can also be obtained with more general definitions of the discrepancy and corresponding definitions for $V(f)$. For the details, see [Niederreiter 1992, Niederreiter & Xing 1998, Sobol' 1998, Larcher 1998, Hellekalek 1998, Hickernell 2000] and the references cited there. When $V(f) < \infty$, this provides a convergence rate of $O(n^{-1}(\ln n)^t)$, which is better than the MC rate $O_p(n^{-1/2})$ *asymptotically*.

However, this theory has practical limitations. Firstly, as soon as t exceeds half a dozen or so, $O(n^{-1}(\ln n)^t)$ becomes smaller than $O_p(n^{-1/2})$ only for excessively large (impractical) values of n. Secondly, the discrepancy $D^*(P_n)$ for a given n and $t > 3$ (say) is usually too difficult to compute. Thirdly, for most functions f encountered in simulation applications, $V(f) = \infty$ or $V(f)$ cannot be computed and the Koksma-Hlawka worst-case error bound would be much much larger than the true error anyway.

The main problem is that for large (or even moderate) t, it becomes impossible to well cover the unit hypercube $[0,1)^t$ with points, because there is too much space to fill up when the dimension is large.

The good news, on the other hand, is that "low-discrepancy" point sets seem to effectively reduce the integration error by a significant amount, compared with MC, even when t is very large, for certain classes of problems. Why is that?

In the remainder of this paper, we provide some explanations and use them to define selection criteria for quasi-random point sets P_n for a fixed n. We concentrate on point sets constructed as the intersection of an *integration lattice* with the unit hypercube, i.e., for which Q_n is a *lattice rule*. The quality of P_n is then measured by the quality of its *lattice structure* and by the quality of the lattice structure of its projections over lower-dimensional subspaces. We define figures of merit that take into account the quality of a selected subset of projections and give examples of rules chosen by optimizing the figures of merit.

In practice, we apply a random shift modulo 1 to P_n before computing the estimator, and repeat this m times, independently, in order to obtain unbiased estimators of μ and of the variance of the estimator of μ. We obtain an exact expression for the variance of the resulting estimator, in terms of the squared coefficients of the Fourier expansion of f. Concrete numerical examples illustrate the effectiveness of the approach.

A detailed and expanded version of this paper is available in [L'Ecuyer & Lemieux 2000].

2 FUNCTIONAL ANOVA

We said that constructing a point set P_n that covers $[0,1)^t$ very well is too difficult when t is large. So we will change that goal to the following: Construct P_n so that for selected subsets of coordinates $I \subset \{1, \ldots, t\}$, the projection of P_n over the subspace determined by the coordinates that belong to I, denoted $P_n(I)$, is very evenly distributed over this $|I|$-dimensional subspace. The reason why this often suffices to have a small integration error is that a small number of terms in the functional ANOVA decomposition of f, which we now explain, often capture a large fraction of the variance σ^2.

The *functional ANOVA decomposition* [Hoeffding 1948, Efron & Stein 1981, Owen 1998] writes $f(\mathbf{u})$ as a sum of orthogonal functions:

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

$$= \mu + \sum_{i=1}^{t} f_{\{i\}}(u_i) + \sum_{i,j=1}^{t} f_{\{i,j\}}(u_i, u_j) + \cdots$$

where $f_I(\mathbf{u}) = f_I(u_1, \ldots, u_t)$ depends only on $\{u_i, \ i \in I\}$, $f_\phi(\mathbf{u}) \equiv \mu$ (ϕ is the empty set),

$$\int_{[0,1)^t} f_I(\mathbf{u})d\mathbf{u} = 0 \quad \text{for } I \neq \phi,$$

$$\int_{[0,1)^{2t}} f_I(\mathbf{u})f_J(\mathbf{v})d\mathbf{u}d\mathbf{v} = 0 \quad \text{for } I \neq J,$$

and the variance σ^2 decomposes as

$$\sigma^2 = \sum_{I \subseteq \{1,\ldots,t\}} \sigma_I^2 = \sum_{\phi \neq I \subseteq \{1,\ldots,t\}} \int_{[0,1)^t} f_I^2(\mathbf{u})d\mathbf{u}.$$

The best mean-square approximation of $f(\cdot)$ by a sum of d-dimensional (or less) functions is $\sum_{|I| \leq d} f_I(\cdot)$.

When the approximation is good for small d, we say that f has a *low effective dimension* in the superposition sense. High-dimensional functions with low effective dimension are frequent in simulation applications. Often, the most important sets I are those that contain successive indices, or a small number of nearby indices. Sometimes, one may *change* f to make this happen [Caflisch & Moskowitz 1995, Caflish, Morokoff, & Owen 1997, Owen 1998, Fox 1999]. This suggests that the point sets P_n should be chosen on the basis of the quality of the distribution of the points over the

well. We will define our selection criteria on this basis.

2.1 LATTICE RULES

An *integration lattice* is a discrete subset of the real space $\mathbb{R}^t$, that contains $\mathbb{Z}^t$ (the integer vectors), and that can be expressed as

$$L_t = \left\{ \mathbf{v} = \sum_{j=1}^{t} z_j \mathbf{v}_j \mid \text{ each } z_j \in \mathbb{Z} \right\}, \quad (5)$$

where $\mathbf{v}_1, \ldots, \mathbf{v}_t$ are linearly independent vectors in $\mathbb{R}^t$ forming a *basis* of the lattice. The *dual lattice* is $L_t^* = \{ \mathbf{h} \in \mathbb{R}^t : \mathbf{h} \cdot \mathbf{v} \in \mathbb{Z} \text{ for all } \mathbf{v} \in L_t \}$, where $\mathbf{h} \cdot \mathbf{v}$ denotes the standard inner product between the vectors $\mathbf{h}$ and $\mathbf{v}$. A *lattice rule* is an integration method that approximates μ by Q_n in (2), using the node set $P_n = L_t \cap [0, 1)^t$, where n is the cardinality of P_n, called the *order* of the rule. For more details about lattice rules, see [Hickernell 1998, L'Ecuyer & Lemieux 2000, Sloan & Joe 1994].

Here, we concentrate on the simple special case of *Korobov rules*, for which P_n can be expressed as

$$P_n = \{ (j/n)\mathbf{v} \bmod 1 : 0 \le j < n \} \quad (6)$$

for $\mathbf{v} = (1, a, \ldots, a^{t-1})$, for some integer a, where $0 < a < n$ and $\gcd(a, n) = 1$. In this case, P_n is in fact equal to the set of all vectors of t successive output values produced by the linear congruential generator (LCG) defined by the recurrence

$$x_i = (a x_{i-1}) \bmod n, \qquad u_i = x_i/n, \quad (7)$$

from all possible initial states x_0. That is, $P_n = \{ (u_0, \ldots, u_{t-1}) : 0 \le x_0 < n \}$. If we also assume that n is a prime number and a is a primitive element modulo n, then the LCG (7) has maximal period length $n - 1$, and the point set P_n can be generated efficiently by running (7) from $x_0 = 1$ (say), taking the first $n - 1$ overlapping vectors of t successive values, and adding the zero vector. Details on LCGs can be found in [Knuth 1998, L'Ecuyer 1998].

For a given subset of coordinates $I = \{i_1, \ldots, i_d\} \subseteq \{1, \ldots, t\}$, the projection $L_t(I)$ of L_t over the d-dimensional subspace determined by I is also a lattice, with dual lattice $L_t^*(I)$. One has $P_n(I) = L_t(I) \cap [0, 1)^t$.

The point set P_n defined by (6) is called *fully projection-regular* if for any non-empty $I \subseteq \{1, \ldots, t\}$, $P_n(I)$ contains as many distinct points as P_n. It is called *dimension-stationary* if $P_n(\{i_1, \ldots, i_d\}) = P_n(\{i_1 + j, \ldots, i_d + j\})$ for all

the quality of $P_n(I)$ only for the sets I for which $i_1 = 1$. These two properties do not hold for typical QMC point sets. However, we have [L'Ecuyer & Lemieux 2000, Lemieux & L'Ecuyer 2000]:

Proposition 1 *A Korobov lattice rule is dimension stationary, and it is fully projection-regular if* $\gcd(a, n) = 1$.

The lattice structure of $L_t(I)$ means that its points are contained in families of equidistant parallel hyperplanes. Choose the family of hyperplanes that are farthest apart among those that cover all the points, and let $d_t(I)$ be the distance between these hyperplanes. We want $d_t(I)$ as small as possible, because large values mean thick slices of empty space, whereas small values mean that the points $P_n(I)$ provide a good and even coverage of the unit hypercube. It turns out that $d_t(I) = 1/\ell_I$ where ℓ_I is the Euclidean length of the shortest nonzero vector in the dual lattice $L_t^*(I)$. Tight upper bounds are available on the best possible value of ℓ_I as a function of n and of the dimension $d = |I|$. They have the form $\ell_I \le \ell_d^*(n) = c_d n^{1/d}$ with explicit constants c_d [Conway & Sloane 1999, L'Ecuyer 1999b]. One can thus normalize each ℓ_I to a figure of merit $\ell_I / \ell_{|I|}^*(n)$ lying between 0 and 1 (the larger the better), which allows comparing the quality of the projections for sets I of different dimensions. The effect of the normalization is to be more demanding on $d_t(I)$ for the projections over lower-dimensional subspaces, i.e., when $|I|$ is small.

The next step is to take the worst-case value of that figure of merit over a selected family of sets I that are considered important. In view of Proposition 1, we can restrict ourselves to the sets I whose first index is $i_1 = 1$. We propose the following general worst-case figure of merit. For arbitrary integers $d \ge 1$ and $t_1 \ge \cdots \ge t_d \ge d$, define

$$M_{t_1, \ldots, t_d} = \min \left[\min_{2 \le s \le t_1} \frac{\ell_s}{\ell_s^*(n)}, \min_{2 \le s \le d, I \in S(s, t_s)} \frac{\ell_I}{\ell_{|I|}^*(n)} \right] \quad (8)$$

where $S(s, t_s) = \{ I = \{i_1, \ldots, i_s\} : 1 = i_1 < \cdots < i_s \le t_s \}$. This measure takes into account the projections over s successive dimensions for all $s \le t_1$, and the projections over no more than d non-successive dimensions that are not too far apart. The rules selected on the basis of this criterion are designed to be uniformly good for a wide range of problems. In contrast, lattice rules provided in earlier tables, e.g., in [Sloan & Joe 1994], are typically chosen by considering only the set $I = \{1, \ldots, t\}$, and are not necessarily good for problems where

3

ordinates of the form $I = \{1, \ldots, s\}$.

To select rules for general usage, we suggest using (8) with d equal to 4 or 5, and t_s decreasing with s. There is a compromise to be made when choosing d and the t_s's: $M_{t_1, \ldots, t_d}$ is a worst case over $(t_1 - d) + \sum_{s=2}^{d} \binom{t_s - 1}{s-1}$ projections, and when this number is too large, the worst-case measure becomes less discriminatory because eventually there are always some projections that are bad. It is also more costly to compute. One could also replace the minimum in (8) by a weighted average of some sort. We adopted the minimum in part because it makes the computations easier and the search for the best rules more efficient. Note that the same criterion could be used as well for selecting linear congruential or multiple recursive pseudorandom number generators. It has been used in the past [L'Ecuyer 1996, L'Ecuyer 1999a, L'Ecuyer 1999b], but only for $d = 1$.

Table 1 gives the best multipliers a that are primitive elements modulo n, in terms of the criteria M_{32} and $M_{32,24,12,8}$, for certain prime numbers n close to powers of 2. For each of the two criteria, for each n, we give an optimal multiplier a, its optimal criterion value (marked with a *), and the value of the other criterion. Note that for $n = 131071$, for example, the best LCG with respect to M_{32} is very bad with respect to $M_{32,24,12,8}$ (certain low-dimensional projections over non-successive indices are very bad).

Table 1. Best LCGs with respect to M_{32} and $M_{32,24,12,8}$.

n	a	M_{32}	$M_{32,24,12,8}$
4093	219	0.66150*	0.13642
	1516	0.39382	0.28399*
8191	1716	0.64854*	0.05243
	5130	0.50777	0.30676*
16381	665	0.65508*	0.15291
	4026	0.50348	0.29139*
32749	9515	0.67356*	0.29319
	14251	0.50086	0.32234*
65521	2469	0.63900*	0.17455
	8950	0.55678	0.34307*
131071	29803	0.66230*	0.03137
	28823	0.44439	0.33946*

3 RANDOM SHIFTS

When P_n is deterministic, the integration error is also deterministic and hard to estimate. To be able to estimate the error, we will use independent random shifts of P_n, as in [Cranley &

$\tilde{P}_n = \{\tilde{\mathbf{u}}_0, \ldots, \tilde{\mathbf{u}}_{n-1}\}$ and $\tilde{Q}_n = (1/n) \sum_{i=0}^{n-1} f(\tilde{\mathbf{u}}_i)$. Repeat this m times, independently, with the same P_n, thus obtaining m i.i.d. copies of $\tilde{Q}_n$, denoted $X_1, \ldots, X_m$. Let $\bar{X} = (X_1 + \cdots + X_m)/m$ and $S_x^2 = \sum_{j=1}^{m} (X_j - \bar{X})^2/(m-1)$.

Proposition 2 *We have $E[\bar{X}] = \mu$ and $E[S_x^2] = \text{Var}[X_j] = m \ \text{Var}[\bar{X}]$.*

We now have an unbiased estimator of μ whose variance can be estimated. From this, we can compute a confidence interval for μ. We hope that with a good choice of P_n, $\text{Var}[X_j]$ will be smaller than $\text{Var}[Q_n]$, the variance of the standard MC estimator when the $\mathbf{u}_i$ are i.i.d. uniform.

Let

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

be the Fourier expansion of f, with Fourier coefficients

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

where $\mathbf{Z}^t$ is the set of all t-dimensional integer vectors.

Proposition 3 *If $\sigma^2 < \infty$, with the MC method,*

$$\text{Var}[Q_n] = \frac{\sigma^2}{n} = \frac{1}{n} \sum_{\mathbf{0} \neq \mathbf{h} \in \mathbf{Z}^t} |\hat{f}(\mathbf{h})|^2. \qquad (9)$$

For a randomly shifted lattice rule,

$$\text{Var}[\tilde{Q}_n] = \sum_{\mathbf{0} \neq \mathbf{h} \in L_t^*} |\hat{f}(\mathbf{h})|^2. \qquad (10)$$

A similar result was given in [Tuffin 1998] under much stronger conditions. The result also stands if we replace f, σ^2, L_t^*, and Q_n by f_I, σ_I^2, $L_t^*(I)$, and $Q_n(I)$, respectively, for $I \neq \phi$: $Q_n(I)$ and $\tilde{Q}_n(I)$ become estimators of $E[f_I(\mathbf{U})] = 0$. Since L_t^* contains exactly $1/n$ of the points of $\mathbf{Z}^t$, the proposition says that the randomly shifted lattice rule reduces the variance compared with MC if and only if the "average" squared Fourier coefficients are smaller over L_t^* than over $\mathbf{Z}^t$. For typical well-behaved problems, the squared Fourier coefficients tend to decrease quickly with the size of $\mathbf{h}$. This suggests (heuristically) that we should avoid having small vectors $\mathbf{h}$ in the dual lattice $L_t^*(I)$ for the sets I deemed important. This is exactly what our criterion (8) is doing!

There are other ways of randomizing P_n, e.g., based on stratification or latin hypercube sampling. Some of them guarantee a variance reduction, but destroy the lattice structure, and do not perform as well empirically according to our experience [Lemieux 2000, L'Ecuyer & Lemieux 2000].

4

This example is taken from [Avramidis & Wilson 1996]. A *stochastic activity network* (SAN) is represented by a directed acyclic graph whose set of arcs $\mathcal{A}$ correspond to *activities*. Each activity $k \in \mathcal{A}$ has a random duration V_k with distribution function $F_k(\cdot)$. The *network completion time* T is the length of the longest path from the source to the sink, which are two special nodes in the network. Such networks are used for planning and scheduling in large complex projects. Suppose we are interested in estimating $\mu = P[T \leq x]$ for a given x. To apply MC, we generate $\{V_k = F_k^{-1}(U_k),\ k \in \mathcal{A}\}$ where the U_k's are i.i.d. $U(0,1)$, we compute the length of the longest path by a standard algorithm [Hillier & Lieberman 1990], and we observe whether or not $T \leq x$. Here, the function f is simply the indicator of the event $\{T \leq x\}$ and the dimension t is the cardinality of $\mathcal{A}$.

The following conditional Monte Carlo (CMC) method has been proposed for this model [Avramidis & Wilson 1996]. Select a set $\mathcal{L} \subseteq \mathcal{A}$ such that each directed path j from the source to the sink contains exactly one activity l_j from $\mathcal{L}$. Then, generate (by simulation) only $\{V_k,\ k \notin \mathcal{L}\}$, and estimate μ by the conditional probability $P[T \leq x \mid \{V_k,\ k \notin \mathcal{L}\}]$. The dimension is thus reduced from $|\mathcal{A}|$ to $|\mathcal{A} \setminus \mathcal{L}|$ and the variance is provably reduced.

For both methods, one can replace the vectors of random numbers $(U_1, \ldots, U_t)$ by a randomly-shifted quasi-random point set P_n as in the previous section.
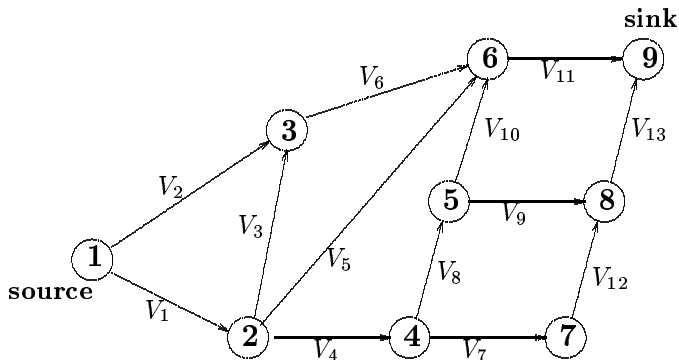

Figure 1. Example of a SAN, taken from Avramidis and Wilson (1996)

We performed experiments with the network shown in Figure 1, with the same set $\mathcal{L}$ and the same probability laws of the activity durations as in [Avramidis & Wilson 1998], Section 4.1. This network has 13 activities. The set $\mathcal{L}$ contains the

We tried standard MC, CMC, the randomly shifted lattice rules (LR), and LR combined with CMC. For LR, we used the rules of Table 1, selected via $M_{32,24,12,8}$, with the values of n given in Table 2 and with $m = 100$ random shifts to estimate the variance. For MC and CMC, we made mn i.i.d. replications, so that the total number of simulation runs was the same for all methods. Table 2 gives the (estimated) factor by which the variance is divided, for each method, compared with standard MC. The combination of LR with CMC (last line) is a clear winner and the gain increases with n. For example, with $n = 65521 \approx 2^{16}$, it takes approximately 3000 times fewer simulation runs with LR+CMC than with standard MC to obtain an estimation with the same precision. The efficiency is thus improved by (roughly) a factor of 3000. We obtained similar results for the other network presented in [Avramidis & Wilson 1996].

Table 2. Estimated variance reduction factors for the SAN example

Method	t	4093	16381	65521
MC	13	1	1	1
LR	13	6	4	24
CMC	8	4	4	4
LR+CMC	8	268	839	3086

(table header has spanning label n over the last three columns)

4.2 *Pricing Asian options*

Consider the pricing of an Asian option on the arithmetic average, for a single asset whose value follows a geometric Brownian motion. If the value of the asset at time u is $S(u)$, one has

$$\frac{S(u_2)}{S(u_1)} = \exp\left[(r - \sigma^2/2)(u_2 - u_1) + \sigma\sqrt{u_2 - u_1}Z\right]$$

where Z is a $N(0,1)$ random variable independent of what happens outside the time interval $[u_1, u_2]$, r is the risk-free appreciation rate, and σ is the volatility. Details about this model can be found, e.g., in [Duffie 1996]. The final value of the option, discounted to time 0, is given by

$$e^{-rT} \max\left(0, \frac{1}{t}\sum_{i=1}^{t} S(t_i) - K\right),$$

where $t_i = iT/t$, and K (the strike price), T (the expiration date), and t are constants.

The values of $S(t_i)$ can be generated by generating t standard normals, so the problem is t-dimensional. Instead of generating $S(t_1), S(t_2), \ldots$

This *Brownian bridge technique* (BB) [Caflisch & Moskowitz 1995] does not reduce the dimension nor the variance for MC, but makes the first few random numbers more important, and thus pushes most of the variance σ^2 into the σ_I^2's for which I contains the first few coordinates, in the functional ANOVA decomposition. This should help when we use the lattice rules.

A well-known variance reduction method for this problem is to use the payoff of the option on the *geometric* average, whose expectation has a closed-form analytic formula, as a control variable (CV) [Kemna & Vorst 1990, Lemieux & L'Ecuyer 1998].

Table 3. Estimated variance reduction factors w.r.t. MC for the Asian-option example

Method	n		
	4093	16381	65521
MC	1	1	1
LR	6	7	12
CV	376	378	381
CV+LR	703	620	597
BB+CV+LR	2488	4876	4958

Table 3 reports the variance reduction factors of LR, CV, CV+LR, and BB+CV+LR, compared with MC, for an example with $\sigma = 0.3$, $r = 0.05$, $K = 55$, $S(0) = 50$, $T = 1$, and $t = 64$. This is an integral in 64 dimensions. We used the same lattice rules as in the previous example, and the same numbers of replications. The combination BB+CV+LR divides the variance by a factor of nearly 5000 compared with MC when $n = 65521$.

5   CONCLUSIONS

We gave a brief overview of some recent ideas and results on randomized lattice rules for variance reduction in Monte Carlo integration. Further results and details can be found in [Lemieux 2000, L'Ecuyer & Lemieux 2000, Hellekalek & Larcher 1998].

Our proposed criterion $M_{t_1,\ldots,t_d}$ for lattice rules is imperfect, but it is convenient and it provides rules that seem to work well in practice. Alternative criteria may consider more subsets I and take a weighted sum instead of a worst case (minimum) as in (8). The weights can be chosen so that the sets I with large size or span are not taken into account unless they are really very bad. Non-Euclidean norms can also be used [Entacher, Hellekalek, & L'Ecuyer 2000, Hickernell

Quasi-Monte Carlo is usually associated with low-discrepancy point sets and sequences such as the so-called (t, m, s)-nets and the sequences of Halton, Sobol', Faure, and Niederreiter, which are motivated mainly by the worst-case error bound provided by the Koksma–Hlawka inequality (4). Lattice rules have also been justified by such worst-case error bounds [Niederreiter 1992, Sloan & Joe 1994]. Our approach here is to use their randomized versions as variance reduction tools instead. This appears more practical, in particular because it provides unbiased mean and variance estimators.

Analogs of lattice rules in a space of polynomials, where the LCG (7) is replaced (essentially) by a linear feedback shift register or Tausworthe-type recurrence, have also been proposed and studied recently [Larcher 1998, L'Ecuyer & Lemieux 1999, Lemieux 2000], and are still under development.

There is also work in progress on *embedded lattice sequences*, where a sequence of lattices with node sets $\{P_{n_i}, i \geq 1\}$ is defined so that n_i divides n_{i+1} (e.g., $n_{i+1} = 2n_i$), $P_{n_i} \subset P_{n_{i+1}}$ for each i, and each node set P_{n_i} is of good quality [Hickernell, Hong, L'Ecuyer, & Lemieux 2000]. This permits one to increase the value of n in a lattice rule without having to start all over again, i.e., the value of n need not be fixed in advance. One stops adding points when the error estimate is deemed small enough.

Finally, a good lattice rule *for a given problem* is one whose dual lattice contains none of the vectors $\mathbf{h}$ for which $|\hat{f}(\mathbf{h})|^2$ is large for this particular problem. Thus, ideally, the choice of the rule should depend on the problem. This suggests *adaptive* lattice sequences, where the choice of the next lattice in the sequence is based on estimates of certain squared Fourier coefficients. This seems like an interesting topic for further investigation.

REFERENCES

Avramidis, A. N. & J. R. Wilson (1996). Integrated variance reduction strategies for simulation. *Operations Research* 44:327–346.

Avramidis, A. N. & J. R. Wilson (1998). Correlation-induction techniques for estimating quantiles in simulation experiments. *Operations Research* 46(4):574–591.

Caflisch, R. E. & B. Moskowitz (1995). Modified Monte Carlo methods using quasi-random sequences. In

York, pp. 1–16. Springer-Verlag.

Caflish, R. E., W. Morokoff, & A. Owen (1997). Valuation of mortgage-backed securities using Brownian bridges to reduce effective dimension. *The Journal of Computational Finance* 1(1):27–46.

Conway, J. H. & N. J. A. Sloane (1999). *Sphere Packings, Lattices and Groups* (3rd ed.). Grundlehren der Mathematischen Wissenschaften 290. New York: Springer-Verlag.

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

Duffie, D. (1996). *Dynamic Asset Pricing Theory* (second ed.). Princeton University Press.

Efron, B. & C. Stein (1981). The jackknife estimator of variance. *Annals of Statistics* 9:586–596.

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

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

Fox, B. L. (1999). *Stategies for Quasi-Monte Carlo.* Boston, MA: Kluwer Academic.

Hellekalek, P. (1998). On the assessment of random and quasi-random point sets. See Hellekalek & Larcher (1998), pp. 49–108.

Hellekalek, P. & G. Larcher (eds.) (1998). *Random and Quasi-Random Point Sets*, Volume 138 of *Lecture Notes in Statistics*. New York: Springer.

Hickernell, F. J. (1998). Lattice rules: How well do they measure up? See Hellekalek & Larcher (1998), pp. 109–166.

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

Hickernell, F. J., H. S. Hong, P. L'Ecuyer, & C. Lemieux (2000). Extensible lattice sequences for quasi-Monte Carlo quadrature. *SIAM Journal on Scientific Computing.* To appear.

Hillier, F. S. & G. J. Lieberman (1990). *Introduction to Operations Research* (Fifth ed.). McGraw-Hill.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distributions. *Annals of Mathematical Statistics* 19:293–325.

Kemna, A. G. Z. & A. C. F. Vorst (1990). A pricing method for options based on average asset values. *Journal of Banking and Finance* 14:113–129.

Knuth, D. E. (1998). *The Art of Computer Programming, Volume 2: Seminumerical Algorithms* (Third ed.). Reading, Mass.: Addison-Wesley.

Larcher, G. (1998). Digital point sets: Analysis and applications. See Hellekalek & Larcher (1998), pp. 167–222.

Law, A. M. & W. D. Kelton (2000). *Simulation Modeling and Analysis* (Third ed.). New York: McGraw-Hill.

L'Ecuyer, P. (1996). Combined multiple recursive random number generators. *Operations Research* 44(5):816–822.

L'Ecuyer, P. (1999a). Good parameters and implementations for combined multiple recursive random number generators. *Operations Research* 47(1):159–164.

L'Ecuyer, P. (1999b). Tables of linear congruential generators of different sizes and good lattice structure. *Mathematics of Computation* 68(225):249–260.

L'Ecuyer, P. & C. Lemieux (1999). Quasi-Monte Carlo via linear shift-register sequences. In *Proceedings of the 1999 Winter Simulation Conference*, pp. 336–343. IEEE Press.

L'Ecuyer, P. & C. Lemieux (2000). Variance reduction via lattice rules. *Management Science.* To appear.

Lemieux, C. (2000). *L'utilisation de règles de réseau en simulation comme technique de réduction de la variance.* Ph. D. thesis, Université de Montréal.

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

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

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. & C. Xing (1998). Nets, (t, s)-sequences, and algebraic geometry. See Hellekalek & Larcher (1998), pp. 267–302.

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. & S. Joe (1994). *Lattice Methods for Multiple Integration.* Oxford: Clarendon Press.

Sobol', I. M. (1998). On quasi-Monte Carlo integration. *Mathematics and Computers in Simulation* 47:103–112.

Tuffin, B. (1998). Variance reduction order using good lattice points in Monte Carlo methods. *Computing* 61:371–378.