# On Selection Criteria for Lattice Rules and Other Quasi-Monte Carlo Point Sets

Christiane Lemieux and Pierre L'Ecuyer

*Département d'Informatique et de Recherche Opérationnelle*
*Université de Montréal, C.P. 6128, Succ. Centre-Ville*
*Montréal, H3C 3J7, Canada*
`e-mail: lecuyer@iro.umontreal.ca`

---

**Abstract**

We define new selection criteria for lattice rules for quasi-Monte Carlo integration. The criteria examine the projections of the lattice over subspaces of small or successive dimensions. Their computation exploits the *dimension-stationarity* of certain lattice rules, and of other low-discrepancy point sets sharing this property. Numerical results illustrate the usefulness of these new figures of merit.

---

## 1 Background on Monte Carlo and Quasi-Monte Carlo

We want to estimate the integral of a function $f$ over the $s$-dimensional unit hypercube $[0,1)^s$, namely

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

by the average value of $f$ over the point set $P_n = \{\mathbf{u}_0, \ldots, \mathbf{u}_{n-1}\} \subset [0,1)^s$,

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

[1] This work has been supported by NSERC-Canada grant No. ODGP0110050 to the second author and by scholarships from NSERC-Canada and FCAR-Québec to the first author. We thank the 2 referees, whose comments helped clarifying the presentation.

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 $[0,1)^s$. The idea of Quasi-Monte Carlo (QMC) is to take the point set $P_n$ more evenly distributed than random points.

Essentially all practical stochastic simulations can be put into the framework of estimating $\mu$ for a given (complicated) $f$, sometimes with a random $s$. Indeed, randomness in simulations is typically generated by transforming (an imitation of) a sequence of i.i.d. $U(0,1)$ random variables, i.e., a random point in $[0,1)^s$ for some $s$. When $s$ is random, one can view the number of dimensions as infinite with only a finite subset of the random numbers being used.

In the MC method, $Q_n$ is an unbiased estimator of $\mu$ with variance $\mathrm{Var}[Q_n] = \sigma^2/n$, where $\sigma^2 = \int_{[0,1)^s} f^2(\mathbf{u})d\mathbf{u} - \mu^2$. If $f$ is square-integrable (i.e., $\sigma^2 < \infty$) one has the central limit theorem: $\sqrt{n}(Q_n - \mu)/\sigma \Rightarrow N(0,1)$, so $|E_n| = O_p(\sigma/\sqrt{n})$. Compared with traditional numerical integration methods, MC requires only a weak condition on $f$, its error $E_n$ can be estimated in a probabilistic sense (e.g., by computing a variance estimator and using the central limit theorem to compute a confidence interval), it can deal with arbitrarily high dimensions $s$, and the convergence of $E_n$ to 0 is at rate $O_p(1/\sqrt{n})$ regardless of $s$.

By placing the points *more evenly* than at random, QMC tries to improve on the $O_p(1/\sqrt{n})$ convergence rate. For a precise definition of "more evenly", we need a well-defined measure of non-uniformity of $P_n$, i.e., a measure of *discrepancy* between the discrete distribution determined by the points of $P_n$ and the uniform distribution over $[0,1)^s$. A *low-discrepancy point set* $P_n$ is one for which the discrepancy measure is significantly smaller than that of a typical random point set. There are several ways of defining the discrepancy, and for a large subclass of these definitions, one can obtain a general bound on the absolute error of the form

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

where $\mathcal{F}$ is a Banach space of functions $f$ with norm $\|\cdot\|$, $V(f) = \|f - \mu\|$ measures the variability of the function $f$, and $D(P_n)$ measures the discrepancy of $P_n$. Hickernell [7] explains how to define a triple $(\mathcal{F}, D, V)$ by choosing a reproducing-kernel Hilbert space $\mathcal{F}$. In this setup, the upper bound (3) is tight in the sense that equality is attained for some worst-case function in $\mathcal{F}$. A popular choice for $D(P_n)$ is the rectangular star discrepancy $D_n^*(P_n)$, defined in terms of rectangular boxes with one corner at the origin (see, e.g., [17] for details). Then, (3) becomes the celebrated Koksma-Hlawka inequality.

Worst-case deterministic bounds as in (3) are nice in principle, but in practice they are almost always useless for the following reasons: (1) Many popular measures of discrepancy (e.g., the rectangular star discrepancy) are practically too hard to compute in general; (2) $V(f)$ is often infinite or is practically impossible to compute; (3) the error is typically much less than the worst-case upper bound, which is often too large to be of practical use. QMC often beats MC in practice, but one must look beyond (3) to understand why.

In the next section of this paper, we look at lattice rules and their randomization by a uniform rotation modulo 1. We provide explicit expressions for the error and for the variance of the randomized estimator. In Section 3, based on a functional ANOVA decomposition of $\mathrm{Var}[E_n]$, we motivate new criteria which take into account the quality of certain low-dimensional projections, and we introduce the notion of a dimension-stationary point set. In Section 4, we propose our new criteria and give specific examples of lattice rules found by maximizing these criteria, and examples of lattice rules that are excellent with respect to projections over successive dimensions, but very bad for projections over certain pairs of non-successive dimensions. In Section 5, we use these rules for two simulation examples, and show that using a criterion that looks at the projections over non-successive dimensions sometimes makes a big difference. These new criteria could also be used for selecting (pseudo)random number generators, as in [12].

## 2 Randomly-Shifted Lattice Rules

For a general point set $P_n$, by using the Fourier expansion of $f$,

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

with *Fourier coefficients*

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

and summing the integration errors of the different terms of this expansion, assuming that the Fourier series of $f$ converges absolutely, one obtains

$$E_n = \frac{1}{n} \sum_{\mathbf{0}\neq\mathbf{h}\in\mathbb{Z}^s} \hat{f}(\mathbf{h}) \sum_{i=0}^{n-1} \exp(2\pi\sqrt{-1}\,\mathbf{h}\cdot\mathbf{u}_i). \tag{4}$$

Examples of discrepancies yielding error bounds (3) via (4) can be found in [3,5,7,8,16,20].

We now suppose that $P_n$ is the intersection of a lattice $L_s$ with the unit hypercube $[0,1)^s$. Then, $Q_n$ is called a *lattice rule* with *node set* $P_n$, and (4) becomes [6,20]

$$E_n = \sum_{\mathbf{0} \neq \mathbf{h} \in L_s^*} \hat{f}(\mathbf{h}) \tag{5}$$

where $L_s^* = \{\mathbf{h} \in \mathbb{Z}^s : \mathbf{k} \cdot \mathbf{h} \in \mathbb{Z} \text{ for all } \mathbf{k} \in L_s\}$ is the *dual* lattice to $L_s$. In view of the expression for $E_n$, the lattice rule can be selected to minimize a discrepancy measure of the form

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

where the $w(\mathbf{h})$ are arbitrary weights decreasing with the size of $\mathbf{h}$ in accordance with how the Fourier coefficients $\hat{f}(\mathbf{h})$ are believed to decrease (e.g., $w(\mathbf{h}) = 1/\|\mathbf{h}\|$ for some definition of the norm $\|\cdot\|$ [16]).

A widely used class of lattice rules are the *Korobov rules*, for which $P_n = \{\mathbf{u}_0, \ldots, \mathbf{u}_{n-1}\}$ where $\mathbf{u}_i = (i\mathbf{v}/n) \bmod 1$, $\mathbf{v} = (1, a, a^2, \ldots, a^{s-1})$, and $a$ is a positive integer. They have the advantage that if $n$ is a prime number and $a$ is a primitive element modulo $n$, $P_n$ becomes the set of all overlapping vectors of successive values produced by the linear congruential pseudorandom number generator (LCG) with modulus $n$ and multiplier $a$, over its entire period length, plus the zero vector. In other words, the points $\mathbf{u}_i$ are easily generated via the recurrence $x_i = (ax_{i-1}) \bmod n$, $u_i = x_i/n$, $\mathbf{u}_i = (u_i, \ldots, u_{i+s-1})$, for $i > 0$, where $x_1 = 1$, and adding $\mathbf{x}_0 = \mathbf{0}$. For general values of $n$ and $a$, $P_n$ is the set of all overlapping vectors of successive values produced by the LCG over all of its cycles, i.e., from all initial values $x_0 \in \{0, \ldots, n-1\}$. For more about LCGs and their lattice structure, we refer the reader to [9,13] and other references therein.

Given the difficulty of using the deterministic error bounds, we turn ourselves to *probabilistic estimates*. The idea is to randomize $P_n$ so that $Q_n$ becomes an unbiased estimator of $\mu$ without losing its low-discrepancy property, and repeat the randomization a certain number of times (independently) in order to obtain an unbiased estimator of the variance. One randomization approach is the *random shift modulo 1*, also called Cranley-Patterson rotation [2]. It works as follows. Generate one point $\mathbf{U}$ uniformly over $[0,1)^s$ 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). Let $\tilde{P}_n = \{\tilde{\mathbf{u}}_0, \ldots, \tilde{\mathbf{u}}_{n-1}\}$. The value $\tilde{Q}_n$ of $Q_n$ computed with the shifted point set $\tilde{P}_n$ turns out to be an unbiased estimator of $\mu$ (because each $f(\tilde{\mathbf{u}}_i)$ is). Let $\tilde{E}_n = \tilde{Q}_n - \mu$. 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 $\mathrm{Var}[Q_n]$. This technique applies to any low-discrepancy point set. Under fairly general conditions [8], one has

$E[D^2(\tilde{P}_n)] = D^2[P_n]$ and $\mathrm{Var}[\tilde{E}_n] = E[\tilde{E}_n^2] = O(D^2[P_n])$, so it is possible to have $\mathrm{Var}[\tilde{E}_n] = O(n^{-2}(\ln n)^{2s})$ by choosing one of the many low-discrepancy point sets for which $D(P_n) = O(n^{-1}(\ln n)^s)$ [17,5]. Tuffin [21] gives stronger conditions under which $\mathrm{Var}[\tilde{E}_n] = O(n^{-2\alpha}(\ln n)^{2\alpha(s-1)+1}(\ln \ln n)^{(2\alpha-1)(s-1)})$ where $\alpha > 1$ indicates the smoothness of $f$.

Our aim here is not to bound the variance but to estimate it and (hopefully) reduce it compared with MC. The next proposition compares the variance of the MC estimator to that of a randomly-shifted lattice rule. It says that the lattice rule beats MC if and only if the average of the squared Fourier coefficients is smaller over $\mathbf{h} \in L_s^* \setminus \{\mathbf{0}\}$ than over $\mathbb{Z}^s \setminus \{\mathbf{0}\}$. The proof can be found in [14]. A different proof of (8) is given in the proof of Theorem 2 of [21], but under the much stronger condition that $f$ has an absolutely convergent Fourier expansion. For typical real-life simulation problems, the latter condition rarely holds although $f$ is usually square-integrable (i.e., the MC estimator has finite variance).

**Proposition 1** *For $f$ square-integrable, with the MC method, one has*

$$\mathrm{Var}[E_n] = \frac{\sigma^2}{n} = \frac{1}{n} \sum_{\mathbf{0} \neq \mathbf{h} \in \mathbb{Z}^s} |\hat{f}(\mathbf{h})|^2. \tag{7}$$

*For a randomly shifted lattice rule, one has*

$$\mathrm{Var}[\tilde{E}_n] = \sum_{\mathbf{0} \neq \mathbf{h} \in L_s^*} |\hat{f}(\mathbf{h})|^2. \tag{8}$$

The lattice rule does not necessarily reduce the variance compared with MC. However, a common assumption is that the squared Fourier coefficients tend to decrease with the norm of $\mathbf{h}$ (with different possibilities for the choice of norm). This brings us back to a discrepancy measure as in (6). If $w(\mathbf{h}) = 1/\|\mathbf{h}\|_2$, the inverse Euclidean norm of $\mathbf{h}$, then minimizing the second $D(P_n)$ (with the sup) is equivalent to maximizing the Euclidean distance between the origin and the nearest nonzero point in $L_s^*$. This criterion is identical to the *spectral test* which is commonly used to rate LCGs [3–5,9].

Lattice rules are typically chosen for a *fixed* $s$ (i.e., different rules for different $s$), based on the minimization of some figure of merit of the form (6) [20]. Tables of rules that are good *uniformly* for a range of values of $s$ are given in [13], for several values of $n$ that are either powers of 2 or primes close to powers of 2. In the next two sections, we motivate and propose more stringent criteria than those of [13], also based on the spectral test but (in contrast to [13]) taking into account the quality of the projections onto low-dimensional subspaces determined by non-successive coordinates. An ANOVA decomposition suggests why these projections are important.

## 3 Quality Criteria Based on Projections over Subspaces

The Hoeffding functional ANOVA decomposition (e.g., [19]) writes $f$ as a sum of orthogonal functions that depend on different subsets of the coordinates:

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

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

$$\sigma^2 = \sum_{I \subseteq \{1,\ldots,s\}} \sigma_I^2 = \sum_{I \subseteq \{1,\ldots,s\}} \sum_{\mathbf{0} \neq \mathbf{h} \in \mathbb{Z}^s} |\hat{f}_I(\mathbf{h})|^2,$$

and for a randomly shifted lattice rule, one has

$$\mathrm{Var}[\tilde{E}_n] = \sum_{I \subseteq \{1,\ldots,s\}} \sum_{\mathbf{0} \neq \mathbf{h} \in L_s^*} |\hat{f}_I(\mathbf{h})|^2, \tag{9}$$

where for $I \neq \phi$, $\sigma_I^2$ is the variance of $f_I(\mathbf{U})$ when $\mathbf{U} : U[0,1)^s$, the $\hat{f}_I(\mathbf{h})$ are the coefficients of the Fourier expansion of $f_I$, and $\hat{f}_I(\mathbf{h}) = 0$ whenever the components of $\mathbf{h}$ do not satisfy: $h_j \neq 0$ if and only if $j \in I$. In this sense, each $\mathbf{h}$ corresponds to a given $I$ (and each non-empty $I$ to infinitely many $\mathbf{h}$).

We say that $f$ has effective dimension $\leq t$ in the *truncation sense* [1,19] if $\sum_{I \subseteq \{1,\ldots,t\}} \sigma_I^2$ is near $\sigma^2$, in the *superposition sense* [1,19] if $\sum_{|I| \leq t} \sigma_I^2$ is near $\sigma^2$, and in the *successive-dimensions* sense if $\sum_{I \subseteq \{i,\ldots,i+t-1\},\, 1 \leq i \leq s-t+1} \sigma_I^2$ is near $\sigma^2$. The first definition means that $f$ is *almost* $t$-dimensional, while the others mean, in a different sense, that $f$ is almost a sum of $t$-dimensional functions. High-dimensional functions that have low effective dimension are frequent in simulation applications. In many cases, the most important sets $I$ are those that contain either successive indices, or a small number of indices that are not too far apart. This fact combined with the expression (9) for $\mathrm{Var}[\tilde{E}_n]$ suggests discrepancy measures of the form (6), but where the sum (or the sup) is restricted to those $\mathbf{h}$ that belong to the subspaces determined by those sets $I$ that are considered important. We propose selection criteria along these lines.

For a point set $P_n = \{\mathbf{u}_0, \ldots, \mathbf{u}_{n-1}\}$ and a set of dimensions $I \subseteq \{1, \ldots, s\}$, define $P_n(I)$ as the projection of $P_n$ over the $|I|$-dimensional subspace determined by the dimensions in $I$. We want $P_n(I)$ to be well distributed for a specified class of sets $I$.

Certain point sets $P_n$ have the property that $P_n(I)$ is the same for several $I$, and this can be exploited to reduce the number of sets $I$ for which the quality of the projection $P_n(I)$ must be examined. We say that $P_n$ is *dimension-stationary* (DS) if $P_n(\{i_1, \ldots, i_t\}) = P_n(\{i_1 + j, \ldots, i_t + j\})$ for all $i_1, \ldots, i_t$ and $j$ such that $1 \le i_1 < \cdots < i_t \le s$ and $1 \le j \le s - i_t$. One way to construct a DS point set is as follows. Consider a recurrence of the form $\xi_i = \tau(\xi_{i-1})$ and $u_i = g(\xi_i)$ for $\tau : \Xi \to \Xi$, $g : \Xi \to [0, 1)$, $\xi_0 \in \Xi$, and $\Xi$ a finite set of cardinality $n$.

**Proposition 2** *If $\tau$ is invertible on $\Xi$ (i.e., each cycle determined by the recurrence is purely periodic), then $P_n = \{\mathbf{u} = (g(\xi_0), \ldots, g(\xi_{s-1})) : \xi_0 \in \Xi\}$ is a DS point set.*

**PROOF.** Since the mapping $\xi_0 \to \xi_i = \tau^i(\xi_0)$ is invertible (and therefore one-to-one), one has $P_n(\{i_1, \ldots, i_t\}) = \{\mathbf{u} = (g(\xi_{i_1 - 1}), \ldots, g(\xi_{i_t - 1})) : \xi_0 \in \Xi\} = \{(g(\xi_{i_1+j-1}), \ldots, g(\xi_{i_t+j-1})) : \xi_j \in \Xi\} = \{(g(\xi_{i_1+j-1}), \ldots, g(\xi_{i_t+j-1})) : \xi_0 \in \Xi\} = P_n(\{i_1 + j, \ldots, i_t + j\})$ for $1 \le j \le s - i_t$.

Purely periodic recurrences are widely used to construct (pseudo)random number generators such as the LCGs, multiple recursive generators, linear feedback shift register generators, etc. [10,11,17], usually with a very large state space $\Xi$ (e.g., often $n > 2^{100}$). Any of these techniques can be used with a smaller $n$ to construct DS QMC point sets $P_n$. In particular, the Korobov rules described earlier are DS provided that $\gcd(a, n) = 1$. The important advantage of DS point sets is that their projections $P_n(I)$ depend only on the *spacings* between the successive indices in $I$, so several projections $P_n(I)$ are identical, and this makes easier the task of constructing a good $P_n$ in terms of the quality of the projections. This property does not hold in general for QMC point sets, e.g., for typical $(t, m, s)$-net constructions.

## 4 Specific Choices of Lattice Rules

We now propose specific selection criteria for Korobov lattice rules (i.e., LCGs for QMC) based on the quality of certain projections $P_n(I)$. Let $L_I$ be the $|I|$-dimensional lattice generated by $P_n(I)$, and $L_I^*$ be the corresponding dual lattice, i.e., the projection of $L_s^*$ over the subspace determined by the indices in $I$. We measure the quality of $P_n(I)$ by the Euclidean length $\ell_I$ of the shortest nonzero vector $\mathbf{h}$ in $L_I^*$. This length is normalized by an absolute upper bound $\ell_t^*(n) = c_t n^{1/t}$ where $t = |I|$ and the $c_t$'s are constants given in [13]. One always has $0 \le \ell_I / \ell_t^*(n) \le 1$, and we prefer values close to 1. With this normalization, we ask for a larger $\ell_I$ when $I$ contains a smaller number of dimensions, i.e, when $t$ is smaller.

If $P_n$ is DS, one has $\ell_{\{i_1,\ldots,i_t\}} = \ell_{\{1, i_2-i_1+1, \ldots, i_t-i_1+1\}}$, and in particular $\ell_I = \ell_t$ whenever $I$ is comprised of $t$ successive coordinates, so it always suffices to restrict our attention to the sets $I$ whose first coordinate is 1. We define the worst-case figure of merit

$$M_{t,k} = \min \left[ \min_{2\le s\le t} \ell_s/\ell_s^*(n), \ \min_{I\in S(k)} \ell_I/\ell_{|I|}^*(n) \right], \tag{10}$$

where $S(k) = \{I = \{i_1, \ldots, i_j\} : |I| \le k,\ i_1 = 1 \text{ and } i_j \le t\}$. For $k = 1$, one has $M_{t,k} = M_t = \min_{2\le s\le t} \ell_s/\ell_s^*(n)$, the figure of merit that was used in [13] to compute tables of lattice rules, for $t = 8, 16, 32$. Clearly, $M_{t,k}$ decreases with $k$, because the set $S(k)$ increases (quickly) with $k$. Note that $1/M_{t,k}$ is a special case of (6) (right side) with $w(\mathbf{h}) = \ell^*_{|I(\mathbf{h})|}/\ell_I$ if $I(\mathbf{h}) \in S(k)$ or $I(\mathbf{h})$ is comprised of $t$ or less successive coordinates, and $w(\mathbf{h}) = 0$ otherwise, where $I(\mathbf{h})$ is the set of nonzero coordinates of $\mathbf{h}$. The criterion $M_{t,k}$ is based on the idea of having good point sets for problems with reasonably low effective dimension in a combination of the successive-dimensions sense (left quantity in brackets) and superposition sense (right quantity in brackets).

For a given $n$, let $M_{t,k}^*$ be the largest value of $M_{t,k}$ obtained from a search over all primitive element modulo $n$. There are cases where $M_{t,k}/M_{t,k}^*$ is small even though $M_{t,k-1}/M_{t,k-1}^*$ is close or equal to 1. This happens quite often for $k = 2$, but more and more rarely when $k$ increases (e.g., very rarely for $k > 4$). When it happens, according to our experience, the randomized QMC estimator $\tilde{Q}_n$ based on the corresponding lattice tends to perform poorly. We give examples below. Also, the time to perform searches based on $M_{t,k}$ increases exponentially with $k$, so when $k > 4$ or 5 it becomes practically impossible to do such searches unless $n$ and $t$ are small. This suggests computing tables of coefficients based on $M_{t,k}$, for $2 \le k \le 4$, instead of $k = 1$. Table 1 gives examples of exhaustive search results for the best multipliers $a$ that are primitive element modulo $n$, for a given prime $n$, with $t = 8$ and $k \le 4$. Stars (*) indicate the best $a$ found with respect to the $M_{8,k}$ in the corresponding column. For each $a$ in the table, we also give the value of $M_{8,k}$ for $1 \le k \le 4$.

For $n = 13331$ and 131071, the best $a$ with respect to $M_8$ performs poorly with respect to $M_{8,k}$ for $k \ge 2$. The point sets constructed with these $a$'s have bad projections over some pairs of dimensions $(i, j)$, $j - i \le 7$. Note that the best $a$'s with respect to $M_{8,3}$ are also good with respect to the other values of $k$. On the other hand, $M_{8,k} \ll M_{8,k}^*$ for the best $a$ with respect to $M_{8,k-1}$ is more frequent for $k = 3$ than for $k = 4$, and still more frequent for $k = 2$. Thus, using $M_{8,3}$ as a criterion seems to be a good compromise. In the next section, we compare the estimators based on lattice rules selected via $M_{8,3}$ and

Table 1
Best $a$'s with respect to $M_{8,k}$, $1 \leq k \leq 4$, for different values of $n$

| $n$ | $a$ | $M_{8,4}$ | $M_{8,3}$ | $M_{8,2}$ | $M_8$ |
|---|---|---|---|---|---|
| 13331 | 3210 | 0.42675* | 0.42675 | 0.44687 | 0.52761 |
| | 9147 | 0.30309 | 0.48263* | 0.53077 | 0.53077 |
| | 1993 | 0.20234 | 0.20234 | 0.65613* | 0.65613 |
| | 7033 | 0.09706 | 0.09706 | 0.09706 | 0.71544* |
| 32749 | 307 | 0.38533* | 0.39577 | 0.41208 | 0.41208 |
| | 8801 | 0.32481 | 0.46549* | 0.46549 | 0.49395 |
| | 3042 | 0.27955 | 0.27985 | 0.65164* | 0.65164 |
| | 219 | 0.22793 | 0.22793 | 0.54037 | 0.71802* |
| 65521 | 332 | 0.38979* | 0.41988 | 0.53533 | 0.56644 |
| | 15409 | 0.22299 | 0.46670* | 0.48026 | 0.48026 |
| | 6060 | 0.22910 | 0.24005 | 0.61743* | 0.61743 |
| | 17364 | 0.28788 | 0.33148 | 0.39281 | 0.70713* |
| 131071 | 9290 | 0.41693* | 0.47516 | 0.47631 | 0.58448 |
| | 38429 | 0.34233 | 0.47986* | 0.47986 | 0.48814 |
| | 52344 | 0.17004 | 0.17004 | 0.66695* | 0.66838 |
| | 43165 | 0.02425 | 0.02425 | 0.02425 | 0.70941* |

$M_8$ on two simulation problems.

## 5 Simulation Examples

*A Polynomial Function*

We take the same example as in [3]. The polynomial used therein has an effective dimension equal to $\alpha = 2$ or $\alpha = 3$ in the superposition sense, but not in the truncation sense (independently of the dimension $s$). We therefore expect the criterion $M_{t,3}$ to be more appropriate than $M_t$. Here, the lattice rules are not randomly shifted since an exact formula for $\mu$ can be computed for this problem. The numbers given in Table 2 are the mean relative errors obtained from 20 i.i.d. replications, as in [3] for $\alpha = 3$. We compare the MC estimator with the lattice rules (LR) estimators derived from the best $a$ with respect to both $M_{8,3}$ and $M_8$.

As it turns out, the best $a$'s with respect to $M_{8,3}$ generally outperform the best with respect to $M_8$. In fact, the latter sometimes do *worse* than MC when their $M_{8,3}$ is too small (e.g., for $n = 13331$ and $n = 131071$, when $s > 10$). Compared with MC, the error is reduced by a factor of at least 10 with LR when $M_{8,3}$ is used. When the best $a$ with respect to $M_8$ performs reasonably well with respect to $M_{8,3}$, both LR estimators have a comparable error. Note that since $s$ goes up to 30 for this problem, it would be preferable

Table 2
Mean relative errors, $\alpha = 3$

| | $n$ | $s = 5$ | $s = 10$ | $s = 15$ | $s = 20$ | $s = 25$ | $s = 30$ |
|---|---|---|---|---|---|---|---|
| MC | 13331 | 6.49e-3 | 4.61e-3 | 4.05e-3 | 3.59e-3 | 2.39e-3 | 3.15e-3 |
| LR, $a$=9147 | 13331 | 2.43e-4 | 4.28e-4 | 4.06e-4 | 3.37e-4 | 1.63e-4 | 2.34e-4 |
| LR, $a$=7033 | 13331 | 7.51e-4 | 3.84e-3 | 5.87e-3 | 6.04e-3 | 4.76e-3 | 4.30e-3 |
| MC | 32749 | 3.14e-3 | 2.80e-3 | 2.08e-3 | 2.54e-3 | 1.71e-3 | 1.70e-3 |
| LR, $a$=8801 | 32749 | 2.65e-4 | 3.75e-4 | 4.35e-4 | 5.38e-4 | 6.45e-4 | 6.82e-4 |
| LR, $a$=219 | 32749 | 1.02e-3 | 6.40e-4 | 4.63e-4 | 4.09e-4 | 4.12e-4 | 4.31e-4 |
| MC | 65521 | 2.59e-3 | 1.98e-3 | 1.88e-3 | 1.37e-3 | 1.00e-3 | 1.23e-3 |
| LR, $a$=15409 | 65521 | 1.18e-4 | 2.59e-5 | 1.69e-5 | 3.22e-5 | 5.91e-5 | 7.73e-5 |
| LR, $a$=17364 | 65521 | 1.39e-4 | 1.86e-5 | 1.39e-5 | 3.98e-5 | 2.00e-5 | 5.53e-5 |
| MC | 131071 | 1.94e-3 | 1.48e-3 | 1.21e-3 | 9.28e-4 | 8.29e-4 | 7.72e-4 |
| LR, $a$=38429 | 131071 | 1.79e-4 | 2.00e-4 | 1.55e-4 | 1.03e-4 | 6.92e-5 | 8.96e-5 |
| LR, $a$=43165 | 131071 | 1.10e-3 | 2.48e-3 | 2.19e-3 | 1.77e-3 | 1.57e-3 | 1.39e-3 |

to choose the multipliers $a$ on the basis of $M_{30,3}$ instead of $M_{8,3}$. Performing extensive searches for such large values of $t$ is on our agenda.

*Pricing Asian Options*

Consider the (less artificial) problem of pricing 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 $\sigma$, strike price $K$, and expiration time $T$. The average is over the values at the $s$ observation points $T - s + 1, \ldots, T$. To simulate each observation of the selling price, one needs $s$ 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 [15].

In the following table, we look at the estimated variance reduction factors (with respect to MC) obtained by the shifted lattice rules based on the multipliers $a$ that optimize $M_{8,3}$ and $M_8$. The parameters of the option are $S(0) = 100$, $r = \ln 1.09$, $\sigma = 0.2$ and $T = 120$. We use 100 random shifts $\mathbf{U}$ to estimate the variance of the lattice rules estimator. When both previously mentioned variance reduction techniques are used, we call this the $ACV$ estimator. Otherwise, we have the *naive* estimator. For MC, we used the same total sample size $100n$ (for a fair comparison).

Here again, we see that choosing $a$ on the basis of $M_8$ can lead to bad results when one of the projections $P_n(I)$ is poorly distributed for some pair or triple $I$. The criterion $M_{8,3}$ avoids such rules and provides estimators that significantly reduce the variance compared with MC for this problem, even in 60 dimensions

Table 3
Estimated variance reduction factors

| $s$ | $n$ | $a$ | $K=90$ | $K=100$ | $K=110$ | $K=90$ | $K=100$ | $K=110$ |
|---|---|---|---|---|---|---|---|---|
| | | | naive estimator | | | ACV estimator | | |
| 10 | 13331 | 9147 | 1863.6 | 1246.1 | 566.8 | 28.7 | 10.0 | 9.0 |
| | 13331 | 7033 | 315.9 | 46.0 | 21.4 | 0.7 | 0.5 | 0.2 |
| | 32749 | 8801 | 3602.8 | 2512.4 | 912.1 | 35.7 | 19.2 | 8.3 |
| | 32749 | 219 | 3052.7 | 1395.5 | 546.7 | 7.3 | 6.6 | 5.3 |
| | 65521 | 15409 | 7851.1 | 4492.0 | 1504.3 | 44.5 | 25.4 | 9.8 |
| | 65521 | 17364 | 3807.8 | 2930.5 | 1024.7 | 41.1 | 13.5 | 7.3 |
| 60 | 13331 | 9147 | 238.5 | 64.3 | 15.8 | 9.9 | 6.5 | 2.4 |
| | 13331 | 7033 | 27.7 | 5.5 | 2.1 | 0.5 | 0.5 | 0.7 |
| | 32749 | 8801 | 267.3 | 80.8 | 16.2 | 12.5 | 6.4 | 2.5 |
| | 32749 | 219 | 234.5 | 58.0 | 21.8 | 9.4 | 5.7 | 2.9 |
| | 65521 | 15409 | 315.1 | 88.3 | 22.6 | 20.0 | 7.2 | 2.9 |
| | 65521 | 17364 | 372.7 | 126.6 | 23.3 | 11.5 | 7.3 | 2.8 |

and when both variance reduction techniques are used. The rule based on $M_8$ does worse than MC for the ACV estimators when $n = 13331$. The variance is sometimes as much as 45 times smaller with the rule based on $M_{8,3}$ compared with that based on $M_8$. When the best $a$ with respect to $M_8$ has a reasonable value of $M_{8,3}$, the two LR estimators have a comparable performance, with a slight advantage to the one based on $M_{8,3}$.

## References


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

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

[3] K. Entacher, P. Hellekalek, and P. L'Ecuyer. Quasi-Monte Carlo node sets from linear congruential generators. In Niederreiter and Spanier [18]. To appear.

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

[5] P. Hellekalek and G. Larcher, editors. *Random and Quasi-Random Point Sets*, volume 138 of *Lecture Notes in Statistics.* Springer, New York, 1998.

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

[7] F. J. Hickernell. Lattice rules: How well do they measure up? In Hellekalek and Larcher [5], pages 109–166.

[8] F. J. Hickernell. What affects accuracy of quasi-Monte Carlo quadrature? In Niederreiter and Spanier [18]. To appear.

[9] D. E. Knuth. *The Art of Computer Programming, Volume 2: Seminumerical Algorithms.* Addison-Wesley, Reading, Mass., third edition, 1997.

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

[11] P. L'Ecuyer. Maximally equidistributed combined Tausworthe generators. *Mathematics of Computation*, 65(213):203–213, 1996.

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

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

[14] P. L'Ecuyer and C. Lemieux. Variance reduction via lattice rules. Submitted, 1999.

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

[16] C. Lemieux and P. L'Ecuyer. A comparison of Monte Carlo, lattice rules and other low-discrepancy point sets. In Niederreiter and Spanier [18]. To appear.

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

[18] H. Niederreiter and J. Spanier, editors. *Monte Carlo and Quasi-Monte Carlo Methods 1998.* Lecture Notes in Computational Science and Engineering. Springer-Verlag, New York, 1999. To appear.

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

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

[21] B. Tuffin. On the use of low-discrepancy sequences in Monte Carlo methods. *Computing*, 61:371–378, 1998.