Variance Reduction via Lattice Rules

Pierre L'Ecuyer

Christiane Lemieux

Département d'Informatique et de Recherche Opérationnelle
Université de Montréal, C.P. 6128, Succ. Centre-Ville
Montréal, Canada, H3C 3J7
http://www.iro.umontreal.ca/∼lecuyer/

November, 1999

Les Cahiers du GERAD
G–99–46

Abstract

This is a review article on lattice methods for multiple integration over the unit hypercube, with a variance-reduction viewpoint. It also contains some new results and ideas. The aim is to examine the basic principles supporting these methods and how they can be used effectively for the simulation models that are typically encountered in the area of Management Science. These models can usually be reformulated as integration problems over the unit hypercube with a large (sometimes infinite) number of dimensions. We examine selection criteria for the lattice rules and suggest criteria which take into account the quality of the projections of the lattices over selected low-dimensional subspaces. The criteria are strongly related to those used for selecting linear congruential and multiple recursive random number generators. Numerical examples illustrate the effectiveness of the approach.

Keywords: Simulation, Variance Reduction, Quasi-Monte Carlo, Low Discrepancy, Lattice Rules

Résumé

Nous survolons les méthodes de réseaux pour l'intégration multiple sur l'hypercube unitaire, avec un point de vue axé sur la réduction de variance. L'article contient aussi quelques nouvelles idées et nouveaux résultats. L'objectif est d'examiner les principes de base de ces méthodes et de voir comment on peut les utiliser pour simuler les modèles rencontrés couramment en sciences de la gestion et en recherche opérationnelle. On peut habituellement reformuler les problèmes d'estimation dans ces modèles comme des problèmes d'intégration d'une fonction sur l'hypercube unitaire en un grand nombre de dimensions (parfois infini). Nous examinons des critères de sélection pour les règles de réseaux et suggérons des critères qui tiennent compte de la qualité des projections des réseaux sur des sous-espaces choisis de faible dimension. Les critères sont fortement liés à ceux utilisés pour sélectionner des générateurs pseudo-aléatoires basés sur des récurrences linéaires. Des exemples numériques illustrent l'efficacité de la méthode.

1 Introduction

The purpose of most stochastic simulations is to estimate the mathematical expectation of some cost function, in a wide sense. Sometimes the ultimate aim is optimization, but the mean estimation problem nevertheless appears at an intermediate stage. Since randomness in simulations is almost always generated from a sequence of i.i.d. $U(0,1)$ (independent and identically distributed uniforms over the interval $[0,1]$) random variables, i.e., by generating a (pseudo)random point in the t-dimensional unit hypercube $[0,1)^t$ if t uniforms are needed, the mathematical expectation that we want to estimate can be expressed as the integral of a function f over $[0,1)^t$, namely

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

If the required number of uniforms is random, one can view t as infinite, with only a finite subset of the random numbers being used. The reader who wants concrete illustrations of this general formulation can look right away at the examples in Section 10.

For small t, numerical integration methods such as the product-form Simpson rule, Gauss rule, etc. (Davis and Rabinowitz 1984), are available to approximate the integral (1). These methods quickly become impractical, however, as t increases beyond 4 or 5. For larger t, the usual estimator of μ is the average value of f over some 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 standard *Monte Carlo* (MC) simulation method, P_n is a set of n i.i.d. uniform random points over $[0,1)^t$. Then, Q_n is an unbiased estimator of μ with variance σ^2/n, i.e., $E[Q_n] = \mu$ and $\mathrm{Var}[Q_n] = \sigma^2/n$, provided that

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

i.e., if f is square-integrable over the unit hypercube. When the variance is finite, we also have the central limit theorem: $\sqrt{n}(Q_n - \mu)/\sigma \to N(0,1)$ in distribution as $n \to \infty$, so the error converges in the probabilistic sense as $|E_n| = O_p(\sigma/\sqrt{n})$, regardless of t. This error can be estimated via either the central limit theorem, or large deviations theory, or some other probabilistic method (e.g., Fishman 1996, Law and Kelton 1991).

But is it really the best idea to choose P_n at random? The *Quasi-Monte Carlo* (QMC) method constructs the point set P_n *more evenly* distributed over $[0,1)^t$ than typical random points, in order to try reducing the estimation error $|E_n|$ and perhaps improve over the $O_p(1/\sqrt{n})$ convergence rate. The precise meaning of "more evenly" depends on how we measure uniformity, and this is usually done by defining a measure of *discrepancy* between the discrete distribution determined by the points of P_n and the uniform distribution over $[0,1)^t$. A *low-discrepancy point set* P_n is a point set for which the discrepancy measure

is significantly smaller than that of a typical random point set. Discrepancy measures are often defined in a way that they can be used, together with an appropriate measure of variability of the function f, to provide a worst-case error bound of the general form:

$$|E_n| \le V(f)D(P_n) \qquad \text{for all } f \in \mathcal{F}, \tag{4}$$

where $\mathcal{F}$ is some class of functions f, $V(f)$ measures the variability of f, and $D(P_n)$ measures the discrepancy of P_n. A special case of (4) is the well-known Koksma-Hlawka inequality, for which $D(P_n)$ is the rectangular star discrepancy and $V(f)$ is the total variation of f in the sense of Hardy and Krause (see Kuipers and Niederreiter 1974 for details). Other discrepancy measures as well as thorough discussions of the concepts involved can be found in the papers of Hellekalek (1998) and Hickernell (1998a, 1998b).

The bad news is that the bounds provided by (4) turn out to be rarely practical, because even though they are tight for the worst-case function, they are very loose for "typical" functions and are usually too hard to compute anyway. The good news is that for many simulation problems, QMC nevertheless reduces the actual error $|E_n|$, sometimes by large amounts, compared with standard MC.

The two main families of construction methods for low-discrepancy point sets in practice are the *digital nets* and the *integration lattices* (Larcher 1998, Niederreiter 1992, Sloan and Joe 1994). The former usually aim at constructing so-called (t, m, s)-nets. A *low-discrepancy sequence* is an infinite sequence of points $P_\infty = \{\mathbf{u}_0, \mathbf{u}_1, \ldots\}$ such that for all n (or for an infinite increasing sequence of values of n; e.g., each power of 2), the point set $P_n = \{\mathbf{u}_0, \ldots, \mathbf{u}_{n-1}\}$ has low discrepancy. In case of the rectangular star discrepancy, this name is usually reserved to sequences for which $D(P_n) = O(n^{-1}(\ln n)^t)$. Explicit sequences that satisfy the latter condition have been constructed by Halton, Sobol', Faure, and Niederreiter. For the details, see Drmota and Tichy (1997), Niederreiter (1992), Niederreiter and Xing (1998), Sobol' (1998), Larcher (1998) and the references cited there. A convergence rate of $O(n^{-1}(\ln n)^t)$ is certainly better than the MC rate $O_p(n^{-1/2})$ *asymptotically*, but this superiority is practical only for small t. For example, for $t = 10$ already, to have $n^{-1}(\ln n)^t < n^{-1/2}$ for all $n \ge n_0$ one needs $n_0 \approx 1.2 \times 10^{39}$.

A *lattice rule* is an integration method that estimates μ by (2) and for which P_n is the intersection of an *integration lattice* with the unit hypercube. We illustrate the idea with the following special case. Consider the simple linear recurrence

$$x_i = (ax_{i-1}) \bmod n, \qquad u_i = x_i/n, \tag{5}$$

where $0 < a < n$. This kind of recurrence, with a very large n, has been used for a long time for constructing linear congruential random number generators (LCGs) (e.g., Knuth 1997, L'Ecuyer 1998). In that context, common wisdom says that n should be several orders of magnitude larger than the total number of random numbers u_i that could be used in a single experiment. Here, we take a small n and let P_n be the set of all vectors of t successive values produced by (5), from all initial states x_0, that is, $P_n = \{(u_0, \ldots, u_{t-1}) : x_0 \in \mathbb{Z}_n\}$, where $\mathbb{Z}_n = \{0, \ldots, n-1\}$. We know (e.g., Knuth 1997) that this P_n has a very regular structure: It is the intersection of a lattice with the unit hypercube $[0, 1)^t$. A lattice rule Q_n using this P_n was first proposed by Korobov (1959) and is called a *Korobov lattice rule*.

Figure 1: All pairs (u_i, u_{i+1}) for the LCGs with $(n, a) = (101, 12)$ (left side) and $(n, a) = (101, 51)$ (right side)

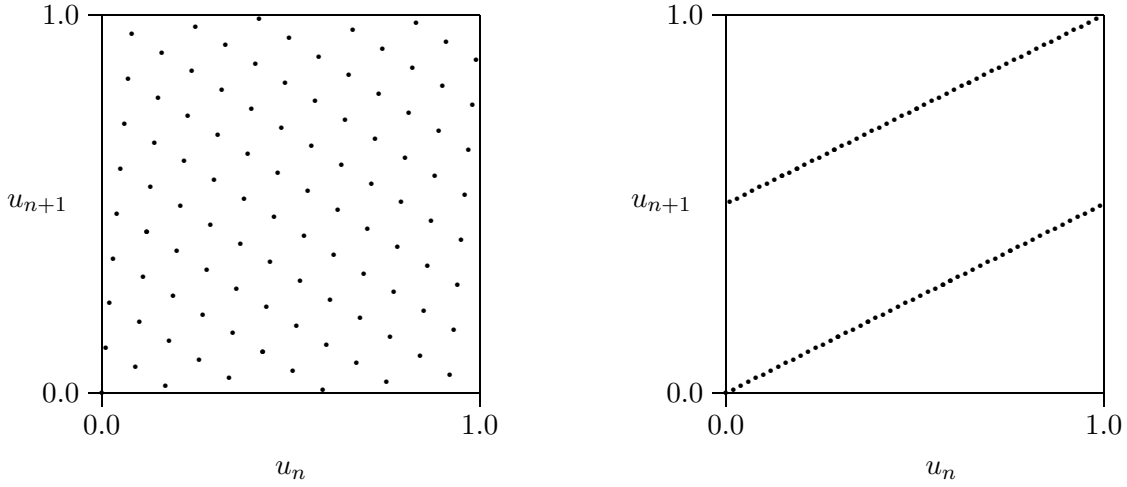

Figure 1 (left) illustrates the lattice structure of the point set P_n for $n = 101$, $a = 12$, and $t = 2$. The points are clearly more regular than typical random points. For a simulation problem that requires only 2 random numbers (a baby example, of course), one can evaluate the function f at the 101 points of P_n and take the average as an estimate of μ. This is a simplified example of QMC by a lattice rule. Using a lattice does not guarantee that the points are well-distributed in the unit hypercube. For instance, Figure 1 (right) shows P_n again for $t = 2$ an $n = 101$, but a changed to 51. This new lattice structure is certainly less attractive, because of the large gaps between the lines. In this case, the lattice rule Q_n would sample the function only on these 2 lines, whereas the one with $a = 12$ would sample more evenly in the unit square.

Some questions that arise regarding QMC via lattice rules: What are proper selection criteria for the lattice parameters? How do we bound or estimate the error E_n? Since error bounds of the form (4) are not very practical, one can consider randomizations of P_n that preserve its uniformity, while making E_n random with mean 0 and providing an unbiased estimator for its variance. Selection criteria for lattice parameters can then be defined by attempting to minimize the variance of E_n for "typical" functions f.

In the next section of this paper, we recall basic definitions and properties of lattices, define lattice rules and their node sets P_n, and examine certain regularity and stationarity properties that the projections of P_n over lower-dimensional subspaces may have. In Section 3 we give error expressions and error bounds for lattice rules. In Section 4 we provide a randomization scheme for a lattice rule, by a uniform rotation modulo 1, and derive explicit expressions for the mean and the variance of the randomized estimator, which we compare to the corresponding expressions for the MC estimator. We also discuss other

randomization approaches. In Section 5 we describe an ANOVA decomposition of f into a sum of lower-dimensional functions. The corresponding decomposition of the variance σ^2 serves to define the concept of effective dimension of f. Selection criteria for lattice rules are discussed in Section 6, where we recall some popular measures of discrepancy and propose a new figure of merit that takes into account the quality of certain low-dimensional projections. This new criterion could also be used for selecting random number generators, as in L'Ecuyer (1999a). In Section 7 we discuss copy rules and explain why we do not recommend them. A polynomial version of lattice rules is introduced in Section 8. Techniques for smoothing the function f and for lowering the effective dimension are outlined in Section 9. In Section 10, we use randomized lattice rules as a variance reduction technique for 3 simulation models for which t is small, medium, and infinite, respectively. The method improves efficiency in all cases.

2 Integration Lattices

We start with a short review on lattices. The reader can find more in, e.g., Conway and Sloane (1988) and Sloan and Joe (1994). The (integration) *lattices* discussed in this paper are discrete subsets of the real space $\mathbb{R}^t$, that contain $\mathbb{Z}^t$ (the integer vectors), and can be expressed as

$$L_t = \left\{ \mathbf{v} = \sum_{j=1}^{t} z_j \mathbf{v}_j \mid \text{ each } z_j \in \mathbb{Z} \right\}, \tag{6}$$

where $\mathbf{v}_1, \ldots, \mathbf{v}_t$ are linearly independent vectors in $\mathbb{R}^t$ which form a *basis* of the lattice. The matrix $\mathbf{V}$ whose ith line is $\mathbf{v}_i$ is the corresponding *generator matrix* of L_t. A lattice L_t shifted by a constant vector $\mathbf{v}_0 \notin L_t$, i.e., a point set of the form $L'_t = \{\mathbf{v} + \mathbf{v}_0 : \mathbf{v} \in L_t\}$, is called a *grid*, or a *shifted lattice*.

The *dual lattice* of L_t is defined as $L_t^* = \{\mathbf{h} \in \mathbb{R}^t : \mathbf{h} \cdot \mathbf{v} \in \mathbb{Z} \text{ for all } \mathbf{v} \in L_t\}$. The *dual* of a given basis $\mathbf{v}_1, \ldots, \mathbf{v}_t$ is the set of vectors $\mathbf{w}_1, \ldots, \mathbf{w}_t$ in $\mathbb{R}^t$ such that $\mathbf{v}_i \cdot \mathbf{w}_j = \delta_{ij}$ ($\delta_{ij} = 1$ if $i = j$, $\delta_{ij} = 0$ otherwise). It is a basis of the dual lattice. These $\mathbf{w}_j$'s are the columns of the matrix $\mathbf{V}^{-1}$, so they can be computed by inverting $\mathbf{V}$.

The determinant of the matrix $\mathbf{V}$ is equal to the volume of the fundamental parallelepiped $\Lambda = \{\mathbf{v} = \lambda_1 \mathbf{v}_1 + \cdots + \lambda_t \mathbf{v}_t : 0 \le \lambda_i \le 1 \text{ for } 1 \le i \le t\}$, and is always equal to the inverse of the density of points, independently of the choice of basis. It is called the determinant of L_t. In other words, the average number of points per unit of volume is $1/\det(L_t) = 1/\det(\mathbf{V}) = \det(\mathbf{V}^{-1})$. This number, called the *density*, is always an integer and is equal to the number of points in *every* cubic box of volume 1 aligned with the axes. The *node set* $P_n = L_t \cap [0,1)^t$ contains exactly $n = 1/\det(L_t)$ points. A *lattice rule* (of integration) of *order* n for μ is a rule of the form (2) with $\{\mathbf{u}_0, \ldots, \mathbf{u}_{n-1}\} = P_n = L_t \cap [0,1)^t$. One can always write

$$P_n = \{((j_1/n_1)\mathbf{v}_1 + \cdots + (j_r/n_r)\mathbf{v}_r) \bmod 1 : 0 \le j_i < n_i \text{ for } i = 1, \ldots, r\}, \tag{7}$$

where the reduction modulo 1 is performed coordinate-wise, the $\mathbf{v}_i$'s are linearly independent *generating vectors*, and $n = n_1 \ldots n_r$. The smallest r for which this holds is called

the *rank* of the lattice rule. Rules of rank $r > 1$ are recommended by Sloan and Joe (1994) based on certain theoretical properties. In Section 7 we explain why we disagree with this recommendation. Elsewhere, we restrict our attention to $r = 1$. For a rule of rank 1, we have

$$P_n = \{(j/n)\mathbf{v} \bmod 1 : 0 \le j < n\} \tag{8}$$

for some vector $\mathbf{v}$. As an important special case, for any LCG defined by (5), the set $P_n = \{(u_0, \ldots, u_{t-1}) : x_0 \in \mathbb{Z}_n\}$ corresponds to a lattice rule of rank 1 with $\mathbf{v} = (1, a, \ldots, a^{t-1})$, which is a Korobov rule, or a rule in Korobov form.

For a rule of rank 1, P_n can be enumerated in a straightforward way by starting with $\mathbf{u} = \mathbf{0}$ and performing $n - 1$ iterations of the form $\mathbf{u} = (\mathbf{u} + \mathbf{v}) \bmod 1$. This requires $O(tn)$ additions modulo 1. If the rule is in Korobov form and if the corresponding LCG has period length $n - 1$ (i.e., n is a prime number and $\nu = n - 1$ is the smallest positive ν for which $a^\nu \bmod n = 1$), then P_n can be enumerated as follows: Start with $x_1 = 1$ and generate the sequence $u_1, u_2, \ldots, u_{n+t-2}$ via (5). Along the way, enumerate $\mathbf{u}_1, \ldots, \mathbf{u}_{n-1}$, the overlapping vectors of successive values. Then add the vector $\mathbf{u}_0 = \mathbf{0}$. This requires $O(n+t)$ multiplications by a, modulo n, plus some overhead to shift the vector components at each iteration, instead of $O(tn)$ additions. The enumeration approach based on the LCG recurrence still works when the LCG has several cycles, but one must run the LCG over each of its cycles, and this becomes more cumbersome as the number of cycles increases.

For a given lattice L_t and a subset of coordinates $I = \{i_1, \ldots, i_d\} \subseteq \{1, \ldots, t\}$, denote by $L_t(I)$ the projection of L_t over the d-dimensional subspace determined by the coordinates in I. This projection is also a lattice, whose density divides that of L_t (there are exactly $\det(L_t(I))/\det(L_t)$ points of L_t that are projected onto each point of $L_t(I)$; in group theory language, $L_t(I)$ corresponds to a coset of L_t). Denote $P_n(I) = L_t(I) \cap [0,1)^d$, the corresponding projection of P_n. For reasons to be explained later, we would like to have not only P_n evenly distributed over $[0,1)^t$, but also $P_n(I)$ evenly distributed over its subspace, at least for certain subsets I deemed important.

Sloan and Joe (1994) call a rank-1 lattice L_t *projection-regular* if all its *principal projections*, $L_t(\{1, \ldots, d\})$ for $1 \le d < t$, have the same density as L_t. This property holds if and only if $\det(L_t(\{1\})) = \det(L_t)$, and implies that the projection $P_n(I)$ contains as many distinct points as P_n whenever I contains 1. We call L_t *fully projection-regular* if $\det(L_t(I)) = \det(L_t)$ for any non-empty $I \subseteq \{1, \ldots, t\}$, i.e., if each $P_n(I)$ contains as many distinct points as P_n. Projection-regularity is easily verified by computing the greatest common divisors (gcd) between n and the coordinates of the generating vector $\mathbf{v}$:

PROPOSITION 1. *A rank-1 lattice L_t with generating vector $\mathbf{v} = (v_1, \ldots, v_t)$ is projection-regular if and only if* $\gcd(n, v_1) = 1$. *It is fully projection-regular if and only if* $\gcd(n, v_d) = 1$ *for* $1 \le d \le t$.

PROOF. The lattice is projection-regular if and only if the 1-dimensional projection $P_n(\{1\})$ contains n distinct points. If $\gcd(v_1, n) = 1$ and $jv_1 \bmod n = iv_1 \bmod n$, then $j - i$ must be a multiple of n, which implies that the points of $P_n(\{1\})$ are all distinct. On

the other hand, if $\gcd(v_1, n) = \rho \neq 1$, then for $j - i = n/\rho$, $(j-i)v_1$ is a multiple of n, so $iv_1 = jv_1 \bmod n$ and therefore $P_n(\{1\})$ contains no more than n/ρ points. This completes the proof of the first part. For the second part, take a 1-dimensional projection over the dth coordinate and use the same argument as in the first part to see that the points of $P_n(\{d\})$ are all distincts if and only if $\gcd(n, v_d) = 1$. This implies that the points of $P_n(I)$ are all distincts for any non-empty I. $\Box$

In particular, a Korobov rule is always projection-regular, since $v_1 = 1$. It is fully projection-regular if $\gcd(a, n) = 1$, e.g., if n is prime and $1 \leq a < n$, or if n is a power of 2 and a is odd. A general rank-1 rule is fully projection-regular, e.g., if n is prime and $1 \leq v_d < n$ for each d, or if n is a power of 2 and each v_d is odd.

Korobov point sets, among others, have the property that several of their projections $P_n(I)$ are identical, so one can assess the quality of a large family of projections by examining only a subset of these projections. More specifically, we say that a point set P_n is *dimension-stationary* if $P_n(\{i_1, \ldots, i_d\}) = P_n(\{i_1 + j, \ldots, i_d + j\})$ for all $i_1, \ldots, i_d$ and j such that $1 \leq i_1 < \cdots < i_d \leq i_d + j \leq t$. In other words, the projections $P_n(I)$ of a dimension-stationary point set depend only on the *spacings* between the indices in I. Every Korobov rule for which $\gcd(a, n) = 1$ is dimension-stationary. More generally, given a recurrence of the form $\xi_i = \tau(\xi_{i-1})$ where $\tau : \Xi \to \Xi$ and Ξ is a finite set, if τ is invertible and $g : \Xi \to [0, 1)$ then $P_n = \{\mathbf{u} = (g(\xi_0), \ldots, g(\xi_{t-1})) : \xi_0 \in \Xi\}$, the set of all (overlapping) output vectors over all the cycles of the recurrence, is a dimension-stationary point set (Lemieux and L'Ecuyer 1999b). Recurrences of this form (with a very large Ξ) are widely used for constructing pseudorandom number generators (e.g., L'Ecuyer 1994, Niederreiter 1992). Their dimension-stationary property is an important advantage when using them in a QMC context. This property does not hold in general for popular QMC point sets such as (typical) (t, m, s)-nets with $t > 0$.

3 Integration Error for Lattice Rules

The Fourier expansion of f is

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

with *Fourier coefficients*

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

Since $\hat{f}(\mathbf{0}) = \mu$, the integration error for a general point set P_n can be written in terms of this expansion as

$$E_n \;=\; \frac{1}{n} \sum_{i=0}^{n-1} (f(\mathbf{u}_i) - \mu)$$

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

$$= \frac{1}{n}\sum_{\mathbf{0}\neq\mathbf{h}\in\mathbf{Z}^t}\hat{f}(\mathbf{h})\sum_{i=0}^{n-1}\exp(2\pi\sqrt{-1}\,\mathbf{h}\cdot\mathbf{u}_i), \tag{9}$$

assuming that we can interchange the summations. In particular, if the Fourier expansion of f is absolutely convergent, i.e., $\sum_{\mathbf{h}\in\mathbf{Z}^t}|\hat{f}(\mathbf{h})| < \infty$, then Fubini's theorem (e.g., Rudin 1974) guarantees that the interchange is valid. Sloan and Osborn (1987) have shown that if P_n is a lattice node set, i.e., $P_n = L_t \cap [0,1)^t$, (9) simplifies to the sum of the Fourier coefficients over the nonzero vectors of the dual lattice:

$$E_n = \sum_{\mathbf{0}\neq\mathbf{h}\in L_t^*}\hat{f}(\mathbf{h}). \tag{10}$$

The proof consists in showing that

$$\sum_{i=0}^{n-1}\exp(2\pi\sqrt{-1}\,\mathbf{h}\cdot\mathbf{u}_i) = \begin{cases} n & \text{if } \mathbf{h}\in L_t^* \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

(Sloan and Joe 1994, Theorem 2.8). If we knew how to efficiently compute [estimate] the Fourier coefficients of f for all $\mathbf{h}\in L_t^*$, we could compute [estimate] the integration error, but this is usually much too complicated in real-life applications.

The error expression (10) immediately suggests a discrepancy measure (or quality criterion) of the form

$$D(P_n) = \sum_{\mathbf{0}\neq\mathbf{h}\in L_t^*} w(\mathbf{h}) \tag{12}$$

or

$$D'(P_n) = \sup_{\mathbf{0}\neq\mathbf{h}\in L_t^*} w(\mathbf{h}) \tag{13}$$

for lattice rules, where the $w(\mathbf{h})$ are arbitrary non-negative weights that decrease with the "size" of $\mathbf{h}$, in a way to be specified. Indeed, for well-behaved (smooth) functions f, $|\hat{f}(\mathbf{h})|$ should tend to decrease with the size of $\mathbf{h}$. (Later on in this section, we will arrive again at the general form of criterion (12–13) by a different route, via a variance minimization argument.) For example, $w(\mathbf{h})$ can be a decreasing function of the norm of $\mathbf{h}$, for some choice of norm. The faster it decreases, the smoother the function (crudely speaking). The specific form of $w(\cdot)$ should reflect our a priori assumptions about the class of functions that we want to consider. An obvious worst-case error bound is then given by:

PROPOSITION 2. *Let $\mathcal{F}$ be the class of functions f such that $|\hat{f}(\mathbf{h})| \le K w(\mathbf{h})$ for all $\mathbf{h}\in L_t^*$, $\mathbf{h}\neq\mathbf{0}$, where K is a constant. Then for all $f\in\mathcal{F}$, $|E_n| \le K D(P_n)$.*

This proposition may look trivial. It can perhaps demystify some worst-case error bounds given in the literature (e.g., Lyness and Sloan 1989, Sloan and Joe 1994). These bounds are often special cases or variants of Proposition 2, with specific choices of $w(\cdot)$.

Hickernell (1998b) provides several error bounds of the form (4) based on variants of (12). For instance, it is easily shown, using (10) and Hölder's inequality, that (4) holds with

$$(D(P_n))^p = \sum_{\mathbf{0}\neq\mathbf{h}\in L_t^*} w(\mathbf{h})^p \tag{14}$$

and

$$(V(f))^q = \sum_{\mathbf{0}\neq\mathbf{h}\in\mathbf{Z}^t} (|\hat{f}(\mathbf{h})|/w(\mathbf{h}))^q, \tag{15}$$

for arbitrary $p, q > 0$ such that $1/p + 1/q = 1$. If we take $p = 2$, $w(\mathbf{h}) = \prod_{j\in I(\mathbf{h})}(\beta_j/|h_j|)^\alpha$ for some positive integer $\alpha > 0$ and arbitrary positive weights $\beta_1, \ldots, \beta_t$, where $I(\mathbf{h})$ denotes the set of nonzero coordinates of $\mathbf{h}$, and we consider the class of functions f whose periodic continuation $\bar{f}$ (defined by $\bar{f}(\mathbf{u}) = f(\mathbf{u} \bmod 1)$ for $\mathbf{u} \in \mathbb{R}^t$) is continuous over the entire space $\mathbb{R}^t$ and has mixed partial derivatives of order α or less that are square integrable over $[0,1)^t$, then $V(f)$ is finite over that class and can be written in terms of the integrals of these mixed partial derivatives. Bounding the partial derivatives can then provide a bound on the integration error, via (4). See Hickernell (1998b) for the details. This upper bound motivates the criterion $\tilde{\mathcal{P}}_{\alpha,p}(P_n)$ to be discussed in Section 6.

From a practical viewpoint, these bounds and those given by Proposition 2 do not resolve the problem of estimating the error, because they require explicit bounds $Kw(\mathbf{h})$ on the Fourier coefficients which must decay quickly enough so that $D(P_n) < \infty$, or we need bounds on the mixed partial derivatives. Such bounds are almost never available. To be on the safer side regarding the assumptions of Proposition 2, we may want to take a $w(\cdot)$ that decreases more slowly, but then the error bounds tend to become too wide. The situation is actually darker: The Fourier expansion of f can be absolutely convergent *only if* the periodic continuation of f is continuous over the entire space $\mathbb{R}^t$. For typical simulation problems encountered in management science, the function $\bar{f}$ is discontinuous at the boundary of the unit hypercube, and often in the interior of the hypercube as well.

What we need is a different way of estimating the error. An attractive solution is to obtain a probabilistic error estimator via independent randomizations of the point set P_n, as described in the next section. Numerical analysts sometimes argue against probabilistic error estimates because they are not 100% *guaranteed*, in contrast to the deterministic bounds. We believe that estimates that we can compute are more useful than bounds that are several orders of magnitude too wide, or that we cannot compute.

Another (highly heuristic) way of assessing the error is to repeat the integration with a sequence of lattice rules that contain an increasing number of points (e.g., doubling n each time), and stop when the approximation Q_n seems to have stabilized. These lattices can be embedded (i.e, $P_{n'} \subset P_n$ if $n' < n$ and if these are the node sets of two of these lattice rules) or not. The problem with this approach is that the error often decreases

in a non-monotone fashion, and may still be very large even if the value of Q_n did not change after we have doubled n. This would occur, for example, if important terms in the error expression (10) correspond to values of $\mathbf{h}$ that belong to none of the dual lattices of the node sets considered so far. For every fixed sequence of rules, it is easy to construct examples for which this happens.

4 Random Shifts and Variance Expressions

A simple way of randomizing P_n without destroying its regular structure is to shift it randomly, modulo 1, with respect to all of the coordinates, as proposed by Cranley and Patterson (1976). Generate one point $\mathbf{u}$ 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). Let $\tilde{P}_n = \{\tilde{\mathbf{u}}_0, \ldots, \tilde{\mathbf{u}}_{n-1}\}$, $\tilde{Q}_n = (1/n)\sum_{i=0}^{n-1} f(\tilde{\mathbf{u}}_i)$, and $\tilde{E}_n = \tilde{Q}_n - \mu$. This can be repeated m times, independently, with the same P_n, thus obtaining m i.i.d. copies of the random variable $\tilde{Q}_n$, which we denote $X_1, \ldots, X_m$. Let $\bar{X} = (X_1 + \cdots + X_m)/m$ and $S_x^2 = \sum_{j=1}^{m}(X_j - \bar{X})^2/(m-1)$. We now have:

PROPOSITION 3. *$E[\bar{X}] = E[X_j] = \mu$ and $E[S_x^2] =$ $\mathrm{Var}[X_j] = m\ \mathrm{Var}[\bar{X}]$.*

PROOF. The first part is quite obvious: Since each $\tilde{\mathbf{u}}_i$ is a random variable uniformly distributed over $[0,1)^t$, each $f(\tilde{\mathbf{u}}_i)$ is an unbiased estimator of μ, and so is their average. Sloan and Joe (1994) give a different proof of their Theorem 4.11. For the second part, which seems new, it suffices to show that the X_j's are pairwise uncorrelated. Without loss of generality, it suffices to show that $\mathrm{Cov}(X_1, X_2) = 0$. Let $\mathbf{u}$ and $\mathbf{u}'$ be the two independent uniforms used to randomly shift the points to compute X_1 and X_2, respectively. Then, for any $i, \ell \in \{0, \ldots, n-1\}$, $\tilde{\mathbf{u}}_i = (\mathbf{u}_i + \mathbf{u}) \bmod 1$ and $\tilde{\mathbf{u}}'_\ell = (\mathbf{u}_\ell + \mathbf{u}') \bmod 1$ are independent and uniformly distributed over $[0,1)^t$, so that $\mathrm{Cov}[f(\tilde{\mathbf{u}}_i),\ f(\tilde{\mathbf{u}}'_\ell)] = 0$. Therefore,

$$\mathrm{Cov}[X_1,\ X_2] = \frac{1}{n^2}\ \mathrm{Cov}\left[\sum_{i=0}^{n-1} f(\tilde{\mathbf{u}}_i),\ \sum_{\ell=0}^{n-1} f(\tilde{\mathbf{u}}'_\ell)\right] = \frac{1}{n^2}\sum_{i=0}^{n-1}\sum_{\ell=0}^{n-1}\ \mathrm{Cov}[f(\tilde{\mathbf{u}}_i),\ f(\tilde{\mathbf{u}}'_\ell)] = 0.$$

□

It should be underlined that Proposition 3 holds for any point set P_n; it does not have to come from a lattice. This variance estimation method, by random shifts modulo 1, therefore applies to any kind of low-discrepancy point set. We also mention that Q_n itself is *not* an unbiased estimator of μ (it is not a random variable). Observe that Proposition 3 holds under *weaker* conditions than (10); the Fourier expansion of f need not be absolutely convergent.

We now know how to estimate the variance, but this variance estimator says nothing about how to determine our lattice selection criteria. Since $\bar{X}$ is a *statistical* estimator of μ, the natural goal is to minimize its variance, i.e., minimize $\mathrm{Var}[\tilde{Q}_n]$. The next proposition expresses this variance in terms of the (squared) Fourier coefficients, both for a lattice rule and for plain MC (for comparison). Tuffin (1998) gives a different proof of (17) (in

the proof of Theorem 2) under the condition that the Fourier expansion of f is absolutely convergent. This is a much stronger condition than the square integrability of f (i.e., finite variance), and it rarely holds for real-life simulation models.

PROPOSITION 4. *If f is square-integrable, with the MC method (i.e., if P_n contains n i.i.d. random points) we have*

$$\mathrm{Var}[\tilde{Q}_n] = \mathrm{Var}[Q_n] = \frac{1}{n} \sum_{\mathbf{0} \neq \mathbf{h} \in \mathbf{Z}^t} |\hat{f}(\mathbf{h})|^2. \tag{16}$$

For a randomly shifted lattice rule, we have

$$\mathrm{Var}[\tilde{Q}_n] = \sum_{\mathbf{0} \neq \mathbf{h} \in L_t^*} |\hat{f}(\mathbf{h})|^2. \tag{17}$$

PROOF. With MC, (16) follows from Parseval's equality (Rudin 1974) and the fact that $\hat{f}(\mathbf{0}) = \mu$. For the randomly shifted lattice rule, if we define the function $g : [0,1)^t \to \mathbb{R}$ by $g(\mathbf{u}) = \sum_{i=0}^{n-1} f((\mathbf{u}_i + \mathbf{u}) \bmod 1)/n$, we get

$$\mathrm{Var}[\tilde{Q}_n] = \mathrm{Var}(g(\mathbf{u})) = \sum_{\mathbf{0} \neq \mathbf{h} \in \mathbf{Z}^t} |\hat{g}(\mathbf{h})|^2, \tag{18}$$

by using the Parseval equality on g. The Fourier coefficients $\hat{g}(\mathbf{h})$ are

$$\begin{aligned}
\hat{g}(\mathbf{h}) &= \int_{[0,1)^t} g(\mathbf{u}) e^{-2\pi\sqrt{-1}\mathbf{h}\cdot\mathbf{u}} d\mathbf{u} \\
&= \int_{[0,1)^t} \left(\frac{1}{n} \sum_{i=0}^{n-1} f((\mathbf{u}_i + \mathbf{u}) \bmod 1) \right) e^{-2\pi\sqrt{-1}\mathbf{h}\cdot\mathbf{u}} d\mathbf{u} \\
&= \frac{1}{n} \sum_{i=0}^{n-1} \int_{[0,1)^t} f((\mathbf{u}_i + \mathbf{u}) \bmod 1) e^{-2\pi\sqrt{-1}\mathbf{h}\cdot\mathbf{u}} d\mathbf{u} \\
&= \frac{1}{n} \sum_{i=0}^{n-1} \int_{[0,1)^t} f(\mathbf{v}_i) e^{-2\pi\sqrt{-1}\mathbf{h}\cdot(\mathbf{v}_i - \mathbf{u}_i)} d\mathbf{v}_i \\
&= \frac{1}{n} \sum_{i=0}^{n-1} e^{2\pi\sqrt{-1}\mathbf{h}\cdot\mathbf{u}_i} \int_{[0,1)^t} f(\mathbf{v}_i) e^{-2\pi\sqrt{-1}\mathbf{h}\cdot\mathbf{v}_i} d\mathbf{v}_i \\
&= \frac{1}{n} \sum_{i=0}^{n-1} e^{2\pi\sqrt{-1}\mathbf{h}\cdot\mathbf{u}_i} \hat{f}(\mathbf{h}) \\
&= \begin{cases} \hat{f}(\mathbf{h}) & \text{if } \mathbf{h} \in L_t^* \\ 0 & \text{otherwise.} \end{cases}
\end{aligned} \tag{19}$$

In the last display, the third equality follows from Fubini's theorem (Rudin 1974) because f is square-integrable over the unit hypercube, the fourth one is obtained by making the change of variable $\mathbf{v}_i = (\mathbf{u}_i + \mathbf{u}) \bmod 1$, and the last one follows from (11). We can now replace $\hat{g}(\mathbf{h})$ by (19) in (18) and this yields the required result. □

The variance is smaller for the randomly shifted lattice rule than for MC if and only if the squared Fourier coefficients are smaller "in the average" over L_t^* than over $\mathbb{Z}^t$. The worst case is when all the nonzero Fourier coefficients of f belongs to L_t^*. The variance of $\tilde{Q}_n$ is then n times larger with the randomly shifted lattice rule than with standard MC. Fortunately, for typical real-life problems, the variance is smaller with the lattice rule than with MC.

Heuristic arguments now enter the scene. A reasonable assumption, similar to the one discussed just after (12–13), is that for well-behaved problems the squared Fourier coefficients should tend to decrease quickly with the size of $\mathbf{h}$, where the size can again be measured in different ways. Small $\mathbf{h}$'s correspond to low-frequency waves in the function f, and are typically more important than the high-frequency waves, which are eventually (for very large $\mathbf{h}$) undetected even by standard MC because of the finite precision in the representation of the real numbers on the computer. The small coordinates in $\mathbf{h}$ also correspond to the most significant bits of the $\mathbf{u}_i$'s, which are usually the most important. This argument leads us to the same general discrepancy measure as in the previous section, namely (12–13). So we are back to the same question: How do we choose w?

Proposition 2 can be rephrased in terms of the variance. This is of course a trivial result, but an important point to underline is that for a given function w such that the sum in (12) converges, the class $\mathcal{F}'$ in the next proposition is generally much larger than $\mathcal{F}$ in Proposition 2.

PROPOSITION 5. *Let $\mathcal{F}'$ be the class of functions f such that $|\hat{f}(\mathbf{h})|^2 \le Kw(\mathbf{h})$ for all $\mathbf{h} \in L_t^*$, $\mathbf{h} \ne \mathbf{0}$, where K is a constant. Then for all $f \in \mathcal{F}'$, $\mathrm{Var}[\tilde{Q}_n] \le KD(P_n)$.*

There are other ways of randomizing P_n than the random shift. Some of them guarantee a variance reduction, but destroy the lattice structure, and do not perform as well as the random shift of P_n for most typical problems, according to our experience. Two of these methods are *stratification* and *latin hypercube sampling* (LHS). One can stratify by partitioning the unit hypercube as follows. For a given basis of L_t, let Λ be the fundamental parallelepiped defined in Section 2, and let $\Lambda_i = (\Lambda + \mathbf{u}_i) \bmod 1$ for each $\mathbf{u}_i \in P_n$. These Λ_i, $0 \le i < n$, form a partition of $[0,1)^t$. For each i, generate a random point $\tilde{\mathbf{u}}_i$ uniformly in Λ_i and adopt the estimator $\tilde{Q}_n$ defined as before, but with these new $\tilde{\mathbf{u}}_i$'s. Since this is *stratified sampling* (Cochran 1977), it follows immediately that $\mathrm{Var}[\tilde{Q}_n]$ is smaller with this scheme than with standard MC (or equal, if f is constant over each Λ_i). Implementing this requires more work than MC and than the random shift of P_n. LHS, on the other hand, constructs the points $\tilde{\mathbf{u}}_i = (\tilde{u}_{i,1}, \ldots, \tilde{u}_{i,t})$ as follows. Let $u_{i,1} = i/n$ for $i = 0, \ldots, n-1$, and let $(u_{0,s}, \ldots, u_{n-1,s})$ be independent random permutations of $(0, 1/n, \ldots, (n-1)/n)$, for $s = 2, \ldots, t$. (This is equivalent to taking the node set of a lattice rule and, for each s, randomly permuting the sth coordinate values of the n points. Such a randomization

completely destroys the lattice structure, except for the unidimensional projections.) Then, let $\tilde{u}_{i,s} = u_{i,s} + \delta_{i,s}/n$ for each (i, s), where the $\delta_{i,s}$ are i.i.d. $U(0, 1)$. The estimator is again $\tilde{Q}_n$. Its variance never exceeds $n/(n-1)$ times that of MC (Owen 1998), and does not exceed the MC variance under the *sufficient* condition that f is monotone with respect to each of its coordinates (Avramidis and Wilson 1996). In the one-dimensional case (and for each one-dimensional projection), LHS is equivalent to the stratification scheme described a few sentences ago. For $s > 1$, however, the s-dimensional projections are not necessarily well distributed under LHS.

Observe that we did not assume $t \leq n$ anywhere so far. Taking $t \gg n$ means that the LCG (5) will cycle several times over the same sequence of values of u_i. However, with the randomly shifted lattice rule this is not a problem because the randomization takes care of shifting the different coordinates differently, which means that the $\tilde{u}_i$ *do not* cycle. Section 10.3 gives an example where we took $t \gg n$.

Rather than analyzing the variance of a randomized lattice for a fixed function, some authors have analyzed the mean square error (MSE) over a space of random functions f. This MSE is equal to the mean square discrepancy, for an appropriate definition of the discrepancy. See, e.g., Woźniakowski (1991), Hickernell (1998b), Hickernell and Hong (1999).

5 Functional ANOVA Decomposition

The functional ANOVA decomposition of Hoeffding (Hoeffding 1948, Efron and Stein 1981, Owen 1998) writes f as a sum of orthogonal functions, where each function depends on a distinct subset I of the coordinates:

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

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

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

and for a randomly shifted lattice rule, one has

$$\mathrm{Var}[\tilde{Q}_n] = \sum_{I \subseteq \{1,\dots,t\}} \sum_{\mathbf{0} \neq \mathbf{h} \in L_t^*} |\hat{f}_I(\mathbf{h})|^2, \tag{20}$$

where for $I \neq \phi$, $\sigma_I^2 = \int_{[0,1)^t} f_I^2(\mathbf{u}) d\mathbf{u}$ is the variance of f_I, 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, the ANOVA decomposition partitions the vectors $\mathbf{h}$ according to the "minimal" subspaces to which they belong, i.e., according to their sets of non-zero coordinates.

We say that f has effective dimension at most d in the *truncation sense* (Caflish, Morokoff, and Owen 1997, Owen 1998) if $\sum_{I\subseteq\{1,\ldots,d\}} \sigma_I^2$ is near σ^2, in the *superposition sense* (Caflish, Morokoff, and Owen 1997, Owen 1998) if $\sum_{|I|\leq d} \sigma_I^2$ is near σ^2, and in the *successive-dimensions* sense if $\sum_{I\subseteq\{i,\ldots,i+d-1\},\, 1\leq i\leq t-d+1} \sigma_I^2$ is near σ^2. The first definition means that f is *almost* d-dimensional (or less), while the others mean, in a different sense, that f is almost a sum of d-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 (20) for $\mathrm{Var}[\tilde{Q}_n]$ suggests discrepancy measures of the form (12) or (13), but where the sum (or the sup) is restricted to those $\mathbf{h}$ that belong to the subspaces determined by the sets I that are considered important. We propose selection criteria along these lines. In Section 9, we mention ways of changing f in order to reduce its effective dimension without changing its expectation.

EXAMPLE 1. For a concrete illustration, consider the 3-dimensional function $f(u_1, u_2, u_3) = 2u_1u_2 + 3u_3^2 + u_2$. The Fourier coefficients of the ANOVA components are $\hat{f}_{\{1\}}(h_1, 0, 0) = \sqrt{-1}/(2\pi h_1)$ if $h_1 \neq 0$, $\hat{f}_{\{2\}}(0, h_2, 0) = \sqrt{-1}/(\pi h_2)$ if $h_2 \neq 0$, $\hat{f}_{\{3\}}(0, 0, h_3) = 3[\sqrt{-1}/(2\pi h_3) + 1/(2\pi^2 h_3^2)]$ if $h_3 \neq 0$, $\hat{f}_{\{1,2\}}(h_1, h_2, 0) = -1/(2\pi^2 h_1 h_2)$ if $h_1 h_2 \neq 0$, and $\hat{f}_I(\mathbf{h}) = 0$ for every other case. The total variance is $\sigma^2 = 56/45$ and it can be decomposed as the sum of $\sigma^2_{\{1\}} = 1/12$, $\sigma^2_{\{2\}} = 1/3$, $\sigma^2_{\{3\}} = 4/5$, and $\sigma^2_{\{1,2\}} = 1/36$ (the other σ_I^2's being 0). Here, the unidimensional ANOVA components $f_{\{3\}}(u_3) = 3u_3^2 - 1$ and $f_{\{2\}}(u_2) = u_2 - 1/2$ account for about 64% and 27% of the total variance, respectively.

6 Selection Criteria for Lattice Rules

We came up with the general selection criteria (12) and (13). It now remains to choose w, and to choose between sum and sup. Two important factors to be considered are: (1) the choice should reflect our idea of the typical behavior of Fourier coefficients in the class of functions that we want to consider and (2) the corresponding figure of merit $D(P_n)$ or $D'(P_n)$ should be relatively easy and fast to compute, so that we can make computer searches for the best lattice parameters. Several choices of w and the relationships between them are discussed, e.g., by Hellekalek (1998), Hickernell (1998b), Niederreiter (1992) and in the references given there.

Historically, a standard choice for w has been $w(\mathbf{h}) = \|\mathbf{h}\|_\pi^{-\alpha}$, a negative power of the product norm $\|\mathbf{h}\|_\pi = \prod_{j=1}^t \max(1, |h_j|)$. With this w, $D(P_n)$ in (12) becomes

$$\mathcal{P}_\alpha(P_n) = \sum_{\mathbf{0}\neq\mathbf{h}\in L_t^*} \|\mathbf{h}\|_\pi^{-\alpha},$$

a special case of (14). Hickernell (1998b) suggests generalizations of $\mathcal{P}_\alpha(P_n)$, incorporating weights and replacing the simple sum in (12) by an $\mathcal{L}_p$-norm. This gives, for instance, the quantity $\tilde{\mathcal{P}}_{\alpha,p}(P_n)$ defined by

$$(\tilde{\mathcal{P}}_{\alpha,p}(P_n))^p = \sum_{\mathbf{0}\neq\mathbf{h}\in L_t^*} \left(\beta_{I(\mathbf{h})}/\|\mathbf{h}\|_\pi\right)^{\alpha p},$$

where $p \geq 1$ and the constants β_I are positive weights that assess the relative importance of the projections $P_n(I)$, i.e., the relative sizes of the σ_I^2's in the ANOVA decomposition.

In the special case of *product-type weights*, of the form

$$\beta_I = \beta_0 \prod_{j\in I} \beta_j,$$

if α is even and $p = 2$, one can write

$$(\tilde{\mathcal{P}}_{\alpha,2}(P_n))^2 = -\beta_0^{2\alpha} + \frac{\beta_0^{2\alpha}}{n} \sum_{\mathbf{u}\in P_n} \prod_{j=1}^{t} \left[1 - \frac{(-4\pi^2\beta_j^2)^\alpha}{2\alpha!} B_{2\alpha}(u_j)\right].$$

where $B_\alpha(\cdot)$ is the Bernoulli polynomial of degree α (e.g., Sloan and Joe 1994). This gives an algorithm for computing $\tilde{\mathcal{P}}_{\alpha,2}(P_n)$ in time $O(nt)$ when α is an even integer and P_n is the point set of a lattice rule. This also means that $(\tilde{\mathcal{P}}_{\alpha,2}(P_n))^2$ can be interpreted in this case as a worst-case variance bound for a class of polynomial functions with certain bounds on the coefficients (Lemieux 1999 provides further details). Note that the $D_{\mathcal{F},\alpha,p}(P)$ of Hickernell (1999) corresponds to $\tilde{\mathcal{P}}_{\alpha,p}(P_n)$, and to $(\mathcal{P}_{2\alpha}(P_n))^{1/2}$ if $p = 2$ and $\beta_j = 1$ for all j, and is a particular case of the discrepancy $D(P_n)$ in (4) and (14).

If the β_j's are less than 1, then β_I tends to decrease with $|I|$, which gives more importance to the lower-dimensional projections of f. In particular, if $\beta_j = \beta < 1$ for all j, β_I decreases geometrically with $|I|$. This means that all the projections f_I over the same number of dimensions $|I|$ are assumed to have the same importance, and the importance decreases with $|I|$. By taking $\beta_j = 1$ for each j, we obtain the classical $\mathcal{P}_\alpha(P_n)$, for which all the projections are given the same weight. With equal weights, the low-dimensional projections are given no more importance than the high-dimensional ones, and (unless t is small) their contribution is diluted by the massive number of higher-dimensional projections. Sloan and Joe (1994) provide tables of parameters for lattices rules with good values of $\mathcal{P}_2(P_n)$ in t dimensions, for t up to 12.

If we take (13) instead of (12), with the same w and with $\alpha = 1$, we get the inverse of the Babenko-Zaremba index, defined as

$$\rho_t = \min_{\mathbf{0}\neq\mathbf{h}\in L_t^*} \|\mathbf{h}\|_\pi, \tag{21}$$

which has also been suggested as a selection criterion for lattice rules, but appears harder to compute than $\mathcal{P}_2(P_n)$. This ρ_t is the limit of $\tilde{\mathcal{P}}_{1,p}^{-1}(P_n)$ as $p \to \infty$. It has been used by Maisonneuve (1972) to compute tables for up to $t = 10$.

Another (natural) choice for $w(\mathbf{h})$ is of course the $\mathcal{L}_p$-norm to some negative power, $w(\mathbf{h}) = \|\mathbf{h}\|_p^{-\alpha}$, where $\|\mathbf{h}\|_p = (|h_1|^p + \cdots + |h_t|^p)^{1/p}$. With $\alpha = 1$ and the criterion (13), $D(P_n)$ becomes the inverse of the $\mathcal{L}_p$-length of the shortest vector $\mathbf{h}$ in the dual lattice, which is equal to the $\mathcal{L}_p$-distance between the successive hyperplanes for the family of parallel equidistant hyperplanes that are farthest apart among those that cover all the points of L_t. For $p = 1$, $\|\mathbf{h}\|_1$ (or $\|\mathbf{h}\|_1 - 1$ in some cases, see Knuth 1997) is the minimal number of hyperplanes that cover all the points of P_n. For $p = 2$ (the Euclidean norm), this is the so-called *spectral test* commonly used for ranking LCGs (Hellekalek 1998, L'Ecuyer 1999b, L'Ecuyer and Couture 1997, Knuth 1997), and we use ℓ_t to denote the length of the shortest vector in this case. Since the density of the vectors $\mathbf{h}$ in L_t^* is fixed, and since we want to avoid the small vectors $\mathbf{h}$ because they are considered the most damaging, maximizing ℓ_t makes sense.

The Euclidean length ℓ_t of the shortest nonzero vector $\mathbf{h}$ is independent of its direction, whereas for the product norm (for ρ_t) the length of $\mathbf{h}$ tends to remain small when $\mathbf{h}$ is aligned with several of the axes and increases quickly when $\mathbf{h}$ is diagonal with respect to the axes. Entacher, Hellekalek, and L'Ecuyer (1999) have proved a relationship between ℓ_t and ρ_t which seems to support the use of ℓ_t. It says (roughly) that a large ℓ_t implies a large ρ_t, but not vice-versa (they provide an example where $\rho_t\sqrt{t} = \ell_t b^{t-1}$ for an arbitrary b):

PROPOSITION 6. *One has* $\rho_t^2 \geq \ell_t^2 - (t-1)$. *The reverse inequality is* $\rho_t^{1/t}\sqrt{t} \leq \ell_t$.

Another important argument favoring ℓ_t is that it can be computed much more quickly than ρ_t or $\mathcal{P}_\alpha(P_n)$, for moderate and large n. Finally, tight upper bounds are available on ℓ_t, of the form $\ell_t \leq \ell_t^*(n) = c_t n^{1/t}$, where the constants c_t can be found in Conway and Sloane (1988) and L'Ecuyer (1999b). One can then define a normalized figure of merit $\ell_t/\ell_t^*(n)$, which lies between 0 and 1 (the larger the better). A similar normalization can be defined for the $\mathcal{L}_p$-distance in general, using a lower bound on the distance provided by Minkowski's general convex body theorem (Minkowski 1911). Hickernell et al. (1999), for example, use this lower bound to normalize the $\mathcal{L}_1$-distance between the successive hyperplanes.

These quantities ℓ_t, ρ_t, $\mathcal{P}_\alpha(P_n)$, etc., measure the structure of the points in the t-dimensional space. In view of the fact that the low-dimensional projections often account for a large fraction of the variance in the ANOVA decomposition in real-life applications, it seems appropriate to examine more closely the structure of the low-dimensional projections $L_t(I)$. Let $L_t^*(I)$ be the dual lattice of $L_t(I)$, i.e., the projection of L_t^* over the subspace determined by the indices in I, let ℓ_I be the Euclidean length of the shortest nonzero vector $\mathbf{h}$ in $L_t^*(I)$, and $\ell_s = \ell_{\{1,\ldots,s\}}$ for $s \leq t$. This length is normalized by the upper bound $\ell_{|I|}^*(n)$. Because $\ell_s^*(n)$ increases with s, the effect of the normalization is to be more demanding regarding the distance between the hyperplanes when the cardinality of I decreases. Assume that L_t is fully projection-regular and dimension-stationary. Then we have $\ell_{\{i_1,\ldots,i_s\}} = \ell_{\{1,\, i_2-i_1+1,\, \ldots, i_s-i_1+1\}}$, and it suffices to compute ℓ_I only for the sets I whose first coordinate is 1.

For arbitrary positive integers $t_1 \geq \cdots \geq t_d \geq d$, consider the worst-case figure of merit

$$M_{t_1,\ldots,t_d} = \min \left[\min_{2 \leq s \leq t_1} \ell_s / \ell_s^*(n), \min_{2 \leq s \leq d} \min_{I \in S(s,t_s)} \ell_I / \ell_{|I|}^*(n) \right], \tag{22}$$

where $S(s,t_s) = \{I = \{i_1,\ldots,i_s\} : 1 = i_1 < \cdots < i_s \leq t_s\}$. This figure of merit takes into account the low-dimensional projections and makes sure that the lattice is good not only in t dimensions, but also in projections over s successive dimensions for all $s \leq t_1$ (usually with $t_1 = t$) and over non-successive dimensions that are not too far apart. This means, to a certain extent, that we can use the same rule independently of the dimension of the problem at hand. In contrast, lattice rules provided in previous tables are typically chosen for a *fixed* t (i.e., different rules are suggested for different values of t, e.g., Sloan and Joe 1994).

The figure of merit $M_{t_1} = \min_{2 \leq s \leq t_1} \ell_s / \ell_s^*(n)$, obtained by taking $d = 1$, has been widely used for ranking and selecting LCGs as well as multiple recursive generators (Fishman 1996, L'Ecuyer 1999a). Tables of good LCGs with respect to this figure of merit, and which can be used as Korobov lattice rules, have been computed by L'Ecuyer (1999b) for $t_1 = 8, 16, 32$, for values of n that are either powers of 2 or primes close to powers of 2. These rules are good *uniformly* for a range of values of s. In Lemieux and L'Ecuyer (1999b), we suggested using $d = 2$ or 3 instead of $d = 1$, with $t_1 = \cdots = t_d$, and gave examples where it makes an important difference in the variance of the estimator $\tilde{Q}_n$. The quantity $M_{t_1,\ldots,t_d}$ is a worst case over $(t_1 - d) + \sum_{s=2}^{d} \binom{t_s - 1}{s-1}$ projections, and this number increases quickly with d unless the t_s are very small. For example, if $d = 4$ and $t_s = t$ for each s, there are 587 projections for $t = 16$ and 5019 projections for $t = 32$. When too many projections are considered, there are inevitably some that are bad, so the worst-case figure of merit is (practically) always small. As a consequence, the figure of merit can no longer distinguish between good and mediocre behavior in the most important projections. Moreover, the time to compute $M_{t_1,\ldots,t_d}$ increases with the number of projections. There is therefore a compromise to be made: We should consider the projections that we think have more chance of being important, but not too many of them. We suggest using the criterion (22) with d equal to 4 or 5, and t_s decreasing with s, both for QMC and for selecting random number generators.

Table 1 gives some values obtained by exhaustive search for the best multipliers a that are primitive element modulo n, for the largest prime numbers n less than certain powers of 2, in terms of the criterion $M_{t_1,\ldots,t_d}$ for selected values of $d, t_1, \ldots, t_d$ given in the table. The last line of the table gives the number of projections considered by each criterion. A star (*) adjacent to the criterion value means that this value is optimal (the best we found) with respect to this criterion. For each parameter set $(d, t_1, \ldots, t_d)$, we give an optimal multiplier a, its optimal criterion value, and also its criterion value for the other parameter sets. Some of the best rules with respect to M_{32} are bad with respect to the criteria that look at projections over non-successive dimensions (e.g., for $n = 8191$ and 131071). The best ones with respect to $M_{32,24,12,8}$ have a relatively good value of M_{32} and are usually good also with respect to $M_{32,24,16,12}$. Of course, since $M_{32,24,16,12}$ looks at the largest number of projections among the three criteria, the best LCGs with respect to this criterion are never bad with respect to the two other criteria.

Table 1: Best a's with respect to $M_{t_1,\ldots,t_d}$ for certain values of $(d, t_1, \ldots, t_d)$ and n.

n	a	M_{32}	$M_{32,24,12,8}$	$M_{32,24,16,12}$
1021	331	0.61872*	0.09210	0.09210
	76	0.53757	0.29344*	0.21672
	306	0.30406	0.26542	0.26542*
2039	393	0.65283*	0.15695	0.15695
	1487	0.49679	0.32196*	0.17209
	280	0.29807	0.25156	0.25156*
4093	219	0.66150*	0.13642	0.13642
	1516	0.39382	0.28399*	0.20839
	1397	0.40722	0.27815	0.27815*
8191	1716	0.64854*	0.05243	0.05243
	5130	0.50777	0.30676*	0.10826
	7151	0.47395	0.28809	0.28299*
16381	665	0.65508*	0.15291	0.14463
	4026	0.50348	0.29139*	0.23532
	5693	0.52539	0.26800	0.25748*
32749	9515	0.67356*	0.29319	0.13061
	14251	0.50086	0.32234*	0.12502
	8363	0.41099	0.29205	0.28645*
65521	2469	0.63900*	0.17455	0.06630
	8950	0.55678	0.34307*	0.20965
	944	0.39593	0.28813	0.26280*
131071	29803	0.66230*	0.03137	0.03137
	28823	0.44439	0.33946*	0.15934
	26771	0.54482	0.29403	0.29403*
No. of projections		31	141	321

7 Rules of Higher Rank

Rules of rank $r > 1$ have been studied by Sloan and Joe (1994) and the references given there. A special case is the *copy rule*, constructed as follows. Divide each of the first r axes of $[0,1)^t$ in η equal parts, partitioning thus the unit hypercube into η^r rectangles of equal volume. Take a rank-1 integration lattice whose node set has cardinality ν, rescale its first r axes so that $[0,1)^t$ is mapped to $[0,1/\eta)^r \times [0,1)^{t-r}$, and make one copy of the rescaled version into each rectangle of the partition. The node set thus obtained has cardinality $n = \nu\eta^r$, and corresponds to a lattice rule called an η^r-*copy rule*. The interest for

these rules stems from the fact that for a fixed value of n, the *average value* of $\mathcal{P}_\alpha(P_n)$ over η^r-copy rules is minimized by taking $r = t$ and $\eta = 2$. Sloan and Joe (1994) made computer searches for good rules in terms of $\mathcal{P}_\alpha(P_n)$ and the best rank-t rules that they found were generally better than their best rank-1 rules, for the same n. Our experiments confirmed this (see our forthcoming Table 2). These results no longer hold, however, if $\mathcal{P}_\alpha(P_n)$ is replaced by another criterion, such as $\tilde{\mathcal{P}}_{\alpha,p}(P_n)$ with unequal weights. This is especially true if the weights are chosen to make the low-dimensional projections more important. For example, if $\beta_1 = \cdots = \beta_t = \beta$ and β is small enough, the optimal η is 1 (Hickernell 1998b).

The limitations of copy rules over low-dimensional projections are easily understood by observing that the node sets of these rules are not projection-regular. For an η^r-copy rule, if $I = \{i_1, \ldots, i_s\} \subseteq \{1, \ldots, r\}$, the projection $P_n(I)$ contains only n/η^{r-s} distinct points. There are exactly η^{r-s} points of P_n projected onto each point of $P_n(I)$. For example, if $n = 2^{18}$ and $r = t = 16$, any unidimensional projection of P_n contains only 8 distinct points repeated 2^{15} times each, any 2-dimensional projection contains 16 distinct points repeated 2^{14} times each, and so on. Such rules are certainly bad sampling schemes in general if the low-dimensional projections of f account for most of the variance in its ANOVA decomposition (e.g., if f is nearly quadratic). As another special case, if we take $r = t$ and $\nu = 1$, so $n = \eta^t$, we obtain a *rectangular rule*, where L_t is the set of all t-dimensional points whose coordinates are multiples of $1/\eta$.

In Table 2, for $t = 12$, we compare the best 2^t-copy rules found by Sloan and Joe (1994) based on criterion $\mathcal{P}_2(P_n)$ (those of rank 12 in the table, with the corresponding ν), and the best rank-1 rules of corresponding orders that we found with criteria $\mathcal{P}_2(P_n)$, $S_{12} = \ell_{12}/\ell^*_{12}$, M_{12}, and $M_{12,8,6}$. For each rule, we give the total number of points n, the value of a, and the value of each criterion. For copy rules, a formula for computing $\mathcal{P}_2(P_n)$ is given by Sloan and Joe (1994), page 107 and Hickernell (1998b), page 150. To compute ℓ_s/ℓ^*_s, we use the fact that for a copy rule of rank t, $\ell_s = \eta\ell_s(\nu)$ and $\ell^*_s = c_s(\nu\eta^t)^{1/s} = \eta^{t/s}\ell^*_s(\nu)$ where $\ell^*_s(\nu)$ and $\ell_s(\nu)$ are the values of ℓ^*_s and ℓ_s for the rank-1 rule of order ν that has been copied. This gives $\ell_s/\ell^*_s = \eta^{1-t/s}\ell_s/\ell^*_s(\nu)$.

Our results agree with the theory of Sloan and Joe (1994): The copy rules of rank 12 have much better values of $\mathcal{P}_2(P_n)$ than the best rank-1 rules. In additional experiments, we found that by going from the best rank-1 rules to the best rank-t rules, the value of $\mathcal{P}_2(P_n)$ improves by a factor that increases with the dimension t. This factor is approximately 1.5 for $t = 4$, 3.2 for $t = 8$, and 6.5 for $t = 12$. The best copy rules of rank $t = 12$ in the table also happen to have a very good value for S_{12} (sometimes better than the best rank-1 rule with respect to S_{12}). However, the copy rules perform very poorly with respect to M_{12} and $M_{12,8,6}$, as expected, because their lower-dimensional projections are bad. It may be interesting to note that if we compare the best rules of rank 1 with respect to $\mathcal{P}_2(P_n)$ with the best rules with respect to S_{12} in the table, the latter perform much better with respect to the two criteria M_{12} and $M_{12,8,6}$.

Table 2: Copy Rules Versus Rank-1 Rules for $t = 12$

rank	criterion	ν	n	a	$\mathcal{P}_2(P_n)$	S_{12}	M_{12}	$M_{12,8,6}$
12	$\mathcal{P}_2(P_n)$	3	12288	1	447*	0.8097*	0.0237	0.0237
1	$\mathcal{P}_2(P_n)$	12281	12281	3636	2930	0.6401	0.0863	0.0187
1	S_{12}	12281	12281	86	3140	0.7574	0.4859	0.3684
1	M_{12}	12281	12281	9948	3160	0.7012	0.6683*	0.1202
1	$M_{12,8,6}$	12281	12281	657	3160	0.6402	0.6031	0.5804*
12	$\mathcal{P}_2(P_n)$	5	20480	2	268*	0.7759	0.0291	0.0184
1	$\mathcal{P}_2(P_n)$	20479	20479	11077	1730	0.6134	0.0728	0.0145
1	S_{12}	20479	20479	54	1900	0.7760*	0.3512	0.3109
1	M_{12}	20479	20479	14700	1900	0.7258	0.6915*	0.2085
1	$M_{12,8,6}$	20479	20479	10741	1880	0.7258	0.5398	0.5398*
12	$\mathcal{P}_2(P_n)$	11	45056	3	121*	0.7266	0.0277	0.0124
1	$\mathcal{P}_2(P_n)$	45053	45053	4928	806	0.6293	0.2334	0.0613
1	S_{12}	45053	45053	87	863	0.7707*	0.3722	0.3722
1	M_{12}	45053	45053	26149	853	0.7266	0.6874*	0.1053
1	$M_{12,8,6}$	45053	45053	5845	857	0.6293	0.5558	0.5542*

8 Polynomial Lattice Rules

The lattice rules discussed so far are based on integration lattices in $\mathbb{R}^t$. This is not the only possibility; one can define lattice rules based on lattices in other spaces. Consider for example the space $\mathbb{F}_2[z]$ of polynomials with coefficients in $\mathbb{F}_2$, the finite field with 2 elements (that is, each coefficient is either 0 or 1 and the arithmetic between the coefficients is performed modulo 2; e.g., Lidl and Niederreiter 1986). Let $P(z) = \sum_{j=0}^{k} a_j z^{k-j} \in \mathbb{F}_2[z]$ be a polynomial of degree k, with $a_k = a_0 = 1$, and consider the linear recurrence

$$p_i(z) = z p_{i-1}(z) \bmod (P(z), 2), \tag{23}$$

where

$$p_i(z) = \sum_{j=1}^{k} c_{i,j} z^{k-j} \tag{24}$$

is a polynomial in $\mathbb{F}_2[z]$, and "mod $(P(z), 2)$" means the remainder of the polynomial division by $P(z)$, with the operations on the coefficients performed in $\mathbb{F}_2$. We now have an LCG in $\mathbb{F}_2[z]$, with modulus $P(z)$ and multiplier z, which has a lattice structure similar to that of usual LCGs (Couture, L'Ecuyer, and Tezuka 1993, Couture and L'Ecuyer 1999). This LCG has maximal period $2^k - 1$ if and only if $P(z)$ is a primitive polynomial over $\mathbb{F}_2$ (Lidl and Niederreiter 1986). The *dual lattice* is the space $\mathcal{L}_t^*$ of multivariate polynomials $\mathbf{h}(z) = (h_1(z), \ldots, h_t(z))$, where $h_s(z) = \sum_{j=0}^{\ell-1} h_{s,j} z^j$, $h_{s,j} \in \mathbb{F}_2$, $\ell \in \mathbb{N}$, and such that

$\sum_{s=1}^{t} h_s(z) z^{s-1} \bmod (P(z), 2) = 0$. Define the "*length*" of $\mathbf{h}(z)$ by $\|\mathbf{h}(z)\| = 2^{\ell_*}$ where $\ell_* = \min\{\ell \geq 0 \mid h_{s,j} = 0 \text{ for all } j \geq \ell \text{ and all } s\}$.

To the polynomial $p_i(z)$, we associate the output value

$$u_i = \sum_{j=1}^{L} y_{i,j} 2^{-j} \tag{25}$$

where L is a positive integer,

$$y_{i,j} = \sum_{l=1}^{k} b_{j,l} c_{i,l} \bmod 2, \tag{26}$$

and the $b_{j,l}$'s are fixed bits forming an $L \times k$ matrix B. The corresponding node set P_n is the set of all vectors $\mathbf{u} = (u_0, \ldots, u_{t-1})$ obtained by taking each of the $n = 2^k$ possibilities for $p_0(z)$ in (24). The bit matrix B should be chosen so that P_n has good uniformity properties and is easy to enumerate. The node set P_n can be randomly shifted by adding a (uniform) random point $\mathbf{u}$ modulo 1, as in Section 4. However, as pointed out to us by R. Couture, the counterpart of the Cranley-Patterson rotation here is to perform a bitwise exclusive-or between the binary expansions of $\mathbf{u}$ and each point of P_n. This yields a randomly scrambled version of P_n, say $\tilde{P}_n$. This randomization of P_n is much simpler than the scrambling proposed by Owen (1997) for nets, and permits one to obtain simple variance expressions. In particular, if we expand the function f as a Walsh series in base 2, with coefficients $\tilde{f}(\mathbf{h}(z))$ where $\mathbf{h}(z)$ runs over the set of multivariate polynomials defined previously, the integration error with P_n and the variance of the estimator with $\tilde{P}_n$ are equal to $\sum_{\mathbf{0} \neq \mathbf{h}(z) \in \mathcal{L}_t^*} \tilde{f}(\mathbf{h}(z))$ (if this series converges absolutely) and $\sum_{\mathbf{0} \neq \mathbf{h}(z) \in \mathcal{L}_t^*} |\tilde{f}(\mathbf{h}(z))|^2$ (if f is square-integrable), respectively (Couture, L'Ecuyer, and Lemieux 1999). This motivates selection criteria for $P(z)$ and B based on the idea of avoiding (again) short vectors $\mathbf{h}(z)$ in the dual lattice, similar to what we discussed in Section 6.

The polynomial lattices defined via (23) are a special case of more general polynomial integration lattices of higher rank which can be defined via a polynomial version of (7). These polynomial lattice rules are in fact equivalent to the digital net constructions of Niederreiter (1992), Sect. 4.4, also discussed by Larcher (1998).

The dual lattice turns out to be useful for studying the following equidistribution properties. By partitioning the interval $[0, 1)$ into 2^ℓ segments of equal length, we determine a partition of the box $[0, 1)^t$ into $2^{t\ell}$ cubic boxes of equal volume. For a given set of indices $I = \{i_1, \ldots, i_s\}$, we say that the projection $P_n(I)$ is *s-distributed to ℓ bits of accuracy* if each box of the partition contains exactly $2^{k-s\ell}$ 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 $2^{s\ell}$ possible $s\ell$-bit strings appears exactly the same number of times. (Of course, this requires $s\ell \leq k$.) To verify this property, it suffices to write a system of linear equations that express these $2^{s\ell}$ bits as a function of $(c_{0,1}, \ldots, c_{0,k})$, and to check that these equations are independent, i.e., that the corresponding matrix has full rank, $s\ell$. The link with the dual lattice is that $P_n(I)$ is s-distributed to ℓ bits of accuracy if and only if $\ell \leq L$ and no nonzero vector in $\mathcal{L}_s^*(I)$

has length 2^ℓ or less, where $\mathcal{L}_s^*(I)$ is the projection of $\mathcal{L}_t^*$ over the subspace determined by I (Couture, L'Ecuyer, and Tezuka 1993). The property can thus be verified by computing the length of the shortest vector in the dual lattice, i.e., by applying the spectral test in the polynomial space (Couture and L'Ecuyer 1999).

To define a criterion, denote by $2^{\ell^*(I)}$ the length of the shortest nonzero vector $\mathbf{h}(z)$ in $\mathcal{L}_t^*(I)$. For positive integers d and $t_1 \geq \cdots \geq t_d$, let $\mathcal{J}(t_1, \ldots, t_d)$ be the class of subsets I such that either $I = \{1, \ldots, s\}$ for $s \leq t_1$, or $I = \{i_1, \ldots, i_s\}$ where $2 \leq s \leq d$ and $1 = i_1 < \cdots < i_s \leq t_s\}$. Define

$$\Delta(t_1, \ldots, t_d) = \max_{I \in \mathcal{J}} \max\left[0,\, \min(L,\, \lfloor k/|I| \rfloor) - \ell^*(I)\right]. \tag{27}$$

We say that the point set P_n is ME$(t_1, \ldots, t_d)$ (where ME stands for *maximally equidistributed*) if $\Delta(t_1, \ldots, t_d) = 0$. This is an extension of the criterion used by L'Ecuyer (1996, 1999c), who selected combined Tausworthe random number generators (these generators turn out to be a special case of (23)–(26)) based on the criterion ME(k). A related criterion 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. The set P_n is a (q, k, t)-net in base 2 if each box of each of these partitions contains exactly 2^q points (Niederreiter 1992 provides the details). 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 large number of partitions. Each box of the partition into $\prod_{j=1}^t 2^{\ell_j}$ rectangles contains exactly 2^q points if and only if $\mathcal{L}_t^*$ contains no vector $(h_1(z), \ldots, h_t(z)) \neq \mathbf{0}$ such that $h_{s,j} = 0$ for all $j \geq l_s$ and $1 \leq s \leq t$. This condition can also be verified by expressing the $\ell_1 + \cdots + \ell_t$ bits that determine in which rectangle the point falls as a linear combination of the bits of the initial state, and checking if the corresponding matrix has full rank.

9 Massaging the Problem

When the function f is fixed, the goal is to find an integration lattice whose dual contains the most important Fourier coefficients. Another way of gaining precision is to change f so that its integral remains the same but its most important Fourier coefficients correspond to vectors $\mathbf{h}$ that are smaller and/or belong to lower-dimensional projections.

A first way of achieving this is to improve the smoothness of $\bar{f}$, the periodic continuation of f, by making nonlinear changes of variables of the form $v_s = \phi_s^{-1}(u_s)$, where $\phi_s : [0, 1) \to [0, 1)$ is smooth and increasing for each s. The integral becomes

$$\mu = \int_{[0,1)^t} g(\mathbf{v}) d\mathbf{v}$$

where $g(\mathbf{v}) = g(v_1, \ldots, v_t) = f(\phi_1(v_1), \ldots, \phi_t(v_t))\phi_1'(v_1) \cdots \phi_t'(v_t)$. By choosing each ϕ_s so that $\phi_s'(0) = \phi_s'(1) = 0$, the periodic continuation of g becomes continuous on the hypercube boundary even if that of f is not. More generally, if the $(d+1)$th derivative of ϕ vanishes on the hypercube boundary, the periodic continuation of g is guaranteed to have a

continuous dth derivative on the boundary. For example, if $\phi(v) = v^3(10-15v+6v^2)$, then both ϕ' and ϕ'' vanish at 0 and 1. These transformation techniques are further discussed in Section 2.12 of Sloan and Joe (1994). These methods should not be applied blindly. A transformation that improves smoothness at the boundary may substantially increase σ^2, the variance of f, e.g., by introducing oscillations inside the hypercube. Finding appropriate ϕ_s's can be hard in practice.

Other types of transformations work by reducing the effective dimension of the problem, by concentrating the variance in the ANOVA decomposition on the σ_I^2's for which I contains only a few small coordinates, or for which I contains only a few coordinates that are close to each other, or something of that kind. That is, concentrating the variance on the subspaces for which the projections $P_n(I)$ are known to have very good uniformity. These methods include the Brownian bridge technique for generating a Brownian motion, special techniques for generating Poisson processes, methods based on principal components analysis, and so on. We refer the reader to Fox (1999). Here we just briefly outline the idea of the Brownian bridge method, which will be used in Section 10.2.

Suppose one has to generate the path of a standard Brownian motion $\{W(\zeta),\, 0 \le \zeta \le T\}$ (with zero trend and variance constant of 1). The standard way is to discretize the time by putting, say, $\zeta_i = i\delta$ for $i = 0, \ldots, t$, where $\delta = T/t$, and then generate $Z_i = (W(\zeta_i) - W(\zeta_{i-1}))/\sqrt{\delta}$, $i = 1, \ldots, t$, which are i.i.d. $N(0,1)$ random variables. If the standard normals are generated by inversion, this requires t uniforms. If the function f is some sort of average over the entire trajectory of W, for instance, then the uniforms used for the early part of the trajectory are slightly more important than those used near the end, because their effect lasts longer. However, the first few uniforms can be made much more important as follows. Generate first $W(T)$, a normal with mean 0 and variance T. Then generate $W(T/2)$, whose distribution conditional on $W(0)$ and $W(T)$ is normal with mean $(W(0) + W(T))/2$ and variance $T/4$, according to the *Brownian bridge formula* (Karatzas and Shreve 1988). By applying the technique recursively, one generates successively $W(T/4)$, $W(3T/4)$, $W(T/8)$, $W(3T/8)$, and so on. The first few values are now very important because they draw a rough sketch of the entire trajectory of W, whereas the values generated later only make minor adjustments to the trajectory. Extensions of this method lead to principal components analysis and other variants, which have been applied successfully in the area of finance (e.g., Acworth, Broadie, and Glasserman 1997, Morokoff 1998).

10 Examples

In the following examples, the random variables are always generated by inversion, so that the dimension t for each problem is equal to the number of random variables that must be generated in one simulation run. For all the examples, we use the lattice rules that minimize the criterion $M_{32,24,12,8}$ in Table 1.

10.1 A Stochastic Activity Network

This example is taken from Avramidis and Wilson (1996). We consider a *stochastic activity network* (SAN), represented by a directed acyclic graph $(\mathcal{N}, \mathcal{A})$, where $\mathcal{N}$ is a set of *nodes* which contains one source and one sink, and $\mathcal{A}$ is a set of arcs corresponding to *activities*. Figure 2 gives an illustration. Each activity $k \in \mathcal{A}$ has a random duration V_k with distribution function $F_k(\cdot)$. Certain *dummy* activities represent precedence relationships and have a duration of 0. We denote by $N(A)$ the number of activities with nonzero duration, $N(P)$ the number of directed paths from the source to the sink, and $C_j \subseteq \mathcal{A}$ the set of activities forming the path j, for $1 \leq j \leq N(P)$. The *network completion time* T is the length of the longest path from the source to the sink.

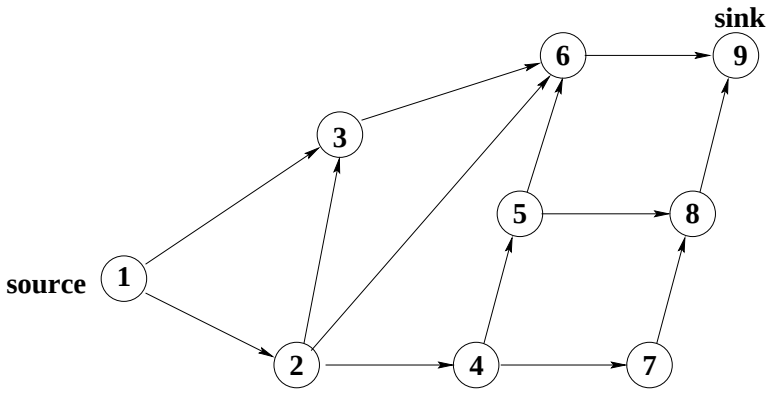

Figure 2: Example of a SAN, taken from Avramidis and Wilson (1996)

We want to estimate $\mu = F_T(x) = P[T \leq x]$ for a given threshold x. With the standard MC or QMC method, this problem has $t = N(A)$ dimensions, since one uniform u_k is needed to generate each activity duration, via $V_k = F_k^{-1}(u_k)$. One can write μ as the integral

$$\mu = F_T(x) = \int_{[0,1)^{N(A)}} \prod_{j=1}^{N(P)} \mathbf{1}\left[\sum_{k \in C_j} F_k^{-1}(u_k) \leq x\right] du_1 \dots du_{N(A)}$$

where $\mathbf{1}$ is the indicator function. Both the dimension of the problem and the variance of the MC estimator can be reduced by applying *conditional Monte Carlo* (CMC), as follows (Avramidis and Wilson 1996). Select a set of activities $\mathcal{L} \subseteq \mathcal{A}$ such that each directed path j from the source to the sink contains exactly one activity l_j from $\mathcal{L}$. This set is called a *uniformly directed cutset*. The idea of QMC is to generate (by simulation) only the durations of the activities in $\mathcal{B} = \mathcal{A} \setminus \mathcal{L}$, and to estimate μ by the conditional probability that $T \leq x$ given those durations. The dimension of the problem is now reduced to $t = N(B)$, where $N(B)$ is the number of non-dummy activities in $\mathcal{B}$. The CMC (unbiased) estimator is

$$Y = P[T \le x \mid \{V_j,\, j \in \mathcal{B}\}] = \prod_{l \in \mathcal{L}} F_l \left[\min_{\{j=1,\ldots,N(P):l_j=l\}} \left(x - \sum_{k \in C_j \setminus \{l_j\}} V_k \right) \right].$$

The t required uniforms for each replication can now be generated either by standard MC or by (randomized) QMC, e.g., via a lattice rule. Avramidis and Wilson (1996) proposed to generate them via Latin Hybercube Sampling (LHS). Note that this setup and methodology applies to estimate the expected length of the longest path in a network in general; it does not have to be a SAN.

Table 3: Estimated variance reduction factors w.r.t. MC for the SAN example

		n		
Method	t	4093	16381	65521
LHS	13	3.2	4.3	3.4
LR	13	6.2	4.2	24.5
MC+CMC	8	4.1	4.1	4.1
LHS+CMC	8	58	56	63
LR+CMC	8	268	839	3086

We performed experiments with the network shown in Figure 2, with the same set $\mathcal{L}$ and the same probability laws of the activity durations as in Avramidis and Wilson (1996), to compare MC, LHS, and a randomly shifted lattice rule (LR), with and without CMC. The set $\mathcal{L}$ contains the 5 arcs that separate the nodes $\{1, 2, 3, 4, 5\}$ from the nodes $\{6, 7, 8, 9\}$. The dimension of the problem is thus $t = 13$ without CMC and $t = 8$ with CMC. We took $x = 90$, which implies $F_T(x) \approx 0.89$. For LR, we used different number of points n, and $m = 100$ random shifts. We made mn i.i.d. replications for MC, for a fair comparison. Table 3 gives the estimated variance reduction factors with respect to the crude MC estimator.

The combination of LR with CMC (last line) is a clear winner here. Moreover, the corresponding variance reduction factor increases almost linearly with n. Its computing time is also *less* than MC for an equivalent total sample size, since both LR and CMC reduce the work in addition to reducing the variance. The combination of LHS with CMC reduces the variance by a non-negligible factor, but this factor is practically independent of n. We performed other experiments with different values of x and with the other network presented in Avramidis and Wilson (1996), and the conclusions were similar.

10.2 Pricing Asian Options

Consider the problem of pricing an asian option on the arithmetic average, for a single asset whose value at time u is denoted by $S(u)$. We assume the Black–Scholes model for the evolution of $S(\cdot)$, with risk-free appreciation rate r, volatility σ, strike price K, and

expiration time T. Under the so-called risk-neutral measure, $S(\cdot)$ obeys the Itô stochastic differential equation

$$dS(\zeta)/S(\zeta) = r d\zeta + \sigma dB(\zeta)$$

where $B(\cdot)$ is a standard Brownian motion. (Details about this model can be found, e.g., in Duffie 1996.) The solution of this equation is

$$S(\zeta) = S(0) \exp\left[(r - \sigma^2/2)\zeta + \sigma B(\zeta)\right].$$

The final value of the option is given by $\max(0,\ (1/t)\sum_{i=1}^{t} S(t_i) - K)$, where $t_i = iT/t$ and t is a fixed constant. The trajectory of $B(\cdot)$ can be generated as described in Section 9, by generating t i.i.d. standard normals. The expected final value, which is the fair price that we want to estimate, can in fact be written as the t-dimensional integral:

$$\mu = \int_{[0,1)^t} \max\left(0,\ \frac{1}{t}\sum_{i=1}^{t} S(0) \exp\left[(r - \sigma^2/2)t_i + \sigma\sqrt{T/t}\sum_{j=1}^{i}\Phi^{-1}(u_j)\right] - K\right) du_1 \dots du_t,$$

where $\Phi(\cdot)$ is the standard normal distribution.

To reduce the variance, one can use the selling price of the option on the *geometric* average as a control variable (Kemna and Vorst 1990) as well as antithetic variates. Numerical results combining these methods with shifted lattice rules are given by Lemieux and L'Ecuyer (1998, 1999a). Glasserman, Heidelberger, and Shahabuddin (1999) also use importance sampling (IS) and stratification (STR) to reduce the variance for this problem. IS and STR are used to generate the product $Y = \mathbf{a} \cdot (Z_1, \dots, Z_t)$, where $\mathbf{a}$ is some "optimal" vector and $Z_1, \dots, Z_t$ are t i.i.d. standard normals. Each Z_i is then generated by conditioning on Y, so the problem thus becomes $(t+1)$-dimensional.

We performed experiments to compare different combinations of the above methods, and their coupling with shifted lattice rules. We denote by CMC the method that generates the Z_i's by conditioning on Y, with $\mathbf{a}$ equal to the optimal drift vector for IS as suggested by Glasserman, Heidelberger, and Shahabuddin (1999), and we apply IS and STR in exactly the same way as these authors (this IS is always combined with CMC). When we combine CMC with LR, we take a rule in $t+1$ dimensions and use the first coordinate of each shifted point to generate the product $\mathbf{a} \cdot (Z_1, \dots, Z_t)$. The Brownian bridge technique is denoted by BB.

Table 4 reports the estimated variance reduction factors with respect to MC, for certain combinations of the methods. The parameters of the option are $\sigma = 0.3$, $r = 0.05$, $K = 55$, $T = 1$ year and $t = 64$. Among the combinations given in the table (and all others that we tried), the winner is CV+IS+CMC+LR. It improves over the MC+IS+CMC+STR combination of Glasserman, Heidelberger, and Shahabuddin (1999) by a factor of approximately 4. One can also observe that CV+LR, which is very simple and easy to implement, already does a decent job. Combining it with BB brings a significant improvement, and adding IS brings another small gain. Our additional experiments with the pricing of Asian options indicated that the effectiveness of CV generally decreases with K and with t,

Table 4: Estimated variance reduction factors w.r.t. MC for the Asian-option example

	n		
Method	4093	16381	65521
MC+IS+CMC+STR	1495	1601	1603
CV+LR	703	620	597
BB+CV+LR	2488	4876	4958
BB+CV+IS+LR	3129	4790	5407
CV+IS+CMC+LR	6073	6206	6847

whereas the effectiveness of IS increases with K (as explained by Glasserman, Heidelberger, and Shahabuddin (1999)). Otherwise, the results were similar to those of Table 4. L'Ecuyer and Lemieux (1999) report preliminary numerical experiments with polynomial lattice rules for the present example.

10.3 A Single Queue

Consider an M/M/1 queue with arrival rate λ and service rate μ. We want to estimate the steady-state probability $p(k)$ that a customer has its sojourn time in the queue larger than k, for some constant k. Simulation is unnecessary for this problem, since it is known that $p(k) = e^{-k\mu(1-\lambda/\mu)}$. However, this simple example allows us to illustrate how lattice rules can be used for infinite-horizon models and how it can be coupled with regenerative simulation. Lindley's equation tells us that

$$T_{i+1} = S_{i+1} + \max(0,\, T_i - A_i)$$

where T_i and S_i are respectively the sojourn time and service time of customer i and A_i is the interarrival time between customers i and $i+1$. We assume that $T_0 = A_0$, so $T_1 = S_1$. The discrete-time process $\{T_i,\, i \geq 0\}$ is a regenerative process with a regeneration epoch at each index i for which $T_i - A_i \leq 0$.

A first approach to estimate $p(k)$ uses a large *truncated horizon*: Simulate a fixed number of customers (say, N, where N is large) and take the average

$$\frac{1}{N}\sum_{i=1}^{N} \mathbf{1}\left(T_i > k\right).$$

This can be replicated a certain number of times, independently, to estimate the variance and compute a confidence interval. It we use 2 uniforms for each customer, one to generate its arrival time and one for its service time, we have a $2N$-dimensional integration problem, for which we can use a $2N$-dimensional lattice rule. (We run a truncated-horizon simulation with each of the n points of the rule.) If we perform m independent random shifts of the rule, we thus simulate a grand total of mnN customers. A second approach is to simulate a

fixed number n of regenerative cycles, using one point from the lattice node set to simulate each regenerative cycle. The dimension t of the problem, which is now twice the number of customers in a cycle, is a random variable with mean $2/(1-\lambda/\mu)$. One can also view the problem as infinite-dimensional, with all but a finite (random) number of the uniforms being unused. Both the truncated horizon and the regenerative method provide biased estimators of $p(k)$. Here, we are not interested in this bias, but only in the variance reduction obtained by applying randomly-shifted lattice rules.

We tried the truncated-horizon estimator on an example with parameter values $\lambda/\mu = 0.6$, $k = 10$ and 20, and $N = 5000$ (so the number of dimensions is $t = 10000$). By using the lattice rule of $n = 1021$ points, the variance was reduced by a factor ranging between 5 and 10 compared with MC. Note that in this example we use nearly 10 times the period length of the LCG (5) to generate each lattice point $\mathbf{u}_i$ (i.e., $t \approx 10n$). However, as explained earlier, the coordinates $\tilde{u}_{i,j}$ of $\tilde{\mathbf{u}}_i$ are not periodic, thanks to the random shift, and the fact that $t \gg n$ poses no difficulty. Moreover, for this model, customers that are far apart in time are almost independent, which means that the important σ_I's in the ANOVA decomposition are those for which $i_d - i_1$ is small, assuming that $I = \{i_1, \ldots, i_d\}$ where $i_1 < \cdots < i_d$. In other words, this problem has an effective dimension much less than $2N$ in the successive-dimensions sense. This is especially true if the traffic intensity λ/μ is small. The effective dimension increases with the traffic intensity, as does the average length of the regenerative cycles. We also tried the regenerative method on this example, with $n = 1021$, and obtained a variance reduction of approximately 3 compared with MC when $k = 10$ and 2 when $k = 20$. With $n = 65521$ points, these factors increased to 3.5 and 2.2, respectively. The variance reduction is less important here than with the truncated-horizon estimator: In the latter case, each simulation gives us a mean-value over many cycles (instead of only one for the regenerative method), and this averaging introduces a smoothness favorable to LR in the function f that corresponds to the integral of the form (1) that we try to estimate.

11 Conclusion

QMC is most often 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, where the concept of discrepancy is in the sense of the rectangular star discrepancy, and where the justification for QMC is based on the worst-case error bound provided by the Koksma–Hlawka inequality (4). Lattice rules, which are an alternative to this framework, have also been traditionally justified by worst-case error bounds. Viewing them as a variance reduction tool seems more practical, however, as we have argued in this paper. Our coverage of lattice rules is of course incomplete. For other viewpoints and results, we refer the reader to the book of Sloan and Joe (1994) and the recent papers of Hickernell.

The criterion $M_{t_1,\ldots,t_d}$ that we have proposed is not perfect, but it is convenient and it provides rules that seem to work well in practice. We admit that the choice of d and of the t_s's is arbitrary and that the corresponding function w in (13) cuts abruptly to zero once we hit the subspaces (or projections) that are not considered by the criterion. An

alternative would be to consider all subsets I for the minimization in (22), but to multiply the constants $\ell^*_{|I|}(n)$ by some weights that decrease smoothly towards 0 with the size and span of I (i.e., $|I|$ and $i_d - i_1$) so that the projections over coordinate sets with large size or span will not be taken into account unless they are really very bad. This smoother scheme could be more complicated to implement than the criterion (22), however, because a larger number of subsets I would have to be examined, and the choice of the weights is still arbitrary.

Among the interesting topics currently under investigation, we mention the concept of *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., $n_{i+1} = 2n_i$) and $P_{n_i} \subset P_{n_{i+1}}$ for each i. The idea is that if the empirical variance (or the other error estimate in use) is still larger than desired after applying the lattice rule with n_i points, one can switch to the lattice rule with n_{i+1} points (e.g., double the number of points) without discarding the work performed so far. One only needs to evaluate the function at the new points. With this kind of lattice sequence, the number of points in the lattice needs not be fixed in advance. To implement this concept, one needs to find a practical way of constructing such a sequence of embedded lattices so that each intermediate node set P_{n_i} is of good quality. Hickernell et al. (1999) have recently proposed one way of doing this. They provide concrete parameters and numerical illustrations.

For a given problem, a good lattice rule is one that kicks out of the dual lattice the most important squared Fourier coefficients in (16). The choice of the rule should therefore (ideally) 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, or on sums of certain bundles of squared coefficients. This deserves further investigation.

Acknowledgments

This work has been supported by NSERC-Canada grant No. ODGP0110050 to the first author and via an FCAR-Québec scholarship to the second author. We thank Raymond Couture, Peter Hellekalek, and Harald Niederreiter for their helpful suggestions and comments.

References

Acworth, P., M. Broadie, and P. Glasserman, "A Comparison of Some Monte Carlo and quasi-Monte Carlo Techniques for Option Pricing," In *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing*, ed. P. Hellekalek and H. Niederreiter, number 127 in Lecture Notes in Statistics, Springer-Verlag, 1997, 1–18.

Avramidis, A. N. and J. R. Wilson, "Integrated Variance Reduction Strategies for Simulation," *Operations Research*, 44 (1996), 327–346.

Caflish, R. E., W. Morokoff, and A. Owen, "Valuation of Mortgage-Backed Securities Using Brownian Bridges to Reduce Effective Dimension," *The Journal of Computational Finance*, 1, 1 (1997), 27–46.

Cochran, W. G., *Sampling Techniques*, second edition, John Wiley and Sons, New York, 1977.

Conway, J. H. and N. J. A. Sloane, *Sphere Packings, Lattices and Groups*, Grundlehren der Mathematischen Wissenschaften 290, Springer-Verlag, New York, 1988.

Couture, R. and P. L'Ecuyer, "Lattice Computations for Random Numbers," *Mathematics of Computation* (1999). To Appear.

Couture, R., P. L'Ecuyer, and C. Lemieux, "Polynomial Lattice Rules," 1999. In preparation.

Couture, R., P. L'Ecuyer, and S. Tezuka, "On the Distribution of k-Dimensional Vectors for Simple and Combined Tausworthe Sequences," *Mathematics of Computation*, 60, 202 (1993), 749–761, S11–S16.

Cranley, R. and T. N. L. Patterson, "Randomization of Number Theoretic Methods for Multiple Integration," *SIAM Journal on Numerical Analysis*, 13, 6 (1976), 904–914.

Davis, P. and P. Rabinowitz, *Methods of Numerical Integration*, second edition, Academic Press, New York, 1984.

Drmota, M. and R. F. Tichy, *Sequences, Discrepancies and Applications*, Lecture Notes in Mathematics, Springer Verlag, New York, 1997.

Duffie, D., *Dynamic Asset Pricing Theory*, second edition, Princeton University Press, 1996.

Efron, B. and C. Stein, "The Jackknife Estimator of Variance," *Annals of Statistics*, 9 (1981), 586–596.

Entacher, K., P. Hellekalek, and P. L'Ecuyer, "Quasi-Monte Carlo Node Sets from Linear Congruential Generators," In *Monte Carlo and Quasi-Monte Carlo Methods 1998*, Lecture Notes in Computational Science and Engineering, Springer-Verlag, 1999. To appear.

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

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

Glasserman, P., P. Heidelberger, and P. Shahabuddin, "Asymptotically Optimal Importance Sampling and Stratification for Pricing Path Dependent Options," *Journal of Mathematical Finance*, 9, 2 (1999), 117–152.

Hellekalek, P., "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*, Springer, New York, 1998, 49–108.

Hickernell, F. J., "A Generalized Discrepancy and Quadrature Error Bound," *Mathematics of Computation*, 67 (1998a), 299–322.

Hickernell, F. J., "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*, Springer, New York, 1998b, 109–166.

Hickernell, F. J., "What Affects Accuracy of Quasi-Monte Carlo Quadrature?," In *Monte Carlo and Quasi-Monte Carlo Methods 1998*, ed. H. Niederreiter and J. Spanier, Lecture Notes in Computational Science and Engineering, Springer-Verlag, 1999. To appear.

Hickernell, F. J. and H. S. Hong, "The Asymptotic Efficiency of Randomized Nets for Quadrature," *Mathematics of Computation*, 68 (1999), 767–791.

Hickernell, F. J., H. S. Hong, P. L'Ecuyer, and C. Lemieux, "Extensible Lattice Sequences for Quasi-Monte Carlo Quadrature," 1999. Submitted.

Hoeffding, W., "A Class of Statistics with Asymptotically Normal Distributions," *Annals of Mathematical Statistics*, 19 (1948), 293–325.

Karatzas, I. and S. Shreve, *Brownian Motion and Stochastic Calculus*, second edition, Springer-Verlag, New York, 1988.

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

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

Korobov, N. M., "The Approximate Computation of Multiple Integrals," *Dokl. Akad. Nauk SSSR*, 124 (1959), 1207–1210. In Russian.

Kuipers, L. and H. Niederreiter, *Uniform Distribution of Sequences*, John Wiley, New York, 1974.

Larcher, G., "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*, Springer, New York, 1998, 167–222.

Law, A. M. and W. D. Kelton, *Simulation Modeling and Analysis*, second edition, McGraw-Hill, New York, 1991.

L'Ecuyer, P., "Uniform Random Number Generation," *Annals of Operations Research*, 53 (1994), 77–120.

L'Ecuyer, P., "Maximally Equidistributed Combined Tausworthe Generators," *Mathematics of Computation*, 65, 213 (1996), 203–213.

L'Ecuyer, P., "Random Number Generation," In *Handbook of Simulation*, ed. J. Banks, Wiley, 1998, 93–137. Chapter 4.

L'Ecuyer, P., "Good Parameters and Implementations for Combined Multiple Recursive Random Number Generators," *Operations Research*, 47, 1 (1999a), 159–164.

L'Ecuyer, P., "Tables of Linear Congruential Generators of Different Sizes and Good Lattice Structure," *Mathematics of Computation*, 68, 225 (1999b), 249–260.

L'Ecuyer, P., "Tables of Maximally Equidistributed Combined LFSR Generators," *Mathematics of Computation*, 68, 225 (1999c), 261–269.

L'Ecuyer, P. and R. Couture, "An Implementation of the Lattice and Spectral Tests for Multiple Recursive Linear Random Number Generators," *INFORMS Journal on Computing*, 9, 2 (1997), 206–217.

L'Ecuyer, P. and C. Lemieux, "Quasi-Monte Carlo via Linear Shift-Register Sequences," In *Proceedings of the 1999 Winter Simulation Conference*, IEEE Press, 1999. To appear.

Lemieux, C., "L'amélioration de l'efficacité des estimateurs en simulation," PhD thesis, Université de Montréal, 1999. In preparation.

Lemieux, C. and P. L'Ecuyer, "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, IEEE Press, 1998, 579–586.

Lemieux, C. and P. L'Ecuyer, "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, Lecture Notes in Computational Science and Engineering, Springer-Verlag, 1999a. To appear.

Lemieux, C. and P. L'Ecuyer, "Selection Criteria for Lattice Rules and Other Low-Discrepancy Point Sets," 1999b. to appear.

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

Lyness, J. N. and I. H. Sloan, "Some properties of rank-2 lattice rules," *Mathematics of Computation*, 53, 188 (1989), 627–637.

Maisonneuve, D., "Recherche et Utilisation des "Bons Treillis", Programmation et Résultats Numériques," In *Applications of Number Theory to Numerical Analysis*, ed. S. K. Zaremba, Academic Press, New York, 1972, 121–201.

Minkowski, H., *Gesammelte Abhandlungen*, volume I and II, Teubner-Verlag, Leipzig, 1911.

Morokoff, W. J., "Generating Quasi-Random Paths for Stochastic Processes," *SIAM Review*, 40, 4 (1998), 765–788.

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

Niederreiter, H. and C. Xing, "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*, Springer, New York, 1998, 267–302.

Owen, A. B., "Scrambled Net Variance for Integrals of Smooth Functions," *Annals of Statistics*, 25, 4 (1997), 1541–1562.

Owen, A. B., "Latin Supercube Sampling for Very High-Dimensional Simulations," *ACM Transactions of Modeling and Computer Simulation*, 8, 1 (1998), 71–102.

Rudin, W., *Real and Complex Analysis*, second edition, McGraw-Hill, New York, 1974.

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

Sloan, I. H. and T. R. Osborn, "Multiple integration over bounded and unbounded regions," *Journal of Computational and Applied Mathematics*, 17 (1987), 181–196.

Sobol', I. M., "On Quasi-Monte Carlo Integration," *Mathematics and Computers in Simulation*, 47 (1998), 103–112.

Tuffin, B., "On the use of low-discrepancy sequences in Monte Carlo methods," *Computing*, 61 (1998), 371–378.

Woźniakowski, H., "Average Case Complexity of Multivariate Integration," *Bulletin of the American Mathematical Society*, 24 (1991), 185–194.