Latin Supercube Sampling for Very High Dimensional Simulations

Art B. Owen
Stanford University

June 1997

Abstract

This paper introduces Latin supercube sampling (LSS) for very high dimensional simulations, such as arise in particle transport, finance and queuing. LSS is developed as a combination of two widely used methods: Latin hypercube sampling (LHS), and Quasi-Monte Carlo (QMC). In LSS, the input variables are grouped into subsets, and a lower dimensional QMC method is used within each subset. The QMC points are presented in random order within subsets. QMC methods have been observed to lose effectiveness in high dimensional problems. This paper shows that LSS can extend the benefits of QMC to much higher dimensions, when one can make a good grouping of input variables. Some suggestions for grouping variables are given for the motivating examples. Even a poor grouping can still be expected to do as well as LHS. The paper also extends LHS and LSS to infinite dimensional problems. The paper includes a survey of QMC methods, randomized versions of them (RQMC) and previous methods for extending QMC to higher dimensions. Furthermore it shows that LSS applied with RQMC is more reliable than LSS with QMC.

1 Introduction

The fundamental problem we consider here is to compute the value of

$$I = \int_{[0,1)^d} f(X)dX.$$

This problem can include, via change of variable, integration over nonrectangular regions, integration with respect to non-uniform probability distributions, and reformulations designed to improve accuracy, such as importance

sampling or periodization. Some simulations described below can be cast in this form with $d = \infty$.

For small dimensions d and smooth integrands f, classical techniques such as those of Davis and Rabinowitz [6] provide an excellent solution. For moderate dimensions, iterated versions of the classical techniques may work well. These are justified by Fubini's theorem and are typically constructed as tensor products of univariate or low dimensional rules.

A working definition of a "high dimensional" problem, is one in which such iterated integration methods are computationally infeasible. This can depend somewhat on f and on the computer at hand, but large changes in computing power are required to make even small increases in the point at which dimension becomes large. For high dimensional problems, simulation methods such as Monte Carlo (MC) and equidistribution or Quasi-Monte Carlo (QMC) are used.

Theory has it that QMC is more accurate than MC at least for enormous simulations, and numerical experiments often find the advantage appears at practical sample sizes. But, the advantages of QMC have often been observed to diminish as d increases. See Morokoff and Caflisch [25, 26, 27] for a discussion of this point.

For a large enough dimension d, it becomes difficult or impossible to even construct a QMC point set, having meaningful equidistribution properties. A working definition of a "very high dimensional" problem, is one for which QMC constructions either are not practically computable or are not especially equidistributed for practical sample sizes.

The purpose of this article is to present LSS as a method, that in favorable cases restores the effectiveness of QMC for very high dimensional problems, while performing at least as well as MC in the unfavorable cases. Throughout this paper, d is the dimension of the original integration problem and s is the dimension of an integration rule, such as a QMC rule. The emphasis is on what to do when $d > s$ or even $d \gg s$.

Some examples of very high dimensional simulations are given next. This section then concludes with an outline of the article.

1.1 Examples of very high dimensional problems

1.1.1 Transport Simulations

Particle transport simulations are used in the design of radiation shields for nuclear reactors (see Spanier [44]). The trajectories of a large number of particles are simulated as they transit from a source towards a target

possibly undergoing a sequence of collisions. Each collision may take 6 or 7 random numbers to describe changes in the particles' positions, velocities and energies and whether they are absorbed. To simulate k collisions thus takes $d = 6k$ or $7k$. Since there is no a priori upper bound on k these problems can be considered infinite dimensional, even though any given particle only undergoes a finite number of collisions.

Light transport problems, arising in computer graphics, have a similar flavor. One can follow a photon from a light source until it splats on the imaging plane, possibly after a number of reflections. See Keller [21] and Guibas and Veach [50].

Solving Laplace's equation with a boundary constraint arises in the design of semiconductors [40]. One approach due to Kakutani [20] is to simulate Brownian paths from a point in the region until they reach the boundary. The solution is the average of the boundary values reached by the paths. The simulation proceeds in a series of steps each of which takes a particle to the edge of a bounding box within the boundary region. There is no a priori limit on the number of steps one particle might require, so again $d = \infty$. For more applications to microelectronics, see Kersch and Morokoff [22].

1.1.2 Financial valuation

The value of some financial derivatives depends on a whole time series of random fluctuations, in a way that cannot be replaced by a closed form expression. For example, collateralized mortgage obligations [4, 37, 38] depend on the interest rate at 360 future time periods (for monthly payments on 30 year mortgages) and on the fraction of mortgage holders prepaying in each of those 360 time periods. Thus $d = 360$ if the prepayment levels are modelled as a deterministic function of interest rates and otherwise $d > 360$.

Similarly, Asian options depend on the average value of a security over a number of time points and some options involve multiple correlated securities at a number of future time points. See Boyle, Broadie and Glasserman [2] and Joy, Boyle and Tan [19] for examples and further references.

1.1.3 Ergodic simulations and transients

Some simulations are conducted by following only a single sample path for a very long run. For such simulations to work, the problem must have an ergodic property such that one long run simulation converges to the ensemble average of many runs. The methods described here are not aimed at this problem per se, but can be of use in studying transient phenomena.

For example, suppose that a queue starts off empty and we want to know the average amount of time required to reach half of its capacity for the first time. It may be better here to simulate a large number of initially empty queues than to sample a single queue (and wait for the initial conditions to recur). Once again $d = \infty$. For a discussion of the initial transiet problem, see Glynn [13].

1.2 Outline of this article

1.2.1 Background

Section 2 reviews MC and QMC integration methods and introduces notation for them. That section also describes randomized quasi-Monte Carlo (RQMC) methods. These are hybrids of QMC and MC, with at least the accuracy of the former, and having the data based error estimation methods of the latter. Section 3 presents an ANOVA (analysis of variance) decomposition of square integrable functions over $[0,1)^d$ for $d < \infty$. This decomposition may be used to explain for which integration problems QMC might be expected to improve on MC. Section 4 presents Latin hypercube sampling for $d < \infty$.

1.2.2 Survey

The two leading families of QMC methods are lattice methods and (t, m, s)-nets. Section 5 describes lattice sampling and RQMC versions of it. Section 6 describes the (t, m, s)-nets and the related (t, s)-sequences and some RQMC versions of them. Section 7 surveys some previous methods for extending (R)QMC to higher dimensions, and gives examples of how to apply them in some of the motivating problems.

1.2.3 New results

Section 8 proposes a new simulation method in which different sets of input variables are each handled by (R)QMC. The method is called Latin supercube sampling (LSS) because it randomizes the run order within sets of input variables in the way that Latin hypercube sampling randomizes stratified input variables, one at a time. Section 9 considers the accuracy of LSS. Theorem 1 gives expressions for the bias and variance of LSS with QMC points. Theorem 2 gives an expression for the variance of LSS using RQMC points. This section also shows how to use replications to estimate the accuracy of LSS, and explains why LSS with RQMC is more reliable

than LSS with QMC. Section 10 considers the case where $d = \infty$. Using martingale truncations, an ANOVA decomposition is developed for square integrable functions on $[0,1)^\infty$. This decomposition is then used to study LHS and LSS for $d = \infty$. Some conclusions are given in Section 11.

2 Monte Carlo, Quasi-Monte Carlo and hybrids

All of the methods we consider in this paper estimate the integral I by

$$\hat{I} = \hat{I}_n = \frac{1}{n}\sum_{i=1}^{n} f(X_i) \tag{1}$$

for carefully chosen points $X_i = (X_i^1, \ldots, X_i^d) \in [0,1)^d$. That is, we do not consider here the effects of weighting the observations unequally. Some importance sampling and periodization techniques appear to weight the observations, but can be written as in (1) by subsuming the weight into f.

2.1 Monte Carlo

The simplest Monte Carlo method for estimating I takes n points X_i independently drawn from the uniform distribution on $[0,1)^d$. In practice this is almost always approximated by deterministic points taken from a pseudorandom number generator, but we will analyze the pseudorandom points as though they were genuinely random.

Under Monte Carlo sampling, and mild conditions on f, the estimator $\hat{I}$ is a random variable with expectation I and variance σ^2/n where $\sigma^2 = \int (f(X) - I)^2 dX$. (Here and elsewhere, when no region of integration is specified, the whole space $[0,1)^d$ is understood.) Thus the error in $\hat{I}$ is of order $n^{-1/2}$ in probability. Classical methods can achieve rates much better than $n^{-1/2}$, for small d and well behaved f, but the MC rate is remarkable in that it holds for all dimensions with only weak conditions on f.

2.2 Quasi-Monte Carlo

The accuracy of MC is adversely affected by gaps and clusters that arise by chance among the X_i. Equidistribution, or Quasi-Monte Carlo methods use deterministic lists of points $X_1, \ldots, X_n$ that are constructed to avoid, to the extent possible, gaps and clusters. For details on QMC methods see the monograph by Niederreiter [29].

To quantify the uniformity of a list of points, one uses a distance between the continuous uniform distribution on $[0,1)^d$ and the discrete uniform distribution taking X_i with probability $1/n$ for $i = 1, \ldots, n$. The most widely studied distance measure is the star discrepancy

$$D_n^* = D_n^*(X_1, \ldots, X_n) = \sup_{0 \le c_j < 1} \left| \prod_{j=1}^{d} c_j - \frac{1}{n} \sum_{i=1}^{n} \prod_{j=1}^{d} 1_{0 \le X_i^j < c_j} \right| \tag{2}$$

which compares the continuous and discrete uniform distributions on hyperrectangles with one corner at the origin. Some other discrepancy measures appear in [29] and Hickernell [16] has further generalized discrepancy.

Star discrepancy is related to integration accuracy by the Koksma-Hlawka inequality

$$|\hat{I} - I| \le D_n^*(X_1, \ldots, X_n) V_{\mathrm{HK}}(f) \tag{3}$$

where $V_{\mathrm{HK}}(f)$ is the total variation of f in the sense of Hardy and Krause. See [29] for the definition of V_{HK}, and [16] for analogues of the Koksma-Hlawka inequality with other notions of discrepancy.

It is possible to construct an infinite sequence of points $X_1, X_2, \ldots$ along which $D_n^* = O(n^{-1}(\log n)^d)$. This proves that one can achieve an asymptotic rate better than that of MC, at least for integrands with $V_{\mathrm{HK}}(f) < \infty$. The error bound $(\log n)^d/n$, increases with n until $n \ge \exp(d)$, so clearly for large d it takes impractically large samples before the asymptote is relevant. Yet empirical studies suggest that QMC is more accurate than MC on some real problems with practical sample sizes. Unfortunately, the superiority of QMC over MC appears to take greater n to set in when d is large. See Morokoff and Caflisch [26, 27], Sarkar and Prasad [41] and van Rensburg and Torrie [49] on these issues.

2.3 Randomized quasi-Monte Carlo

A serious drawback with QMC methods is that there is no practical way to estimate the size of $\hat{I} - I$ from the function evaluations $f(X_1), \ldots, f(X_n)$. Estimating V_{HK} from data appears to be extremely difficult, and in any event, the inequality (3) can be quite conservative.

By comparison, in the basic MC method above, the estimate

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (f(X_i) - \hat{I})^2$$

has expected value σ^2 and the central limit theorem allows one to construct approximate confidence intervals for I. For more complicated MC methods, one can replicate the whole procedure independently a number of times.

Hybrid methods have been developed to combine the best of MC and QMC. The history of this idea is outlined in [34]. The general approach is to introduce some randomness into a QMC procedure, while retaining the equidistribution properties of the QMC method. If the resulting estimate $\hat{I}$ is unbiased, then by using several independently randomized estimates one can use standard statistical methods to combine the estimates and estimate the variance of the result.

3 Functional ANOVAs

This section describes some ways to decompose the integrand f into a sum of simpler functions. We begin with notation. The set $\mathcal{A} = \{1, 2, \ldots, d\}$ denotes the coordinate axes of $[0,1)^d$. We assume here that $d < \infty$, treating the infinite dimensional case in Section 10. The letter u denotes a subset of $\mathcal{A}$, $|u|$ is the cardinality of u and $-u$ is the complement of u with respect to $\mathcal{A}$. We use these subsets as superscripts: $[0,1)^u$ denotes the space of values for those components X^j of X, with $j \in u$. Similarly X^u denotes the coordinate projection of a point X onto $[0,1)^u$ and, in integrals, $dX^u = \prod_{j\in u} dX^j$. The case $u = \emptyset$ can require special attention, either by specifying a convention, or by restricting some operations to $|u| > 0$.

Under the mild conditions that $\int f(X)^2 dX < \infty$ and f is measurable, we can write $f(X)$ as a sum of 2^d orthogonal functions, one for each subset of the input axes

$$f(X) = \sum_{u\subseteq\{1,2,\ldots,d\}} f_u(X^u). \tag{4}$$

Here f_u only depends on the components of X in X^u, though for convenience we may write it as $f_u(X)$, or as $f_u(X^v)$ for $v \supseteq u$, functions that are constant over values of X^j with $j \in v - u$.

The ANOVA decomposition is orthogonal in that $\int f_u(X) f_v(X) dX = 0$ whenever $u \neq v$. The functions f_u are defined recursively by

$$f_u(X^u) = \int_{Z:Z^u=X^u} \Big(f(Z) - \sum_{v\subset u} f_v(X^u) \Big) dZ^{-u} \tag{5}$$

where the sum is over strict subsets $v \neq u$. If $j \in u$, then $\int_0^1 f_u(X) dX^j = 0$ at any value of $X^{u-\{j\}}$.

Examples make equation (5) clearer. First of all $f_\emptyset = \int f(Z)dZ = I$. Then $f_{\{j\}}(X^j) = \int_{Z:Z^j=X^j}(f(Z)-I)\prod_{k\neq j}dZ^k$ and so on. The function $f_{\{j\}}(X^j)$, called the "main effect of X^j" can be thought of as the average effect of the j'th variable on the response function $f(X)$. The function $f_{\{j,k\}}(X^{\{j,k\}})$ describes the joint effect, or interaction, of variables X^j and X^k on the response, beyond their individual contributions $f_{\{j\}}$ and $f_{\{k\}}$. More generally f_u is the $|u|$-factor interaction among variables X^u.

The term ANOVA is an acronym for "analysis of variance". The decomposition is completely analogous to the one used in experimental statistics, for functions over Cartesian products of finite sets. The variance being analyzed is $\sigma^2 = \int(f(X)-I)^2dX$ and one easily finds that

$$\sigma^2 = \sum_u \sigma_u^2 = \sum_{|u|>0}\sigma_u^2$$

where σ_u^2 is the variance of f_u. That is $\sigma_u^2 = \int f_u(X^u)^2dX^u$ if $u\neq\emptyset$, and $\sigma_\emptyset^2 = 0$. This ANOVA decomposition for functions is due to Hoeffding. Takemura [46] gives a history. Notable contributions were made by Efron and Stein [8], Stein [45] and Wahba [51]. The notation used here is based on Owen [32]. Hickernell [16] has proposed a family of generalizations.

The best (in mean square) constant approximation to f is $I = f_\emptyset$. Stein [45] notes that the best additive approximation to f is

$$f_{\mathrm{add}}(X) = I + \sum_{j=1}^d f_{\{j\}}(X^j). \tag{6}$$

Furthermore the best approximation to f using sums of functions depending on m or fewer components is $\sum_{|u|\leq m} f_u(X^u)$, and the best approximation to f using variables contained in the set u is $\sum_{v\subseteq u} f_v$. We note for later use that

$$\int_{[0,1)^v} f(X)dX^v = \sum_{u\subseteq -v} f_u(X). \tag{7}$$

For the integration rule given by (1) we have $\hat{I} = \sum_u \hat{I}_u$ where

$$\hat{I}_u = \frac{1}{n}\sum_{i=1}^n f_u(X_i).$$

Easily $\hat{I}_\emptyset = I$ and so $\hat{I} - I = \sum_{|u|>0}\hat{I}_u$.

3.1 Effective dimension

Caflisch, Morokoff and Owen [4] define two notions of the effective dimension of an integrand, using the ANOVA decomposition.

Definition 1 *The effective dimension of f, in the superposition sense, is the smallest integer d_S such that $\sum_{|u|\le d_S}\sigma_u^2 \ge 0.99\sigma^2$.*

A method with good uniformity in every d_S dimensional projection of $X_1,\dots,X_n$ can be expected to work well with functions of effective dimension d_S (or smaller), in the superposition sense.

Definition 2 *The effective dimension of f, in the truncation sense, is the smallest integer d_T such that $\sum_{u\subseteq\{1,2,\dots,d_T\}}\sigma_u^2 \ge 0.99\sigma^2$.*

A method with good uniformity in the first d_T input variables of $X_1,\dots,X_n$ can be expected to work well with functions of effective dimension d_T (or smaller), in the truncation sense.

To illustrate the difference, a linear function has $d_S = 1$ but can have $d_T = d$. Clearly the cutoff threshold 0.99 is arbitrary and could be replaced by another.

4 Latin hypercube sampling

For one dimensional integration, the midpoint rule uses integration points given by

$$A_i = \frac{i-0.5}{n}, \quad i = 1,\dots,n. \tag{8}$$

Here and below we follow the convention that A_i denote the points of an integration rule used in constructing another integration rule. The points X_i denote the integration rule that is used to construct $\hat{I}$.

If we view the midpoint rule as a one dimensional quasi-Monte Carlo sampling scheme, the following stratified sampling method

$$A_i = \frac{i-V_i}{n}, \quad i = 1,\dots,n \tag{9}$$

where the V_i are independent $U(0,1]$ random variables, can be thought of as an RQMC method. In (9) each interval $[(k-1)/n, k/n)$ for $k = 1,\dots,n$ has exactly one uniformly randomly located point in it.

A variant of Latin hypercube sampling can be constructed from d midpoint rules, by randomizing their run order

$$X_i^j = A_{\pi_j(i)} = \frac{\pi_j(i) - 0.5}{n}. \tag{10}$$

The π_j are independent uniform random permutations of the integers $1, \ldots, n$. Each of the $n!$ possible permutations has an equal chance of being used.

The original definition of Latin hypercube sampling of McKay, Conover and Beckman [24], was based on randomizing the run order of d independently stratified samples A^j

$$X_i^j = A^j_{\pi_j(i)} = \frac{\pi_j(i) - V^j_{\pi_j(i)}}{n} = \frac{\pi_j(i) - U_i^j}{n}. \tag{11}$$

Here $U_1^j, \ldots, U_n^j$ are $V_1^j, \ldots, V_n^j$ in random order, and so the nd random variables U_i^j are also independent with the $U(0,1]$ distribution. The centered version in equation (10) was originally due to Patterson [39] in the setting of agricultural experiments, while the version in equation (11) was motivated by computer experiments.

In Latin hypercube sampling each coordinate projection of the X_i is nearly uniform. It should therefore be no surprise that when f is nearly a sum of univariate functions of the coordinates of X, that Latin hypercube sampling does well.

Under Latin hypercube sampling (11), $\hat{I}$ is a random variable with $E(\hat{I}) = I$ and

$$\begin{aligned} V_{\mathrm{LHS}}(\hat{I}) &= \frac{1}{n}\int \left(f(X) - f_{\mathrm{add}}(X)\right)^2 dX + o\left(\frac{1}{n}\right) \\ &= \frac{1}{n}\left(\sigma^2 - \sum_{j=1}^{d} \sigma^2_{\{j\}}\right) + o\left(\frac{1}{n}\right) \end{aligned} \tag{12}$$

where f_{add} is the additive function closest to f in mean square. See Stein [45] for details.

If the integrand is not nearly a sum of univariate functions, then Latin hypercube sampling cannot be expected to improve upon Monte Carlo. But it never does much worse than Monte Carlo either. For any f with $\int f^2 dX < \infty$, we have

$$V_{\mathrm{LHS}}(\hat{I}) \le \frac{n}{n-1} V_{\mathrm{MC}}(\hat{I}). \tag{13}$$

Latin hypercube sampling with $n > 1$ observations is never worse than

Monte Carlo sampling with $n-1$ observations. See Owen [35] for details.

5 Lattice Methods and Randomizations

The definitive reference on lattice methods is the book by Sloan and Joe [42]. The simplest form of lattice method is the method of good lattice points, or the number theoretic method. Another good reference for these points is Hua and Wang [17] and Fang and Wang [9] discuss applications in statistics.

For the method of good lattice points, in dimension s,

$$X_i^j = \left\{ \frac{i\, g_j}{n} \right\}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, s \tag{14}$$

where $\{z\} = z - \lfloor z \rfloor$ is the fractional part of z (or "z modulo 1") and g_j are integers. The points X_i belong to a lattice. A judicious choice for n and the g_j can produce a sequence of points with good equidistribution properties in the unit cube, hence the term "good" in "good lattice points".

Good lattice points and their generalizations are extremely well suited to integrating periodic functions with rapidly decaying Fourier coefficients. Combined with techniques for replacing non-periodic integrands by periodic ones having the same integral they are one of the leading families of Quasi-Monte Carlo integration methods.

5.1 Cranley-Patterson Rotations

Cranley and Patterson [5] randomize the good lattice points as follows. Let A_i^j be the good lattice points for $1 \le i \le n$ and $1 \le j \le s$. Let U^j be independent $U[0,1)$ random variables. Then let

$$X_i^j = \{A_i^j + U^j\}. \tag{15}$$

If one maps the unit interval $[0,1)$ onto the unit circle by identifying the endpoints, the result of the Cranley-Patterson randomization is to rotate the n points $A_1^j, \ldots, A_n^j$ to the right by a distance of U^j, with wrap-around. Taking r replicates of these n points and letting $r \to \infty$ produces errors at the Monte Carlo error rate $O((nr)^{-1/2})$ but perhaps with a much more favorable constant than has simple Monte Carlo. Joe [18] uses the same randomization on more general lattice rules.

5.2 Orthogonal rotations

The Cranley-Patterson rotations are natural for periodic functions on the unit cube. However, sometimes points A_i^j are transformed from independent uniform random variables to independent $N(0,1)$ random variables $Z_i^j = \Phi^{-1}(A_i^j)$, where Φ is the normal distribution function, prior to use. Applying Cranley-Patterson rotations to the A_i does not lead to a natural randomization of the Z_i. One might instead produce a random orthogonal matrix Q (see Devroye [7]) and apply it to the rows of A to produce row vectors $Z_i = A_i Q$. This technique gets expensive if s is large because generating Q takes $O(s^3)$ work.

6 Nets and Randomizations

We begin with more notation. We describe nets as sequences of points in the cube $[0,1)^s$, where in most applications $s = d$. When $s > d$ one simply selects d of the dimensions to use. When $s < d$ one can apply the compromise methods of Sections 7 and 8.

The integer $b \geq 2$ is used throughout as a base for representing points in $[0,1)$. Thus $X_i^j = \sum_{k=1}^{\infty} x_{ijk} b^{-k}$ where x_{ijk} are integers with $0 \leq x_{ijk} < b$.

6.1 (t,m,s)-nets, (t,s)-sequences and (λ,t,m,s)-nets.

Here we describe equidistribution methods known as (t,m,s)-nets and (t,s)-sequences. These have been developed by Sobol', Faure, and Niederreiter and a comprehensive discussion of them appears in the monograph by Niederreiter [29] on which this subsection is based. Some graphical displays of these input points appear in [23]. We also make two minor variations on Niederreiter's terminology in Definitions 3 and 4.

An elementary interval of $[0,1)^s$ in base b is defined as a set of the form

$$B = \prod_{j=1}^{s} \Big[\frac{t_j}{b^{k_j}}, \frac{t_j+1}{b^{k_j}}\Big) \tag{16}$$

for nonnegative integers k_j and $t_j < b^{k_j}$. We use Tezuka's [47] term "b-ary box" for "elementary interval in base b".

The b-ary box B in (16) is a hyperrectangle of volume b^{-m} where $m = \sum_{j=1}^{s} k_j$. In addition to the volume of a b-ary box, the "effective dimension" of the b-ary box is a useful quantity.

Definition 3 *The effective dimension of the b-ary box B defined in (16) is*

$$\delta = \delta(B) = \sum_{j=1}^{s} 1_{k_j > 0}.$$

For a point X to belong to B requires nontrivial constraints on exactly $\delta(B)$ of the coordinates X^j. The effective dimension of the b-ary box is the same as the effective dimension of the function equal to one in that b-ary box and zero outside of it (in the superposition sense, possibly requiring a threshold higher than 0.99). The unit cube $[0,1)^s$ is the unique b-ary box of effective dimension 0. No b-ary box of effective dimension δ can have volume larger than $b^{-\delta}$.

Ideally each b-ary box of volume V should have nV points of the integration rule. Let $t \geq 0$ and $m \geq 0$ be integers. A finite sequence $X_1, \ldots, X_n \in [0,1)^s$ with $n = b^m$ is a (t, m, s)-net in base b if every b-ary box of volume b^{t-m} contains exactly b^t points of the sequence. The net property starts to become relevant at $n = b^{t+1}$ where it constrains the equidistribution over some b-ary boxes of effective dimension 1. It takes at least $n = b^{t+s}$ points before the net property applies to any b-ary boxes of effective dimension s.

The infinite sequence $X_1, X_2, \ldots \in [0,1)^s$ is a (t, s)-sequence in base b if for all $m \geq 0$ and all $k \geq 0$ the finite sequence $X_{kb^m+1}, \ldots, X_{(k+1)b^m}$ is a (t, m, s)-net in base b. Niederreiter [29] discusses existence and construction of (t, m, s)-nets and (t, s)-sequences.

Smaller values of t imply better equidistribution properties for both (t, m, s)-nets and (t, s)-sequences. An advantage of using nets taken as the first $n = b^m$ points of a (t, s)-sequences is that one can later increase n through a sequence of values $n = \lambda b^m$, $1 \leq \lambda < b$, and find that all of the points used in $\hat{I}_{\lambda b^m}$ are also used in $\hat{I}_{(\lambda+1)b^m}$. As n increases through this sequence of values, any b-ary box of volume V eventually contains nV of the points, and once such a b-ary box is balanced this way, it remains balanced as n increases.

The initial λb^m points of a (t, s)-sequence are well equidistributed but are not a (t, m, s)-net, unless λ is a power of b. Owen [35] introduces the following definition to describe such point sets.

Definition 4 *Let m, t, λ be integers with $m \geq 0$, $0 \leq t \leq m$, and $1 \leq \lambda < b$. A sequence of λb^m points in $[0,1)^s$ is called a (λ, t, m, s)-net in base b if every b-ary box of volume b^{t-m} contains λb^t points of the sequence and no b-ary box of volume b^{t-m-1} contains more than b^t points of the sequence.*

Numerical integration by averaging over the points of a (t, m, s)-net has an error that is $O(n^{-1}(\log n)^{s-1})$, for functions of bounded variation in the sense of Hardy and Krause. See Niederreiter [29] for this result and some related ones. The error attained along a fixed (t, s)-sequence is $O(n^{-1}(\log n)^s)$.

6.2 Base b scrambling of the unit cube

The upper bound from the Koksma-Hlawka inequality is very hard to estimate, and is usually quite conservative, since it applies to the worst possible integrand f for the given points $X_1, \ldots, X_n$. For this reason Owen [34] suggested randomizing the points of a (t, m, s)-net or (t, s)-sequence. By independently repeating the randomization and noting how the resulting answers differ, it is possible to judge statistically the accuracy of an answer.

For this to work it is necessary that the randomization preserve the (t, m, s)-net or (t, s)-sequence structure. Below is a geometric description of the scrambling method. For an algebraic description see [34, 35]. For a discussion of computational issues see [34].

Begin by partitioning the unit cube $[0, 1)^s$ along the X^1 axis into b parallel b-ary boxes of the form $[a/b, (a + 1)/b) \times [0, 1)^{s-1}$ for $a = 0, \ldots, b - 1$. Then randomly shuffle those b-ary boxes replacing them in one of the $b!$ possible orders, each such order having probability $1/b!$. Next take each such b-ary box in turn, partition it into b congruent b-ary boxes of volume b^{-2} along the X^1 axis, and randomly shuffle those boxes. Then repeat this process on b^2 b-ary boxes of volume b^{-3}, b^3 b-ary boxes of volume b^{-4} and so, ad infinitum. In practice this can stop when the b-ary boxes are narrow compared to machine precision. The full scrambling involves applying the above operations along the other $s - 1$ axes $X^2, \ldots, X^s$ as well. All of the many permutations used are to be statistically independent, and in practice they are generated through pseudo-random numbers.

If A or A_i is a point in $[0, 1)^s$, let X or X_i be that point's location after the scrambling. A sequence (X_i) thereby inherits certain equidistribution properties of a sequence (A_i) and the individual points in it are uniformly distributed on $[0, 1)^s$. Owen [34, 35] proves the following two propositions.

Proposition 1 *If (A_i) is a (λ, t, m, s)-net in base b then (X_i) is a (λ, t, m, s)-net in base b with probability 1.*

Proposition 2 *Let A be a point in $[0, 1)^s$ and let X be the scrambled version of A as described above. Then X has the uniform distribution on $[0, 1)^s$.*

A consequence of Proposition 2 is that $\hat{I}$ is a random variable with expectation I. This holds even if the underlying A_i are not points of a net. Because of Proposition 1, any theorems describing the accuracy of $\hat{I}$ based on a (λ, t, m, s)-net with points A_i also holds for the scrambled points X_i.

The randomization in scrambled nets can lead to improved accuracy due to some error cancellation. This appeared to be the case in the simulated examples in [34] and in [36] it is shown that

$$V_{\mathrm{PNET}}(\hat{I}) = O(n^{-3}(\log n)^{s-1}) \tag{17}$$

under mild smoothness conditions on f. Thus the typical error is of order $n^{-3/2}(\log n)^{(s-1)/2}$, an improvement on the rate attained by unrandomized nets. Hickernell [15] shows that scrambled nets have an $\mathcal{L}^2$-star discrepancy that attains the best possible rate.

The variance in scrambled net simulation satisfies

$$V_{\mathrm{PNET}}(\hat{I}) = \sum_{|u|>0} V_{\mathrm{PNET}}(\hat{I}_u). \tag{18}$$

The results in [36] suggest that $V_{\mathrm{PNET}}(\hat{I}_u)$ may become appreciably smaller than $V_{\mathrm{MC}}(\hat{I}_u)$ at around $n = b^{t+|u|}$. Moreover $V_{\mathrm{PNET}}(\hat{I}_u) \le 2.7183 V_{\mathrm{MC}}(\hat{I}_u)$ for any scrambled net with $t = 0$ in any dimension $s \ge 1$ with any integrand f, so least favorable integrands can't make randomized nets much worse than simple Monte Carlo.

6.3 Cranley-Patterson Randomization

It is also possible to apply the Cranley-Patterson randomization of Section 5.1 to points from (t, m, s)-nets and (t, s)-sequences. See Tuffin [48] for details. This randomization does not preserve equidistribution over b-ary boxes, but it does produce unbiased estimates $\hat{I}$ whose variance can be estimated by replication.

6.4 Nets in very high dimensions

There are limitations to the applicability of nets in very high dimensions. Consider $d = 50$. For a $(0, s)$-sequence to be usable requires $s \ge 50$, which in turn requires $b \ge 50$. Niederreiter [28] gives constructions of $(0, s)$-sequences for prime powers $b \ge s$. The natural choice for $d = 50$, is the smallest one, $b = 53$. Because 53 is a prime, the resulting sequence is a Faure sequence [10]. Natural sample sizes with such a base are powers of 53. The superior

asymptotic rate of convergence of these may be expected to set in after $n = 53^{50} \doteq 1.64 \times 10^{86}$. Unless $n \geq 53^2 = 2809$ the net property does not gaurantee balance for any b-ary box of effective dimension larger than one.

Another approach is to use a (t, s)-sequence with smaller b, and in the tradeoff take a larger t. The most widely studied version of this strategy takes $b = 2$, the smallest possible base. For $b = 2$ and $s = 50$, the presently best possible value of t is 77, by a construction of Niederreiter and Xing [30]. The superior asymptotic rate of convergence with this sequence may be expected to set in after $n = b^{t+s} = 2^{127} \doteq 1.70 \times 10^{38}$. This is the smallest value of n for which we can partition the unit cube into b-ary boxes of effective dimension 50 and be sure that each such box has at least one point in it.

Widely used software for generating nets has a limit on the dimension. For instance, the code of Bratley, Fox and Niederreiter [3] has by default a limit of $d = 12$.

The asymptotic advantage of nets, scrambled or otherwise, in dimension 50 appears to take an impractically large number of observations to set in, and the matter is worse for $d = 1000$ or $d = \infty$. Yet as described in Section 11 and seen in [4, 37, 38] it is possible for high dimensional functions to be integrated accurately. This may be because the functions involved have a low effective dimension.

7 Padding Techniques

The full input sequence for the simulation is a n by d matrix X_i^j of numbers between 0 and 1. An (R)QMC method may be used for s of the columns of this matrix. Some simple techniques below may be used to "pad out" the matrix.

7.1 Padding by Monte Carlo

A natural solution is to pad out the input matrix with independent $U[0, 1)$ random variables. Suppose for example, that A_i are the points of an (R)QMC method in s dimensions. Then one might take $X_i^j = A_i^j$ for $1 \leq j \leq s$ and $X_i^j = U_i^j$ for $j > s$ where U_i^j are independent $U[0, 1)$ random variables, independent of the randomization, if any, used in constructing A. This approach is taken by Spanier [44] and Okten [31] studies the resulting discrepancy.

The success of this method will depend in part on making a good choice of which s variables to use (R)QMC on. One might try to identify the most

important subset u of s variables, perhaps by their ANOVA contribution $\sum_{v\subseteq u}\sigma_v^2$. These are not necessarily the first s variables in any natural order, and it may take subject matter knowledge, guesswork and experimentation to make a good choice. A more realistic goal may be to pick those variables with a large value of $\sum_{|v|\le d^\star, v\subseteq u}\sigma_v^2$, where $d^\star$ is the largest dimensionality in which good equidistribution may be expected of the projected (R)QMC points.

7.2 Padding by Latin hypercube sampling

It is possible to pad out (R)QMC in a way that is better than simple Monte Carlo. For any choice n, one can pad with a Latin hypercube sample of n rows and $d-s$ columns. This is even true if $d > n$. In simulations with large enough d (d can be infinite) it may take a "sampling without replacement" trick to generate only those rows of the column X^j that the simulation really uses. Owen [33] gives an example in which a randomized orthogonal array sample is padded out with a Latin hypercube sample.

Padding with Latin hypercube samples instead of independent samples improves the accuracy of integration for the main effects in the $d-s$ variables not sampled by a net. In principle, this may make one want to change which subset u of s variables to apply (R)QMC to. Consider a variable that has an enormous but purely additive effect on f. That variable is handled poorly by Monte Carlo but Latin hypercube sampling handles it as well as most (R)QMC methods do. If one is padding by MC then this variable should be among the (R)QMC variables, but if one is padding by LHS then it should not be among the (R)QMC variables. A natural choice would be the subset u for which $\sum_{1<|v|,v\subseteq u}\sigma_v^2$ or perhaps $\sum_{1<|v|\le d^\star, v\subseteq u}\sigma_v^2$ is maximized.

A disadvantage of Latin hypercube padding is that one cannot add more simulation points to a Latin hypercube sample of size n and still have a Latin hypercube sample. One remedy, used by Caflisch, Morokoff and Owen [4] is to use r independent Latin hypercube samples of n' rows and $d-s$ columns where $n = rn'$. Using blocks of Latin hypercube samples allows one to select sample sizes that are multiples of the block size n'. It makes sense to choose a block size which is a divisor of the natural sample sizes for the (R)QMC method being used. For nets, a power of b seems reasonable.

7.3 Padding with (R)QMC

For simplicity suppose that $d = ks$. It would clearly not work to repeat an s dimensional (R)QMC method k times taking $X_i = (A_i, A_i, \ldots, A_i)$, where

A_i are the s dimensional points of an (R)QMC method. For example, the points X_i^j and X_i^k where $j = k \bmod s$ would then lie on a diagonal in the $X^{\{j,k\}}$ plane.

Similarly, using k independent randomizations of a single underlying QMC point set cannot be expected to work well in practice either. For randomized nets, suppose X^j and X^k are both randomizations of A^1 taken from a (t, m, s)-net in base b with $m > t$. There are n/b points with $0 \le A_i^1 < 1/b$. If $0 \le A_i^1 < 1/b$, then $\tau_j/b \le X_i^j < (\tau_j + 1)/b$ *and* $\tau_k/b \le X_i^k < (\tau_k + 1)/b$ for some randomly chosen $0 \le \tau_j, \tau_k < b$. This implies that a box $[\tau_j/b, (\tau_j + 1)/b) \times [\tau_k/b, (\tau_k + 1)/b)$ of volume b^{-2} has a fraction b^{-1} of the points in the $X^{\{j,k\}}$ plane. The "diagonal" referred to in the previous paragraph is rearranged, but not enough to produce a uniform distribution in the $X^{\{j,k\}}$ plane. Similarly, the randomization of Cranley-Patterson (Section 5.1) produces points $(\{A_i^1 + U^j\}, \{A_i^1 + U^k\})$ so that the diagonal line is shifted to a new random line with "wraparound".

Another tempting trick that fails is to apply various nets in relatively prime bases. One might consider using Faure sequences, $(0, p)$-sequences in base p for primes $p = 2, 3, 5, \ldots$ until the sum of primes used is greater than or equal to d. This does with Faure sequences what the Halton sequence does with generalized van Der Corput sequences. One naively expects, for example, that given an input X^j generated in base 2 and an input X^k generated in base 3 that all boxes of the form $[\tau_j/2, (\tau_j + 1)/2) \times [\tau_k/3, (\tau_k + 1)/3)$ for integers $\tau_j \in \{0, 1\}$, $\tau_k \in \{0, 1, 2\}$ would each have one sixth of the points $X_i^{\{j,k\}}$, when n is a multiple of 6. They don't. Faure [11] (Section 6) suggests something similar but reports in a personal communication that he thinks it's not a good idea. As explained by Niederreiter (personal communication), there is no reason to expect this method to work, because unlike the Halton setting, there is no analogue of the Chinese remainder theorem to apply here.

7.4 Reducing Effective Dimension

When one plans to use (R)QMC points on some inputs and padding for the others, one may be able to do more than simply choose the important variables for QMC. In some cases one can rewrite the integrand in a way that puts a greater amount of the variation into a small number of inputs. For example, when the inputs are used to construct Brownian paths a Brownian bridge encoding as in [4] makes the first few input variables more important than the regular encoding.

Acworth, Broadie and Glasserman [1] use an encoding based on princi-

pal components for Brownian motion. The first 5 principal components of Brownian motion explain about 96% of the variation in the Brownian path. Thus it may pay to use an (R)QMC method on these variables but not on the others.

Fox [12] describes some additional settings where hybrids may be applied. For discrete event simulation, with a Poisson arrival process, Fox suggests first drawing the number of arrivals, then the median arrival time, then a sequence of intermediate times. Fox recommends using QMC on the earlier variables and MC for the remainder. Another possibility he considers is recursively splitting the time interval and using binomial sampling to determine how many observations to put in each subinterval.

Yet another recommendation is for simulations requiring finite independent samples $Z_1, \dots, Z_d$ from some distribution. There Fox suggests using QMC to generate some or all of the order statistics and then MC to generate the remaining order statistics, if any, as well as the random allocation of order statistics to sample values.

8 Latin Supercube Sampling

8.1 Introduction

In Latin supercube sampling (LSS) one takes a list of QMC point sets or RQMC point sets and randomizes their run order the same way that Latin hypercube sampling randomizes the run order of the midpoint and stratified techniques.

Let $\mathcal{A} = \cup_{r=1}^{k} \mathcal{A}_r$ with $\mathcal{A}_r \cap \mathcal{A}_{r'} = \emptyset$ for $r \neq r'$ be a partition of the simulation variables into k nonempty subsets. Letting $s_r = |\mathcal{A}_r|$ we have $s_r \geq 1$ and $\sum_{r=1}^{k} s_r = d$. Let $\mathcal{X}_i^r \in [0,1)^{\mathcal{A}_r}$ for $i = 1, \dots, n$ and $r = 1, \dots, k$. In practice these $\mathcal{X}_i^r$ will ordinarily be points of an s_r dimensional (R)QMC method. For $r = 1, \dots, k$ let $\pi_r(i)$ be independent uniform random permutations of the integers $1, \dots, n$. Then a Latin supercube sample (LSS) is formed by taking

$$X_i = (\mathcal{X}_{\pi_1(i)}^1, \mathcal{X}_{\pi_2(i)}^2, \cdots, \mathcal{X}_{\pi_k(i)}^k). \tag{19}$$

Equation (19) assumes that the variables in $\mathcal{A}_r$ come before those in $\mathcal{A}_{r'}$ whenever $r < r'$. This is not at all necessary. The variables can be interleaved in any order, taking $X_i^{\mathcal{A}_r} = \mathcal{X}_{\pi_r(i)}^r$. The ordering assumption is made solely to simplify notation.

In words, the first s_1 columns in the LSS are obtained by randomly

permuting the run order of the QMC points $\mathcal{X}_1^1, \ldots, \mathcal{X}_n^1$, the next s_2 columns come from an independent permutation of the run order of $\mathcal{X}_i^2$ and so on.

The best results may be expected if one can arrange that variables that interact most strongly are grouped into a subset $\mathcal{A}_r$. In the extreme case, suppose that $\sigma_u^2 = 0$ unless $u \subseteq A_r$ for some r. For simplicity suppose also that each $s_r = s$. In this case the error in LSS is the sum of k (R)QMC errors, one from each subset of axes. One then has an integration rule for a d dimensional function, that converges at a rate usually seen in s dimensional problems. More realistically, one might be able to arrange for *most* of the variance in the function f to take place within subsets $\mathcal{A}_r$ of axes. Then the residual

$$f(X) - I - \sum_{r=1}^{k} \sum_{\substack{0<|u| \\ u \subseteq \mathcal{A}_r}} f_u(X)$$

is small, and Section 9 shows that this residual contributes to the integration error at essentially the Monte Carlo rate.

How could LSS randomization of QMC work when padding with multiple RQMC randomizations, as in Section 7.3 does not work? Consider again input variables X^j and X^k which are both randomizations of the same QMC column A^1. Suppose that $j \in \mathcal{A}_1$ and $k \in \mathcal{A}_2$. With RQMC randomization alone, X_i^j and X_i^k are both images of the same point A_i^1. By randomizing run order, LSS breaks the link between $X_i^j = A^1_{\pi_1(i)}$ and $X_i^k = A^1_{\pi_2(i)}$, because $\pi_1(i)$ is independent of $\pi_2(i)$.

8.2 Partitions

The best way to partition axes is problem specific and may take skill to guess. In addition to the variation considerations above, software engineering complications arise. It may be much more convenient to group together variables used within a software module than to group together variables that span several modules.

For example, in particle transport simulations, it may be best for the variables that determine the r'th collision to be drawn from the same randomization of the (R)QMC points. Or, it might be better to use one (R)QMC sequence for the change in x component direction for s consecutive collisions, another for the change in y component direction and so forth. Which is better depends on which input variables "interact most" and this can be expected to depend on what response is being measured.

In a financial simulation with k Brownian paths, it may make sense to select 5 principal components of each path, apply an (R)QMC method

to each of them with LSS and then pad out the other variables by LHS. Alternatively, it may be better to group the k first principal components together then the k second components and so on.

In a discrete event simulation each source of arrivals could receive its own set of (R)QMC variables. Similarly the variables describing service times could be bundled within one or more (R)QMC sets.

9 LSS Accuracy

9.1 Finite k, QMC

We begin with Latin supercube sampling of a finite number of finite sets of variables. Let $\mathcal{X}^r \in [0,1)^{s_r}$, where $1 \le s_r < \infty$ for $r = 1, \ldots, k < \infty$. Suppose that the points $\mathcal{X}_i^r$ for $i = 1, \ldots, n$ are a nonrandom s_r dimensional integration scheme and that $d = \sum_{r=1}^k s_r$. The motivating case is where one has an s_r dimensional QMC method for each set of points $\mathcal{X}_i^r$.

For independent random permutations π_r of 1 through n, let

$$X_i = (\mathcal{X}^1_{\pi_1(i)}, \mathcal{X}^2_{\pi_2(i)}, \cdots, \mathcal{X}^k_{\pi_k(i)})$$

define an integration point in $[0,1)^s$. We are interested in the error $\hat{I} - I$ where $\hat{I} = n^{-1}\sum_{i=1}^n f(X_i)$ as always.

The integration error may be decomposed as

$$\hat{I} - I = (\hat{I} - I_G) + (I_G - I) \tag{20}$$

where

$$I_G = \frac{1}{n^k}\sum_{i_1=1}^n \sum_{i_2=1}^n \cdots \sum_{i_k=1}^n f(\mathcal{X}^1_{i_1}, \mathcal{X}^2_{i_2}, \cdots, \mathcal{X}^k_{i_k}) \tag{21}$$

is the average of f over a perhaps very large grid containing all n^k integration points that could possibly be sampled by LSS. Define

$$\sigma_G^2 = \frac{1}{n^k}\sum_{i_1=1}^n \sum_{i_2=1}^n \cdots \sum_{i_k=1}^n \Big(f(\mathcal{X}^1_{i_1}, \mathcal{X}^2_{i_2}, \cdots, \mathcal{X}^k_{i_k}) - I_G\Big)^2 \tag{22}$$

and

$$\sigma^2_{G(r)} = \frac{1}{n}\sum_{i_r=1}^n \left(\frac{1}{n^{k-1}}\sum_{i_1=1}^n \cdots \sum_{i_{r-1}=1}^n \sum_{i_{r+1}=1}^n \cdots \sum_{i_k=1}^n f(\mathcal{X}^1_{i_1}, \cdots, \mathcal{X}^k_{i_k}) - I_G\right)^2.$$

Lemma 1 *Under LSS with deterministic QMC point sets,*

$$E(\hat{I} - \hat{I}_G) = 0, \tag{23}$$

and for $n > 1$

$$V(\hat{I} - \hat{I}_G) \leq \frac{1}{n-1}\Big(\sigma_G^2 - \sum_{r=1}^{k} \sigma_{G(r)}^2\Big) \tag{24}$$

$$V(\hat{I} - \hat{I}_G) \geq \frac{1}{n}\left(1 - \frac{1}{(n-1)^2}\right)\Big(\sigma_G^2 - \sum_{r=1}^{k} \sigma_{G(r)}^2\Big). \tag{25}$$

Proof: Consider the function h defined on $[0,1)^k$ through

$$h(z_1, \ldots, z_k) = f(\mathcal{X}^1_{1+\lfloor z_1 n \rfloor}, \ldots, \mathcal{X}^k_{1+\lfloor z_k n \rfloor}),$$

where as usual $\lfloor x \rfloor$ denotes the smallest integer less than or equal to x. Now for any permutations $\pi_1, \ldots, \pi_k$ of $1, \ldots, n$ the Latin hypercube sample with $Z_i^j = (\pi_j(i) - U_i^j)/n$, the centered version of Latin hypercube sampling with $Z_i^j = (\pi_j(i) - 1/2)/n$, and the Latin supercube sample with $X_i = (\mathcal{X}^1_{\pi_1(i)}, \cdots, \mathcal{X}^k_{\pi_k(i)})$ all yield the same function values $Y_i = f(X_i) = h(Z_i)$.

That randomly centered Latin hypercube sampling of h is unbiased establishes (23). The other two results follow from equation (3.2) of [33] which treats the variance of centered Latin hypercube samples, as a special case of randomized orthogonal arrays. The quantity σ_G^2 is the variance of h and $\sigma_{G(r)}^2$ is the discrete version $\sigma_{\{r\}}^2$ for h. ∎

The performance of LSS sampling depends on the quantities I_G, σ_G^2 and $\sigma_{G(r)}^2$, which involve approximate integrals of f and related quantities, based on the QMC point sets $\mathcal{X}_i^r$. Let

$$\mathcal{E}_n^r = \mathcal{E}_n^r(f) = \sup_{X \in X^{-\mathcal{A}^r}} \left| \int_{Z : Z^{-\mathcal{A}} = X} f(Z) d\mathcal{Z}^r - \frac{1}{n}\sum_{i=1}^{n} f(\mathcal{X}^1, \ldots, \mathcal{X}_i^r, \ldots, \mathcal{X}^k) \right|$$

denote the worst error obtained by the integration rule $\mathcal{X}_1^r, \ldots, \mathcal{X}_n^r$ in integrating f over $\mathcal{X}^r$ with the other components of X held fixed. If the QMC point sets are really superior to MC, then we should have $\mathcal{E}_n^r = o(n^{-1/2})$.

It is well known that the integration error in a product of one dimensional rules is essentially the sum of the one dimensional integration rule errors. See Davis and Rabinowitz [6], Chapter 5.6, who cite Haber [14]. Here we present a version for $\hat{I}_G$, a product of multidimensional rules.

Proposition 3 $|\hat{I}_G - I| \leq \sum_{r=1}^{k} \mathcal{E}_n^r$.

Proof: By definition of $\mathcal{E}_n^k$,

$$\begin{aligned}\hat{I}_G &\leq \left(\frac{1}{n^{k-1}}\sum_{i_1=1}^{n}\cdots\sum_{i_{k-1}=1}^{n}\int_{[0,1)^{\mathcal{A}_r}} f(\mathcal{X}_{i_1}^1,\cdots,\mathcal{X}_{i_{k-1}}^{k-1},\mathcal{X}^k)d\mathcal{X}^k\right)+\mathcal{E}_n^k \\ &= \int_{[0,1)^{\mathcal{A}_r}}\left(\frac{1}{n^{k-1}}\sum_{i_1=1}^{n}\cdots\sum_{i_{k-1}=1}^{n} f(\mathcal{X}_{i_1}^1,\cdots,\mathcal{X}_{i_{k-1}}^{k-1},\mathcal{X}^k)d\mathcal{X}^k\right)+\mathcal{E}_n^k.\end{aligned}$$

Making $k-1$ more comparisons between integrals with respect to $\mathcal{X}^r$ and integration rules $\mathcal{X}_i^r$ gives $I_G \leq I + \sum_{r=1}^{k}\mathcal{E}_n^r$. A similar argument gives $I_G \geq I - \sum_{r=1}^{k}\mathcal{E}_n^r$. ■

Theorem 1 *In Latin supercube sampling, if* $\mathcal{E}_n^k(f) = o(n^{-1/2})$, *then*

$$E_{\mathrm{LSS}}(\hat{I}) = I + o(n^{-1/2}) \tag{26}$$

If also $\mathcal{E}_n^r(f_u f_v) = o(n^{-1/2})$ *for all* $u, v \subseteq \mathcal{A}$, *then*

$$V_{\mathrm{LSS}}(\hat{I}) = \frac{1}{n}\left(\sigma^2 - \sum_{r=1}^{k}\sum_{u\subseteq\mathcal{A}_r}\sigma_u^2 + o\left(n^{-1/2}\right)\right). \tag{27}$$

Proof: Equation (26) follows from equation (20), Proposition 3 and Lemma 1. If also $\mathcal{E}_n^r(f_u f_v) = o(n^{-1/2})$ for all $u, v \subseteq \mathcal{A}$, then $\mathcal{E}_n^r(f^2) = o(n^{-1/2})$ and so

$$\sigma_G^2 = \sigma^2 + o(n^{-1/2}).$$

Similarly, assuming each $\mathcal{E}_n^r(f_u f_v) = o(n^{-1/2})$, and making use of equation (7) with $v = -\mathcal{A}_r$ we find

$$\begin{aligned}\sigma_{G(r)}^2 &= \frac{1}{n}\sum_{i_r=1}^{n}\left(\int_{[0,1)^{-\mathcal{A}_r}} f(\mathcal{X}^1,\cdots,\mathcal{X}_{i_r},\cdots,\mathcal{X}^k)dX^{-\mathcal{A}_r} - I_G\right)^2 + o(n^{-1/2}) \\ &= \frac{1}{n}\sum_{i_r=1}^{n}\left(\sum_{0<|u|,u\subseteq\mathcal{A}_r} f_u(\mathcal{X}_i^r)\right)^2 + o(n^{-1/2}) \\ &= \sum_{u\subseteq\mathcal{A}_r}\sigma_u^2 + o(n^{-1/2}).\end{aligned}$$

Substituting into the variance results of Lemma 1 establishes equation (27). ■

Comparing LSS through (27) with LHS through (37) we see that both methods remove all components σ_u^2 with $|u| = 1$ from the variance of $\hat{I}$. The

latter method also removes those higher order components corresponding to $u \subseteq \mathcal{A}_r$.

As for the case of padding we can obtain a more realistic finite sample approximation

$$V_{\mathrm{LSS}}(\hat{I}) \doteq \frac{1}{n}\Bigl(\sigma^2 - \sum_{r=1}^{k} \sum_{|u|\le d^\star, u\subseteq\mathcal{A}_r} \sigma_u^2\Bigr). \tag{28}$$

Here $d^\star$ is the largest dimension in which one can expect the superiority of QMC to have set in. The results in [36] suggest that $d^\star$ could be as large as $m-t$ for a (t,m,s)-net in base b.

9.2 Finite k, RQMC

Suppose that the RQMC points $\mathcal{X}_i^r$ are used instead of QMC points. We assume that as a result of their randomization, RQMC points satisfy $\mathcal{X}_i^r \sim U[0,1)^{s_r}$. This implies that $X_i \sim U[0,1)^d$. We also assume that the randomization preserves (or enhances) the accuracy of the underlying points. This is easiest to establish when the randomization applied preserves the equidistribution properties of the QMC method used. Thus $\mathcal{E}_n^r(f) = o(n^{-1/2})$.

Theorem 2 *In Latin supercube sampling, with RQMC point sets, if each* $X_i \sim U[0,1)^d$, *then*

$$E_{\mathrm{RLSS}}(\hat{I}) = I. \tag{29}$$

If also $\mathcal{E}_n^r(f) = o(n^{-1/2})$ *and* $\mathcal{E}_n^r(f_u f_v) = o(n^{-1/2})$ *for all* $u, v \subseteq \mathcal{A}$, *then*

$$V_{\mathrm{RLSS}}(\hat{I}) = \frac{1}{n}\Bigl(\sigma^2 - \sum_{r=1}^{k} \sum_{u\subseteq\mathcal{A}_r} \sigma_u^2 + o\left(n^{-1/2}\right)\Bigr) + o(n^{-1}). \tag{30}$$

Proof: Equation (29) follows because each $X_i \sim U[0,1)^d$. There are two sources of randomization in RLSS, the RQMC randomization of the underlying QMC points, and the LSS randomizations of their run orders. Let $\mathcal{Q}$ denote all the random variables used in construction of the RQMC points $\mathcal{X}_i^r$. Given $\mathcal{Q}$, we may apply conclusions of Theorem 1,

$$\begin{aligned} V_{\mathrm{RLSS}}(\hat{I}) &= E_{\mathrm{RQMC}}\left(V_{\mathrm{LSS}}(\hat{I} \mid \mathcal{Q})\right) + V_{\mathrm{RQMC}}\left(E_{\mathrm{LSS}}(\hat{I} \mid \mathcal{Q})\right) \\ &= E_{\mathrm{RQMC}}\left(\frac{1}{n}\Bigl(\sigma^2 - \sum_{r=1}^{k} \sum_{u\subseteq\mathcal{A}_r} \sigma_u^2 + o\left(n^{-1/2}\right)\Bigr)\right) + V_{\mathrm{RQMC}}(\hat{I}_G) \end{aligned}$$

$$= \frac{1}{n}\Big(\sigma^2 - \sum_{r=1}^{k}\sum_{u\subseteq\mathcal{A}_r}\sigma_u^2 + o\left(n^{-1/2}\right)\Big) + o(n^{-1}). \quad \blacksquare$$

The asymptotic variance formula in Theorems 1 and 2 are essentially the same. In Theorem 1 there is an asymptotically negligible bias whereas in Theorem 2 there is no bias, though randomness in $\hat{I}_G$ may add an (asymptotically negligible) amount to the variance.

9.3 RQMC versus QMC in LSS

In LSS with QMC points the difference $\hat{I} - \hat{I}_G$ makes an asymptotically negligible contribution to bias, whereas with RQMC points the effect of $\hat{I} - \hat{I}_G$ is an asymptotically negligible increase in variance. This section explains why RQMC is preferred in practice.

While $\hat{I} - \hat{I}_G$ is asymptotically negligible in LSS, a given value of n may not be large enough that $|I - \hat{I}_G| \ll |\hat{I} - \hat{I}_G|$. In this case the QMC version of LSS may be misleading, in a way that the RQMC version is not.

The accuracy of LSS can be estimated using multiple independent estimates $\hat{I}_1, \ldots, \hat{I}_r$ of I. Standard statistical methods may then be used to estimate the variance in equation (27) or (30), and estimate the variance of the pooled estimate $I^\star = 1/r\sum_{h=1}^{r}\hat{I}_h$, by

$$\frac{1}{r(r-1)}\sum_{h=1}^{r}\left(\hat{I}_h - I^\star\right)^2. \tag{31}$$

In the case of QMC, the value of $\hat{I}_G - I$ is constant in each replicate and $E(I^\star - I) = E(\hat{I}_h - I) = I_G - I$. Increasing the number r of replicates will not decrease this source of error. Furthermore the variance estimate (31) will not reflect the error $\hat{I}_G - I$, and hence will be misleadingly small.

By contrast, in the case of RQMC, each independent replication of the LSS permutations can be done with independently generated QMC points. Then $\hat{I}_G - I$ has mean zero and varies independently from replicate to replicate. The variance contribution of $\hat{I}_G - I$ thus decreases with r. Furthermore the fluctuations in $\hat{I}_G - I$ are captured in (31).

In some cases we might be interested in comparing the magnitude of $\hat{I} - \hat{I}_G$ with that of $\hat{I}_G - I$. With the RQMC version of LSS both errors contribute to variance. Standard statistical experimental designs can be used to compare these variance components. One can vary the RQMC points r_1 times and for each of them vary the LSS randomization r_2 times. Then an analysis of the $r_1 r_2$ replicates would allow us to infer whether n is

large enough that $|I - \hat{I}_G| \ll |\hat{I} - \hat{I}_G|$.

10 Infinite Dimensional versions

10.1 Issues when $d = \infty$

The infinite dimensional case needs to be treated with some extra care. For example, in finite dimensional Latin hypercube sampling, one may proceed by showing that

$$V_{\mathrm{LHS}}(\hat{I}) = \frac{1}{n} V(f(X_1)) + \frac{n-1}{n} \mathrm{Cov}_{\mathrm{LHS}}(f(X_1), f(X_2)).$$

Then $\mathrm{Cov}_{\mathrm{LHS}}(f(X_1), f(X_2)) = E((f(X_1) - I) E_{\mathrm{LHS}}(f(X_2) - I | X_1))$, and given X_1 the location of X_2 is uniformly distributed over a set of volume $(1 - 1/n)^d$. But if $d = \infty$ this set has volume zero, calling into question averages over it.

An ANOVA on infinitely many dimensions also requires some care, because there are uncountably many subsets of $\{1, 2, \ldots\}$ among which to partition the variance. Also if $|u| = \infty$ then the definition of f_u in equation (5) involves an uncountable sum over all proper subsets of u. What would we make of an ANOVA effect f_u for $|u| = \infty$, such as an interaction among all components X^j for which j is prime? Fortunately, we only need to use finite subsets u, and there are only countably many of these.

10.2 Martingale truncation

In the infinite dimensional examples of Section 1.1, one expects that f should be "almost finite dimensional", in that the first s dimensions for some possibly large s should capture virtually all of the important variation in f. This expectation is borne out, whenever $\int_{[0,1)^\infty} f(X)^2 dX < \infty$.

Let $1{:}s$ denote the set $\{1, 2, \ldots, s\}$ of leading variables. There are a number of ways to approximate an infinite dimensional f by a function of $X^{1:s}$. One could simply replace each $X^{s+1}, X^{s+2}, \ldots$ by a convenient value such as 0.5 or 0.0, effectively turning f into a function of s variables. Or one can replace $f(X)$ by its minimum (or maximum) value over $X^{s+1}, X^{s+2}, \ldots$ with $X^{1:s}$ held fixed. Or one can replace $f(X)$ by its expectation over $X^{s+1}, X^{s+2}, \ldots$ with $X^{1:s}$ held fixed. This latter approximation is most convenient theoretically because it allows the use of martingale methods. See Williams [53]. We do not give a rigorous treatment of martingales here.

Now define

$$f^s(x^1,\dots,x^s) = E(f(X)|X^1 = x^1,\dots,X^s = x^s), \tag{32}$$

where expectation is taken over independent $X^j, j \geq s+1$ having a $U[0,1)$ distribution. When $1{:}s \subseteq u$ the function $f^s(X^u)$ is taken to mean $f^s(X^{1:s})$ ignoring any coordinates in $1{:}s - u$. For $s = 0$ we take $f^0(X^u) = I$ for any u. The sequence $f^s(X), s \geq 0$ is a martingale, by Levy's upward theorem (Chapter 14.2 of [53]), when the X^j are independent $U[0,1)$ random variables. The key martingale property is that $E(f^{s+1}(X)|X^{1:s}) = f^s(X)$. In this paper f^2 always means the square of f and never means f^s with $s = 2$.

Because we assume $E(f(X)^2) < \infty$ the martingale is bounded in L^2. Thus as $s \to \infty$, we have $f^s(X) \to f(X)$ in L^2 and pointwise. The L^2 convergence means that

$$\lim_{s\to\infty} E\left((f(X) - f^s(X))^2\right) = 0. \tag{33}$$

The pointwise convergence is "almost sure". That is $\lim_{s\to\infty} f^s(X) = f(X)$ holds on a subset of $[0,1)^\infty$ having probability one.

10.3 ANOVA with $d = \infty$

To study functions over infinite dimensional domains we first truncate the dimension, replacing f by some f^s, and then apply a finite dimensional ANOVA to f^s. We begin by defining ANOVA terms for f and for the truncated functions f^s. Then Proposition 4 shows that the two definitions are compatible.

The infinite dimensional ANOVA enjoys two key properties of the finite dimensional ANOVA: Proposition 5 shows that the infinite dimensional function is still a sum of its ANOVA components, and Proposition 6 shows that the variance decomposes into a sum of terms for each component.

Definition 5 *For $d = \infty$ and $|u| < \infty$ define f_u by equation (5). For $|u| = \infty$ take $f_u(X) = 0$ for all X.*

Definition 6 *For $u \subseteq 1{:}s$ let f^s_u be the ANOVA term for u obtained by replacing f by f^s in $d = s$ dimensional ANOVA definition (5). For $u \not\subseteq 1{:}s$, take $f^s_u = 0$.*

Proposition 4 *If $s < \infty$ and $u \subseteq 1:s$, then $f^s_u = f_u$.*

Proof: Consider first $u = \emptyset$. For $s \geq 1$,

$$f^s_\emptyset = \int_{[0,1)^s} f^s(X^{1:s}) dX^{1:s} = E(f(X)) = f_\emptyset$$

by the definition (32) of f^s. Now suppose that $f^s_u = f_u$ whenever $|u| < k$, for $k \geq 0$. Then from the definition (32) of f^s and by (5), we get $f^s_u = f_u$ for $u \subseteq 1{:}s$, for $|u| \leq k$. By induction on $|u|$, it now follows that $s < \infty$ and $u \subseteq 1:s$ implies $f^s_u = f_u$ ■.

In the finite dimensional case, the identity $f = \sum_u f_u$ holds everywhere simply by the definition of the highest order interaction $f_{\{1,\ldots,d\}}$. For $d = \infty$ the decomposition is more subtle. Definition 5 asserts that infinite order terms vanish. But it remains to prove that

$$f(X) = \sum_{0 \leq |u| < \infty} f_u(X). \tag{34}$$

Proposition 5 *If $E(f^2) < \infty$ then equation (34) holds almost surely and in L^2.*

Proof: In the chain

$$\begin{aligned} f(X) &= \lim_{s\to\infty} f^s(X) \\ &= \lim_{s\to\infty} \sum_{u \subseteq 1:s} f^s_u(X) \\ &= \lim_{s\to\infty} \sum_{u \subseteq 1:s} f_u(X) \\ &= \sum_{|u|<\infty} f_u(X) \end{aligned}$$

the first inequality holds almost surely and in L^2 because the martingale is bounded in L^2, the second is exact from the finite dimensional ANOVA, the third follows from Proposition 4, and the fourth follows from two different ways of listing all finite subsets of the positive integers. ■

Proposition 6 *If $E(f^2) < \infty$ then*

$$\sigma^2 = \int (f(X) - I)^2 dX = \sum_{1 \leq |u| < \infty} \int f_u(X)^2 dX. \tag{35}$$

Proof: For any $s \geq 1$ we may write

$$\begin{aligned}\sigma^2 &= \int (f(X) - f^s(X) + f^s(X) - I)^2 \, dX \\ &= \int (f(X) - f^s(X))^2 dX + \int (f^s(X) - I)^2 dX \qquad (36) \\ &\geq \sum_{\substack{|u|>0 \\ u \subseteq 1:s}} \int f_u(X)^2 dX.\end{aligned}$$

Taking the limit as $s \to \infty$, we find $\sigma^2 \geq \sum_{1 \leq |u| < \infty} \int f_u(X)^2 dX$.

For $\epsilon > 0$ choose s_0 so that $\int (f^s(X) - f(X))^2 dX < \epsilon$ whenever $s \geq s_0$. The decomposition (36) now leads to $\sum_{1 \leq |u| < \infty} \int f_u(X)^2 dX > \sigma^2 - \epsilon$. Letting $\epsilon \to 0$ establishes (35). ∎

10.4 LHS with $d = \infty$

Because each X_i is individually $U[0,1)^\infty$ it follows that $E_{\mathrm{LHS}}(\hat{I}) = E_{\mathrm{MC}}(\hat{I}) = I$. Where it makes no difference, the subscript LHS or MC may be dropped from the expectation symbol. Because they are both unbiased, LHS and MC can be compared through their variances. We show below that the variance under LHS is a sum of contributions from each ANOVA term.

The main point of this subsection is to extend two results from finite dimensional LHS to the infinite dimensional case: Lemma 2 shows that LHS never has a much larger variance than MC, and an informal description of the results of Lemma 3 below is that

$$V_{\mathrm{LHS}}(\hat{I}) \doteq \frac{1}{n}\Big(\sigma^2 - \sum_{j=1}^{\infty} \sigma^2_{\{j\}}\Big). \qquad (37)$$

So Latin hypercube sampling of the infinite dimensional cube also removes the variance contribution of an additive approximation to the integrand.

Using the ANOVA decomposition of Section 10.3 we write

$$\hat{I} - I = \frac{1}{n} \sum_{i=1}^{n} \sum_{1 \leq |u| < \infty} f_u(X_i) = \sum_{1 \leq |u| < \infty} \hat{I}_u.$$

The integration error in infinite dimensions is a sum of contributions from different ANOVA terms. Because the ANOVA terms are orthogonal, these contributions are uncorrelated under MC. This also holds under LHS.

Proposition 7 *If* $\int_{[0,1)^\infty} f(X)^2 dX < \infty$ *then*

$$V_{\mathrm{LHS}}(\hat{I}) = \sum_{1\le|u|<\infty} V_{\mathrm{LHS}}(\hat{I}_u).$$

Proof: First we give the steps in the proof, then we justify them.

$$\begin{aligned}
V_{\mathrm{LHS}}(\hat{I}) &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{i'=1}^{n}\sum_{1\le|u|<\infty}\sum_{1\le|u'|<\infty} E(f_u(X_i)f_{u'}(X_{i'})) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{i'=1}^{n}\sum_{1\le|u|<\infty} E(f_u(X_i)f_u(X_{i'})) \\
&= E\left(\sum_{1\le|u|<\infty}\left(\frac{1}{n}\sum_{i=1}^{n} f_u(X_i)\right)^2\right) \\
&= \sum_{1\le|u|<\infty} V_{\mathrm{LHS}}(\hat{I}_u).
\end{aligned}$$

The first, third and fourth equalities above follow from expanding and collecting terms. To demonstrate the second equality, we need to show that $E(f_u(X_i)f_{u'}(X_{i'})) = 0$ whenever $u \ne u'$.

Let $u \ne u'$ be finite subsets. Without loss of generality there is some $j_0 \in u$ with $j_0 \notin u'$. The desired result follows because f_u integrates to zero over X^{j_0}. Formally

$$\begin{aligned}
E(f_u(X_i)f_{u'}(X_{i'})) &= E\left(E\left(f_u(X_i)f_{u'}(X_{i'}) \mid X_{i'}^j, j\in v, X_i^j, j\in u-\{j_0\}\right)\right) \\
&= E\left(f_{u'}(X_{i'})E\left(f_u(X_i) \mid X_i^j, j\in u-\{j_0\}\right)\right) \\
&= 0
\end{aligned}$$

regardless of whether $i = i'$. ∎

Lemma 2 *For* $n > 1$ *and* $1 \le d \le \infty$,

$$V_{\mathrm{LHS}}(\hat{I}) \le \frac{n}{n-1} V_{\mathrm{MC}}(\hat{I}).$$

Proof: If $\int f(X)^2 dX = \infty$ then $V_{\mathrm{MC}}(\hat{I}) = \infty$ and the inequality is trivial, though worthless. Suppose that $\int f(X)^2 dX < \infty$. If $d < \infty$ the inequality holds by (13). If $d = \infty$, the inequality follows by the expansion in Proposition 7 followed by a term by term application of (13). ∎

Lemma 3 *If* $\int_{[0,1)^\infty} f(X)^2 dX < \infty$ *then for any* $\epsilon > 0$

$$V_{\mathrm{LHS}}(\hat{I}) \le \frac{1}{n}\left(\sigma^2 - \sum_{j=1}^{\infty} \sigma_{\{j\}}^2 + \epsilon\right) + o\left(\frac{1}{n}\right), \tag{38}$$

and

$$V_{\mathrm{LHS}}(\hat{I}) \ge \frac{1}{n}\left(\sigma^2 - \sum_{j=1}^{\infty} \sigma_{\{j\}}^2 - \epsilon\right) + o\left(\frac{1}{n}\right). \tag{39}$$

Proof: It follows from Proposition 7 that

$$V_{\mathrm{LHS}}(\hat{I}) = \sum_{j=1}^{\infty} V_{\mathrm{LHS}}(\hat{I}_{\{j\}}) + \sum_{1<|u|<\infty} V_{\mathrm{LHS}}(\hat{I}_u). \tag{40}$$

By Lemma 2, the second term in (40) may be bounded by

$$\frac{n}{n-1} \sum_{1<|u|<\infty} V_{\mathrm{MC}}(\hat{I}_u) = \frac{1}{n-1}\left(\sigma^2 - \sum_{j=1}^{\infty} \sigma_{\{j\}}^2\right).$$

We split the first term of (40) into two pieces. For any fixed s, we find that $\sum_{j=1}^{s} V_{\mathrm{LHS}}(\hat{I}_{\{j\}}) = o\,(1/n)$ by applying equation (12) to the additive integrand $\sum_{j=1}^{s} f_{\{j\}}$. Given $\epsilon > 0$, choose s_0 so that $E((f - f^s)^2) < \epsilon$ whenever $s \ge s_0$. For such s, making another comparison to V_{MC} gives

$$\sum_{j=s+1}^{\infty} V_{\mathrm{LHS}}(\hat{I}_{\{j\}}) \le \frac{1}{n-1} \sum_{j=s+1}^{\infty} \int f_{\{j\}}^2 dX \le \frac{1}{n-1} E((f - f^s)^2) < \frac{\epsilon}{n-1}.$$

Putting the pieces together, and using $1/(n-1) = 1/n + o(1/n)$ establishes equation (38). Similar techniques establish equation (39). ∎

10.5 Infinite LSS

Suppose that $k = \infty$. The martingale construction used in Section 10 applies to LSS by taking $f^r = E(f(X)|\mathcal{X}^1, \ldots, \mathcal{X}^r), r = 0, 1, \ldots, \infty$. Here the truncation is from an infinite to a finite number of QMC rules covering input variables that capture virtually all of the variation in f. Thus we can expect good results from LSS even with $k = \infty$.

In Theorems 1 and 2, no special use was made of the fact that all the (R)QMC rules had finite dimension s_r. The only problem with having some $s_r = \infty$ is that it may be impossible to find an (R)QMC method in infinite

dimensions with $\mathcal{E}_n^r = o(n^{-1/2})$. Some recent work by Wasilkowski and Wozniakowski [52] shows that the error rate $O(n^{-0.677})$ is attainable for any dimension $d < \infty$. If this can be extended to $d = \infty$ then LSS with $s_r = \infty$ would work well. There would remain interesting issues in deciding how to partition the input variables to good effect with finite n, when one or more subsets in the partition can have infinitely many inputs.

11 Conclusions

In view of results like Bahvalov's theorem (given in [6]) numerical integration in high dimensions is known to be intractable. This means that whatever method we're using, there are integrands, perhaps even smooth ones, on which we'll get bad results. Sloan and Wozniakowski [43] give another intractability result for high dimensions, where smoothness means rapidly decaying Fourier coefficients. Of course, intractability does not mean that we'll always get bad results in practice.

When good results are obtained in integrating a high dimensional function, we should conclude first of all that an especially tractable integrand was tried and not that a generally successful method has been found. A secondary conclusion is that we might have made a very good choice in selecting an integration method to exploit whatever features of f made it tractable. For example, even if f is virtually linear, simple Monte Carlo will get a bad result if f has a large variance, while some other methods will do very well.

In this paper we've considered methods that can exploit integrands of low effective dimension. Latin hypercube sampling works well on integrands that are largely one dimensional in the superposition sense. Methods based on (R)QMC with padding work well on integrands with low dimensional structure among the variables treated by (R)QMC and, if padded by Latin hypercube sampling, one dimensional structure among the other variables.

Latin supercube sampling allows the practitioner to exploit still more structure in the integrand. If much of the variation is concentrated within groups of a few variables, especially if it is concentrated within low dimensional subsets of those variables, then LSS allows us to exploit that structure by grouping those variables within (R)QMC groups.

It is not reasonable to expect that these methods will be able to turn high dimensional integration, in general, into a tractable problem. But it may turn out that some broad classes of integration problems have their variation concentrated among several small subsets of input variables. And

it may often be possible for practitioners to engineer their integrands in order to concentrate the variation in such subsets.

Acknowledgements

This paper has benefitted greatly from the comments of three referees and the associate editor. This work was supported by grants from the National Science Foundation and the Alcoa Foundation.

References

[1] P. Acworth, M. Broadie, and P. Glasserman. A comparison of some Monte Carlo techniques for option pricing. In H. Niederreiter, P. Hellekalek, G. Larcher, and P. Zinterhof, editors, *Monte Carlo and quasi-Monte Carlo methods '96*. Springer, 1997. Forthcoming.

[2] P. Boyle, M. Broadie, and P. Glasserman. Monte carlo methods for security pricing. *Journal of Economics, Dynamics and Control*, 1996. In press.

[3] P. Bratley, B. L. Fox, and H. Niederreiter. Implementation and tests of low-discrepancy sequences. *ACM Transactions on Modeling and Computer Simulation*, 2(3):195–213, July 1992.

[4] R. E. Caflisch, W. Morokoff, and A. B. Owen. Valuation of mortgage backed securities using brownian bridges to reduce effective dimension. Technical report, Department of Mathematics, UCLA, 1997.

[5] R. Cranley and T. Patterson. Randomization of number theoretic methods for multiple integration. *SIAM Journal of Numerical Analysis*, 13:904–914, 1976.

[6] P. J. Davis and P. Rabinowitz. *Methods of Numerical Integration (2nd Ed.)*. Academic Press, San Diego, 1984.

[7] L. Devroye. *Non-uniform Random Variate Generation*. Springer, 1986.

[8] B. Efron and C. Stein. The jackknife estimate of variance. *Annals of Statistics*, 9:586–596, 1981.

[9] K.-T. Fang and Y. Wang. *Number Theoretic Methods in Statistics*. Chapman and Hall, London, 1994.

[10] H. Faure. Discrépance de suites associées à un système de numération (en dimension s). *Acta Arithmetica*, 41:337–351, 1982.

[11] H. Faure. Méthodes quasi-Monte-Carlo multidimensionnelles. *Theoretical Computer Science*, 123:131–137, 1994.

[12] B. Fox. Generating Poisson processes by quasi-Monte Carlo. Technical report, SIM-OPT Consulting, 872 Timber Lane, Boulder, CO 80304, 1996.

[13] P. Glynn. Some new results on the initial transient problem. Technical report, Stanford University, EESOR Department, 1995.

[14] S. Haber. Numerical evaluation of multiple integrals. *SIAM Review*, 12:481–526, 1970.

[15] F. J. Hickernell. The mean square discrepancy of randomized nets. *ACM Trans. Model. Comput. Simul.*, 6:274–296, 1996.

[16] F. J. Hickernell. A generalized discrepancy and quadrature error bound. *Mathematics of Computation*, 1997. To appear.

[17] L. Hua and Y. Wang. *Applications of number theory to numerical analysis.* Springer, Berlin, 1981.

[18] S. Joe. Randomization of lattice rules for numerical multiple integration. *Journal of Computational and Applied Mathematics*, 31:299–304, 1990.

[19] C. Joy, P. Boyle, and K.-S. Tan. Quasi-Monte Carlo methods in numerical finance. *Management Science*, 42(6):926–938, 1996.

[20] S. Kakutani. Two-dimensional Brownian motion harmonic functions. *Proc. Imp. Acad. (Tokyo)*, 20:706–714, 1944.

[21] A. Keller. A Quasi-Monte Carlo Algorithm for the Global Illumination Problem in a Radiosity Setting. In H. Niederreiter and P. J.-S. Shiue, editors, *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing*, pages 239–251, New York, 1995. Springer-Verlag.

[22] A. Kersch and W. Morokoff. *Transport simulation in microelectronics.* Birkhäuser, Basel, 1995.

[23] J. Koehler and A. Owen. Computer experiments. In S. Ghosh and C. Rao, editors, *Handbook of Statistics, 13: Design and Analysis of Experiments*.

[24] M. D. McKay, R. J. Beckman, and W. J. Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239–45, 1979.

[25] W. Morokoff and R. Caflisch. A Quasi-Monte Carlo approach to particle simulation of the heat equation. *SIAM Journal on Numerical Analysis*, 30:1558–1573, 1993.

[26] W. Morokoff and R. Caflisch. Quasi-random sequences and their discrepancies. *SIAM J. Sci. Stat. Computing*, 15:1251–1279, 1994.

[27] W. Morokoff and R. Caflisch. Quasi-Monte Carlo integration. *J. Comp. Phys.*, 122:218–230, 1995.

[28] H. Niederreiter. Point sets and sequences with small discrepancy. *Monatshefte fur mathematik*, 104:273–337, 1987.

[29] H. Niederreiter. *Random Number Generation and Quasi-Monte Carlo Methods*. S.I.A.M., Philadelphia, PA, 1992.

[30] H. Niederreiter and C. Xing. Low-discrepancy sequences and global function fields with many rational places. *Finite Fields and Their Applications*, 2:241–273, 1996.

[31] G. Okten. A probabilistic result on the discrepancy of a hybrid-Monte Carlo sequence and applications. Technical report, Mathematics Department, Claremont Graduate School, 1996.

[32] A. B. Owen. Orthogonal arrays for computer experiments, integration and visualization. *Statistica Sinica*, 2:439–452, 1992.

[33] A. B. Owen. Lattice sampling revisited: Monte Carlo variance of means over randomized orthogonal arrays. *The Annals of Statistics*, 22:930–945, 1994.

[34] A. B. Owen. Randomly permuted (t, m, s)-nets and (t, s)-sequences. In H. Niederreiter and P. J.-S. Shiue, editors, *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing*, pages 299–317, New York, 1995. Springer-Verlag.

[35] A. B. Owen. Monte Carlo variance of scrambled equidistribution quadrature. *SIAM Journal of Numerical Analysis*, 34(5), 1997. In press.

[36] A. B. Owen. Scrambled net variance for integrals of smooth functions. *Annals of Statistics*, 25, 1997. In press.

[37] S. Paskov. New methodologies for valuing derivatives. In S. Pliska and M. Dempster, editors, *Mathematics of Derivative Securities*. Isaac Newton Inst., Cambridge U. Press, 1996.

[38] S. Paskov and J. Traub. Faster valuation of financial derivatives. *J. Portfolio Manag.*, pages 113–120, 1995.

[39] H. D. Patterson. The errors of lattice sampling. *J.R.S.S.-B*, 16:140–149, 1954.

[40] M. N. O. Sadiku. *Numerical techniques in electromagnetics*. CRC Press, 1992.

[41] P. K. Sarkar and M. A. Prasad. A comparative study of pseudo and quasi random sequences for the solution of integral equations. *Journal of Computational Physics*, 68:66–88, 1987.

[42] I. H. Sloan and S. Joe. *Lattice Methods for Multiple Integration*. Oxford Science Publications, Oxford, 1994.

[43] I. H. Sloan and H. Wozniakowski. An intractability result for multiple integration. Technical report, University of New South Wales, School of Mathematics, 1995.

[44] J. Spanier. Quasi-Monte Carlo Methods for Particle Transport Problems. In H. Niederreiter and P. J.-S. Shiue, editors, *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing*, pages 121–148, New York, 1995. Springer-Verlag.

[45] M. Stein. Large sample properties of simulations using Latin hypercube sampling. *Technometrics*, 29(2):143–51, 1987.

[46] A. Takemura. Tensor analysis of ANOVA decomposition. *Journal of the American Statistical Association*, 78:894–900, 1983.

[47] S. Tezuka. *Uniform random numbers: theory and practice*. Kluwer Academic Publishers, Boston, 1995.

[48] B. Tuffin. On the use of low discrepancy sequences in Monte Carlo methods. Technical Report 1060, I.R.I.S.A., Rennes, France, 1996.

[49] E. J. J. van Rensburg and G. M. Torrie. Estimation of multidimensional integrals: is Monte Carlo the best method? *Journal of Physics A: Math. Gen.*, 26:943–953, 1993.

[50] E. Veach and L. Guibas. Bidirectional estimators for light transport. In *5th Annual Eurographics Workshop on Rendering*, pages 147–162, June 13–15 1994.

[51] G. Wahba. *Spline Models for Observational Data*. CBMS-NSF, SIAM, Philadelphia, 1990.

[52] H. Wasilkowki and G. Wozniakowski. The exponent of discrepancy is at most 1.4778.... Technical report, Columbia University, Department of Computer Science, 1996.

[53] D. Williams. *Probability with Martingales*. Cambridge University Press, Cambridge, 1991.