Alternative Sampling Methods for Estimating Multivariate Normal Probabilities

Zsolt Sándor[1]
Erasmus University Rotterdam

and

Péter András
University of Newcastle

Abstract

We study the performance of alternative sampling methods for estimating multivariate normal probabilities through the GHK simulator. The sampling methods are randomized versions of some quasi-Monte Carlo samples (Halton, Niederreiter, Niederreiter-Xing sequences and lattice points) and some samples based on orthogonal arrays (Latin hypercube, orthogonal array and orthogonal array based Latin hypercube samples). In general, these samples turn out to have a better performance than Monte Carlo and antithetic Monte Carlo samples. Improvements over these are large for low-dimensional (4 and 10) cases and still significant for dimensions as large as 50.

Key words: Simulation; Quasi-Monte Carlo; (t, m, s)-net; Lattice Points; Multinomial Probit

JEL Classification: C150; C350

Econometric Institute Report EI 2003-05

[1]Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands. E-mail: sandor@few.eur.nl, Phone: +31 10 408 1416, Fax: +31 10 408 9162.

1. Introduction

The Monte Carlo method for estimating high-dimensional integrals has received much attention in the recent econometric literature, especially applied to integrals arising from multivariate normal probabilities. This method replaced quadrature methods, which are cumbersome if the integral has dimension higher than five (Geweke, 1996). In the econometric literature the first attempt of estimating normal probabilities was by Lerman and Manski (1981), who employed a simple frequency simulator of the probabilities. This was followed by several improvements later on (e.g., McFadden, 1989 and Stern, 1992). In the early 90's Geweke, Hajivassiliou and Keane as well as researchers from other fields (e.g., Genz, 1992) independently developed a simulator, known by econometricians as the GHK simulator. Hajivassiliou et al. (1996), using Monte Carlo sampling, and Vijverberg (1997), using antithetic Monte Carlo sampling, find this simulator to have the best performance in comprehensive simulation studies. The performance is measured by the precision of the integral estimate.

The GHK simulator is based on sampling recursively from the truncated normal distribution, which on its turn implies sampling from the uniform distribution. Samples from the uniform distribution are usually obtained by generating so-called pseudo-random numbers on the computer. In the late 50's, about one decade after the systematic development of the Monte Carlo method started, some researchers turned to replacing pseudo-random numbers by deterministic numbers. These deterministic numbers, called quasi-Monte Carlo samples, were typically constructed using number theoretic methods (e.g., Korobov, 1959, Halton, 1960, Sobol', 1967). Later, randomized versions of these were developed (e.g., Cranley and Patterson, 1976, Owen, 1995), which made it possible to compare their performance directly to the performance of pseudo-random numbers.

Parallel to the development of quasi-Monte Carlo sequences, samples generated with semi-deterministic methods were developed in the statistical literature. A class of these methods is based on orthogonal arrays, which are matrices with a certain combinatorial structure. Some of the pioneering work was done by Patterson (1954), who developed lattice sampling. While lattice samples are not exactly uniform, McKay et al. (1979) developed Latin hypercube samples that are dependently uniformly distributed. Latin hypercube sampling was generalized by Owen (1992) and Tang (1993) for orthogonal arrays. Owen (1997a) reveals connections between sampling based on orthogonal arrays and randomized versions of quasi-Monte Carlo samples, and shows the superiority of these over Monte Carlo samples when applied for integration.

In spite of the work cited above and several other works in numerical mathematics and computational statistics that show the superiority of quasi-Monte Carlo sampling in different specific cases, its application to problems in econometrics is rather rare. Geweke (1996) and Judd (1998) present some of the quasi-Monte Carlo methods at a descriptive level. Some

published work employing quasi-Monte Carlo exists in the field of financial economics; we refer to Boyle et al. (1997), who use deterministic, and Tan and Boyle (2000), who employ randomized quasi-Monte Carlo samples. Other applications (Bhat, 2001a, 2001b, Train, 2000, 2002) employ the Halton sequence or its randomized version (described in section 2.4 below) for estimating probabilities arising in logit models with random coefficients (Revelt and Train, 1998). In a more recent work Sándor and Train (2002) compare several different quasi-Monte Carlo samples for these models.

No comprehensive analysis of these samples is available for estimating multivariate normal probabilities in cases of interest in econometrics. In this paper we attempt to fill some of the gap in the literature by studying the performance of samples based on orthogonal arrays and randomized quasi-Monte Carlo samples for estimating integrals arising from multivariate normal probabilities. We present these sampling methods in section 2. We also provide comparisons and mention some existence and construction issues there. In section 3 we present results on the variance of integral estimates for these samples. Here we also provide intuitive arguments why these methods are expected to work well for integration. Section 4 describes the GHK simulator that estimates multivariate normal probabilities. The main contribution of the paper is presented in section 5 as the outcome of a simulation study of the samples applied to multivariate normal probabilities. The results show substantial improvements in precision of the integral estimates over Monte Carlo and antithetic Monte Carlo sampling for most of the samples presented. We compare the performance of the different samples and make recommendations for their future use. We conclude the paper and discuss topics of interest for future research in section 6.

2. Sampling methods

In this section we present sampling methods based on orthogonal arrays and quasi-Monte Carlo sampling. After the first subsection containing some preliminaries, in the second subsection we present orthogonal array, Latin hypercube and orthogonal array based Latin hypercube sampling. Then in subsequent subsections we describe three randomized quasi-Monte Carlo samples, namely, randomized (t, m, s)-nets, randomized Halton sequences and randomly shifted lattice points. In the last subsection we provide some comparisons of the samples and discuss some existence and construction issues.

2.1. Preliminaries

We introduce briefly the Monte Carlo (MC hereafter) method for integration formally. Let

$$I = \int_{[0,1)^s} f(x)\, dx. \tag{2.1}$$

be the integral of the real-valued function f defined on $[0,1)^s$. An estimate of this integral is

$$\widehat{I} = \frac{1}{n}\sum_{i=1}^{n} f(x_i), \tag{2.2}$$

where $(x_i)_{i=1}^n$ is a set of points from $[0,1)^s$. If the elements of the sequence have the uniform distribution on $[0,1)^s$ then, under the condition that the integral exists, we have $\mathrm{E}\left(\widehat{I}\right) = I$, that is, the estimator is unbiased, no matter the elements of the sequence are independent or not. The MC method uses such an estimate of I by taking the sequence (x_i) independently uniformly distributed on $[0,1)^s$.

Intuitively, the estimate is more precise the better the finite sample $(x_i)_{i=1}^n$ approximates the continuous uniform distribution. We refer to the quality of this approximation as the equidistribution property of the sample. In other words, we expect samples with a better equidistribution property to estimate integrals more precisely. We illustrate this in Figure 1A.

[FIGURE 1 ABOUT HERE]

The figure displays a random uniform sample of 9 points in the unit square. We can estimate the integral of a function defined on this set by using this sample. Note, however, that there is no point of the sample in the small squares $[0,1/3)^2$ and $[2/3,1)^2$. Therefore, if the function in these small squares has values that are very different from the average, then the estimate of the integral will not be precise. This deficiency can be avoided by, for example, using a sample that has one point in each of the small squares. Then we say that this latter sample has a better equidistribution property than that from Figure 1A. In section 2.3 we give a formal definition of the notion equidistribution property.

2.2. Samples Based on Orthogonal Arrays

In this subsection we present orthogonal array sampling introduced by Owen (1992) that generalizes Latin hypercube sampling (McKay et al., 1979), which we also present, and orthogonal array based Latin hypercube sampling introduced by Tang (1993). First we need to introduce orthogonal arrays. Orthogonal arrays (OA's hereafter) are matrices whose elements are nonnegative integers with a certain combinatorial structure. They have been used especially in designing computer experiments and are still intensively researched (see, e.g., Hedayat, Sloane and Stufken, 1999).

Let A be an $n \times s$ matrix with elements a_{ij} in the set $\{0,1,...,b-1\}$. A is called an OA of strength $m \leq s$ if in every $n \times m$ submatrix of A the b^m possible distinct rows appear exactly once. Then obviously $n = b^m$. Such an OA is denoted $\mathrm{OA}(n,s,b,m)$. The maximal strength of A is the largest m for which A has strength m.

An OA sample (Owen, 1992) constructed from the OA(n, s, b, m) with elements a_{ij} has the elements

$$x_{ij} = \frac{\pi_j(a_{ij}) + u_{ij}}{b},\ i = 1, ..., n,\ j = 1, ..., s, \tag{2.3}$$

where π_j is a uniform random permutation of $0, ..., b-1$, u_{ij} is a uniform random variable on $[0, 1)$, and the s permutations and ns random variables are mutually independent. So an OA sample attaches a randomly determined permutation to the elements of each column. All x_{ij} are in the interval $[0, 1)$, have the uniform distribution and are mutually independent. The permuted elements $\pi_j(a_{ij})$ of the OA still form an OA(n, s, b, m). In practice often one needs only one sample. In this case the random permutations can be omitted. An OA sample constructed from an OA$(9, 2, 3, 2)$ is presented in Figure 1B. Note that the criterion for a 9-point sample to be an OA sample is that each small square of area $1/9$ contains exactly one point of the sample.

A Latin hypercube (LH hereafter) is an OA with maximal strength 1, that is, OA$(b, s, b, 1)$. Such a LH is a $b \times s$ matrix whose columns are permutations of its symbols $0, 1, ..., b-1$. Then a LH sample (McKay et al., 1979) corresponding to this is given by

$$x_{ij} = \frac{\pi_j(i) + u_{ij}}{b},\ i = 0, 1, ..., b-1,\ j = 1, ..., s, \tag{2.4}$$

where π_j and u_{ij} are of the same type as in (2.3). Though the arguments of $\pi_j(\cdot)$ seem to be different in the two formulas, they are in fact the numbers $0, 1, ..., b-1$ in both. Thus the uniform distribution and independence properties of the OA sample are also valid for the LH sample. Figure 1C displays the points of a LH sample. The criterion for a 9-point sample to be a LH sample is that each rectangle of sides with length 1 and $1/9$ contains exactly one point of the sample. Hence a LH sample has better equidistribution property than a random sample. Still, as shown in Figure 1C, for a LH sample it is not guaranteed that all small squares of volume $1/9$ contain exactly one point of the sequence.

For the sample size $n = b^m$, the elements of a LH take n values while the elements of an OA(b^m, s, b, m) take only b values. This implies that the LH sample has a better equidistribution property than the OA sample along its one-dimensional components. So even though the OA sample has attractive properties due to the combinatorial structure, its rougher one-dimensional parts may weaken its equidistribution property.

This problem was addressed by Tang (1993) who developed a sampling scheme based on OA's transformed into LH's while preserving their combinatorial structure. The new structure obtained is called OA based LH. A description of the scheme follows. Let a_{ij} be the elements of an OA of maximal strength $m \geq 1$ with $n = b^m$ rows. First we apply a random permutation π to the elements of the OA in the sense that we replace each element a_{ij} by $\pi(a_{ij})$; then we randomly permute the s columns of the obtained matrix. Each column j of this matrix contains each symbol $a \in \{0, 1, ..., b-1\}$ exactly b^{m-1} times. If a given a

is the k'th among the b^{m-1} pieces of a's in column j then we replace a by $ab^{m-1}+\pi_{ja}(k)$, where π_{ja} is a random permutation of $0,1,...,b^{m-1}-1$. All the π_{ja} permutations are assumed to be mutually independent. Denote the elements of the obtained matrix by c_{ij}. Then each c_{ij} takes each value $0,1,...,b^m-1$ with the same probability and the elements belonging to any column of the matrix (c_{ij}) take all these values. Hence (c_{ij}) is a LH with b^m symbols.

An OA based LH sample is defined similarly to the samples above by adding a uniform random variable to each element and dividing by n:

$$x_{ij}=\frac{c_{ij}+u_{ij}}{n},\ i=1,...,n,\ j=1,...,s. \tag{2.5}$$

Similarly to the other two samples, all x_{ij} are in the interval $[0,1]$, have the uniform distribution and are mutually independent. Figure 1D presents the points of a OA based LH based on the OA

$$\begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 2 & 2 & 2 \\ 0 & 1 & 2 & 0 & 1 & 2 & 0 & 1 & 2 \end{bmatrix}',$$

where all permutations are the identical permutation. Note that the criterion for a sample of 9 points to be an OA based LH sample is that it satisfies the criteria for both OA and LH samples. Therefore, OA based LH samples have a better equidistribution property than either of the other two samples.

2.3. Randomized (t,m,s)-nets

In this subsection we discuss (t,m,s)-nets and their randomization. The notion of (t,m,s)-net was introduced by Niederreiter (1987) based on the previous work by Sobol' (1967) and Faure (1982), while the methods of randomization we present were developed by Owen (1995), Matoušek (1998) and Hong and Hickernell (2001).

(t,m,s)-nets. A (t,m,s)-net is a set of points from $[0,1)^s$ that satisfy certain equidistribution property, namely that all subintervals of $[0,1)^s$ of a certain type contain a given number of points of the sequence. This equidistribution property ensures that the sequence approximates closely the continuous uniform distribution on $[0,1)^s$.

The subintervals considered by Niederreiter (1987) are of the type

$$E=\prod_{j=1}^{s}\left[\frac{t_j}{b^{k_j}},\frac{t_j+1}{b^{k_j}}\right), \tag{2.6}$$

where b is a positive integer, k_j are nonnegative integers and t_j integers with $0\le t_j<b^{k_j}$. Then E is a hyperrectangle of volume $b^{-(k_1+...+k_s)}$ and is called elementary interval in base b.

Let t and m be nonnegative integers with $t<m$. A finite sequence $a_1,...,a_n\in[0,1)^s$ with $n=b^m$ is called a (t,m,s)-net in base b if every elementary interval in base b of volume

b^{t-m} contains exactly b^t points of the sequence. Note that for the same m, s and b smaller values of t imply a better equidistribution property of the net.

The equidistribution property of nets can be extended to infinite sequences in the following way. Let t be a nonnegative integer. A sequence $a_1, a_2... \in [0,1)^s$ is a (t,s)-sequence in base b if for all $m \geq 0$ and all $k \geq 0$ the finite sequence $a_{kb^m+1}, ..., a_{(k+1)b^m}$ is a (t,m,s)-net in base b.

(t,m,s)-nets in base b are typically constructed through appropriate so-called generator matrices $C_1, ..., C_s$. Each generator matrix is of size $m \times m$ and its elements are from the set $\{0,1,...,b-1\}$. The elements of a (t,m,s)-net in base b are obtained in the following way. We write each $i \in \{0,1,...,b^m-1\}$ in base b representation

$$i = d_{m-1}...d_0 = \sum_{k=0}^{m-1} d_k b^k,$$

where $d_k \in \{0,1,...,b-1\}$ for $k = 0,1,...,m-1$. Denote the i-th member of the (t,m,s)-net in base b by $a_i = (a_{i1}, ..., a_{is})$. The element of this corresponding to dimension j is defined as

$$a_{ij} = \left(b^{-1}, ..., b^{-m}\right) C_j \left(d_0, ..., d_{m-1}\right)', \tag{2.7}$$

where the operations of vector and matrix multiplication are done in a special way. Namely, if b is a prime number then the operations should be done modulo b, that is, the remainder of the division by b should be computed. If b is not prime but the power of a prime then the operations should be done in the field with b elements. This way, $C_j \left(d_0, ..., d_{m-1}\right)'$ is a vector with elements in $\{0,1,...,b-1\}$, and these elements be the digits of the base b representation of a_{ij}. The definition of the element a_{ij} shows that the generator matrices fully determine a (t,m,s)-net. Constructing the generator matrices is not an easy task, and therefore, presenting it is beyond the scope of this paper. For constructions that we use in this paper we refer to Niederreiter (1988) and Niederreiter and Xing (1996)

Randomization. We treat two methods for randomizing (t,m,s)-nets. The first was developed by Owen (1995), while the second by Matoušek (1998) and Hong and Hickernell (2001). Suppose that $a_1, ..., a_n$ is a (t,m,s)-net in base b. We can write each element a_{ij} of each member a_i of this sequence in base b representation, that is, $a_{ij} = \sum_{k=1}^{\infty} a_{ijk} b^{-k}$, with the digits $a_{ijk} \in \{0,1,...,b-1\}$. The randomization should produce a sequence with uniformly distributed elements that retain the (t,m,s)-net property.

We first describe the randomization procedure developed by Owen (1995), and we do so for a generic point $a = (a_1, ..., a_s) \in [0,1)^s$. Suppose that $a_j = \sum_{k=1}^{\infty} a_{jk} b^{-k}$. A randomized version of a is the vector x, whose elements have the following digits:

$$\begin{aligned} x_{j1} &= \pi_j\left(a_{j1}\right), \; x_{j2} = \pi_{ja_{j1}}\left(a_{j2}\right), \; x_{j3} = \pi_{ja_{j1}a_{j2}}\left(a_{j3}\right), ..., \\ x_{jm} &= \pi_{ja_{j1}a_{j2}...a_{jm-1}}\left(a_{jm}\right), \end{aligned} \tag{2.8}$$

where $\pi_{ja_{j1}a_{j2}...a_{jk}}$ for $k = 1, ..., m-1$ is a uniform random permutation of $\{0, 1, ..., b-1\}$. So the permutation used for randomizing the k'th digits of the elements belonging to the same dimension depends on whether the previous $k-1$ digits of these elements' base b representation are equal. We show below with a simple example why this is necessary.

We illustrate randomization in Figure 1D. Note that the sample in this figure, besides being an OA based LH, is also a $(0, 2, 2)$-net in base 3. Take the horizontal dimension and let it correspond to $j = 1$. Since the first digits in base b representation determine to which rectangle of size $1/3 \times 1$ the points belong, permuting the first digits means that we permute the places of the three large rectangles of size $1/3 \times 1$ by a uniform random permutation π_1 of $0, 1, 2$. The second digits in base b representation determine to which rectangle of size $1/9 \times 1$ within the large rectangle the point belongs. Hence permuting the second digits means that we permute the three small rectangles of size $1/9 \times 1$ within the large rectangles, and we do so by using three independent random permutations π_{10}, π_{11}, π_{12}. More precisely, the permutation π_{1a} for $a = 0, 1, 2$ is used for the large rectangle $\left[\frac{a}{3}, \frac{a+1}{3}\right) \times [0, 1)$. Therefore, the permutations π_{1a} of the second decimal digits depend on the first decimal digit a.

The randomization defined above permutes the first m digits in base b representation of a number. This way we obtain points whose coordinates are from the set $\left\{0, \frac{1}{b^m}, \frac{2}{b^m}, ..., \frac{b^m-1}{b^m}\right\}$. This is similar to $\frac{c_{ij}}{b^m}$ that we obtain in the OA based LH samples in (2.5). In order to make these coordinates uniform random on $[0, 1)$, similar to (2.5), we add to them $\frac{u_{ij}}{b^m}$, where u_{ij} is uniform random on $[0, 1)$. This way we obtain randomized (t, m, s)-nets that inherit the equidistribution property of the original nets and contain points that are uniformly distributed. Due to the latter property the integral estimates are unbiased and their standard deviations can be estimated. These properties, proved in Owen (1995), are stated in the following propositions.

PROPOSITION 1. ***Equidistribution of randomization****: If $(a_i)_{i=1}^{b^m}$ is a (t, m, s)-net in base b, then a randomized version $(x_i)_{i=1}^{b^m}$ of it is also a (t, m, s)-net in base b with probability 1. The same is true for (t, s)-sequences.*

PROPOSITION 2. ***Uniformity of randomization****: If a is a point in $[0, 1)^s$ then its randomly permuted version x has the uniform distribution on $[0, 1)^s$.*

The second method for randomizing (t, m, s)-nets was motivated by the fact that Owen's method is fairly slow in practical applications. This is due to the feature that the random permutations corresponding to a digit depend on the previous digits. Let $L_1, ..., L_s$ be lower triangular $m \times m$ matrices with random diagonal elements chosen uniformly from $\{1, 2, ..., b-1\}$ and the other elements chosen random uniformly from $\{0, 1, ..., b-1\}$. Let $e_1, ..., e_s$ be random vectors $m \times 1$ with elements chosen uniformly from $\{0, 1, ..., b-1\}$.

Then the method developed by Matoušek (1998) and Hong and Hickernell (2001), called random linear scrambling, defines

$$c_{ij} = \left(b^{-1}, ..., b^{-m}\right) \left[L_j C_j \left(d_0, ..., d_{m-1}\right)' + e_j\right],$$

as the randomized version of a_{ij}, where the operations are done in the specific way explained below equation (2.7). Hence $c_{ij} \in \left\{0, \frac{1}{b^m}, ..., \frac{b^m-1}{b^m}\right\}$, which can be made uniform by taking $x_{ij} = c_{ij} + \frac{u_{ij}}{b^m}$ with u_{ij} a uniform random number on $[0, 1)$. Matoušek (1998) and Hong and Hickernell (2001) show that random linear scrambling yields randomized (t, m, s)-nets with the equidistribution property of the original nets and contain points that are uniformly distributed. This type of randomization operates in a similar way as Owen's randomization but yields (t, m, s)-nets that are less random. This can be seen from the example of $(0, 2, 2)$-nets in base 3 (Figure 1D) by noting that the second digits in base 3 representation are not permuted by independent random permutations. In spite of the fact that linearly scrambled (t, m, s)-nets are less random, they are likely to be as useful as (t, m, s)-nets randomized by Owen's method. This was shown for a particular subclass, that is, $(0, m, s)$-nets, by Matoušek (1998, Proposition 3.1).

2.4. Randomized Halton sequences

In this subsection we present sequences proposed by Halton (1960) and their randomization developed by Wang and Hickernell (2000). Halton sequences are similar to (t, s)-sequences in that they manipulate the digits of numbers in certain base representations.

Halton sequences. First we show how Halton sequences are defined in one dimension and then extend it to several dimensions. We write any nonnegative integer i in base b as

$$i = d_m ... d_0 = \sum_{k=0}^{m} d_k b^k,$$

where $d_k \in \{0, 1, ..., b-1\}$ for $k = 0, 1, ..., m$. The i-th member of the base b Halton sequence is defined by

$$H_b(i) = 0.d_0 ... d_m \text{ (in base } b) = \sum_{k=0}^{m} d_k b^{-k-1}.$$

That is, we write the base b representation of the number i, reverse the order of its digits and put a decimal point in front of them. The result is a number between 0 and 1 that is by definition the i-th member of the one-dimensional base b Halton sequence.

The multi-dimensional Halton sequence can be obtained by generating several one-dimensional Halton sequences corresponding to bases that are prime numbers. More precisely, we take the first s prime numbers $p_1, ..., p_s$, generate the corresponding one-dimensional

Halton sequence and use these to form the s-dimensional Halton sequence:

$$x_i = \left(H_{p_1}(i), ..., H_{p_s}(i)\right), \ i = 0, 1, ...$$

As noted by Niederreiter (1992, Remark 4.38), all one-dimensional components of this sequence are $(0, 1)$-sequences in the corresponding bases. This makes them closely related to LH samples. However, correlation between two one-dimensional components is not controlled for in the latter. Since the one-dimensional Halton sequences are generated taking bases that are prime numbers, and hence mutually relative primes, the Halton sequence is expected to have lower correlations between its one-dimensional components.

Randomization. The randomization of Halton sequences as introduced by Wang and Hickernell (2000) is based on a recursive relation that holds for the Halton sequence. This relation translates the recursion from i to $i + 1$ into $H_b(i)$ and $H_b(i + 1)$ in a natural way.

Formally, let $x_0 \in [0, 1)$ with the base b representation $x_0 = \sum_{k=0}^{\infty} d_k b^{-k-1}$. Define the expression

$$T_b(x_0) = (1 + d_h) b^{-h-1} + \sum_{k \geq h} d_k b^{-k-1},$$

where $h = \min\{k : d_k \neq b - 1\}$. Then we can define the sequence $(T_b^i(x_0))$ by

$$\begin{aligned} T_b^i(x_0) &\equiv T_b\left(T_b^{i-1}(x_0)\right) \text{ for } i \geq 1 \text{ and} \\ T_b^0(x_0) &\equiv x_0. \end{aligned}$$

Note that with $x_0 = 0$ the above sequence is exactly the one-dimensional Halton sequence in base b. Further, if the starting term can be written as a finite sum $x_0 = \sum_{k=0}^{m} d_k b^{-k-1}$ yielding $x_0 = 0.d_0...d_m$ (in base b) and denoting the corresponding integer $i_0 = d_m...d_0$ (in base b) then $x_0 = H_b(i_0)$ and $T_b^i(x_0) = H_b(i_0 + i)$ for $i \geq 1$. That is, if the starting term of the sequence $(T_b^i(x_0))$ can be written as a finite sum, then the sequence is the same as the Halton sequence of which first i_0 elements are skipped.

Randomized Halton sequences are defined as the above sequences having their starting point random. More precisely, let $x_0 \in [0, 1)$ have the uniform distribution. The randomized one-dimensional Halton sequence in base b is defined by $x_i = T_b^i(x_0)$ for $i = 1, 2, ...$ For the s-dimensional Halton sequence let now $x_0 = (x_{01}, ..., x_{0s}) \in [0, 1)^s$ have the uniform distribution. Then the randomized s-dimensional Halton sequence is defined by

$$x_i = \left(T_{p_1}^i(x_{01}), ..., T_{p_s}^i(x_{0s})\right) \text{ for } i = 1, 2, ... \tag{2.9}$$

Note that according to the remarks from the previous paragraph randomized Halton sequences can also be defined as the deterministic Halton sequences described above by skipping randomly a number of initial terms. Wang and Hickernell (2000) show that the elements of a randomized Halton sequence with a uniform random starting point are uniform.

PROPOSITION 3. *If $x_0 \in [0,1)^s$ is a uniform random vector then x_i defined by (2.9) has the uniform distribution on $[0,1)^s$ for any $i \geq 1$.*

In practice one cannot use a uniformly distributed starting point since its base b representation generally has infinite number of digits. However, if b^m is sufficiently large, where m is the number of digits used in base b representation, then truncating each starting uniform random number by omitting its digits from $m+1$ on, we obtain numbers that approximate uniform numbers fairly well. We return to this problem in section 5 and explain there how we proceeded in practice.

2.5. Randomly shifted lattice points

Lattice points were introduced by Korobov (1959), Bakhvalov (1959) and Hlawka (1962) and further developed to a systematic theory by Sloan and his collaborators (see the references in Sloan and Joe, 1994). The randomization of the lattice points by random shifting was proposed by Cranley and Patterson (1976).

Lattice points. The simplest type of lattice points (in the literature referred to as rank-1 lattice points) are defined as

$$x_i = \left(\left\{ \frac{ig_1}{n} \right\}, ..., \left\{ \frac{ig_s}{n} \right\} \right) \text{ for } i = 0, 1, ..., n-1, \tag{2.10}$$

where n is the number of lattice points, g_j, $j = 1, ..., s$, are positive integers, typically taken to be relative prime with n, and the symbol $\{x\}$ means the fractional part of the number x, that is, $\{x\} = x - \lfloor x \rfloor$. According to their definition, lattice points are easy to generate, provided that we know $g = (g_1, ..., g_s)$.

Procedures for obtaining appropriate g vectors are based on minimizing the integration error. In this paper we treat two criteria regarding the integration error, the worst-case integral error (see, e.g., Sloan and Joe, 1994), commonly denoted by P_α, and the spectral test (Entacher et al. 2000). Both criteria are based on the Fourier series representation of the integrand. Fourier series are helpful in expressing the integration error of lattice points because lattice points are especially suited for integrating periodic functions. We give more detail on this in section 3.3. Since it is rather difficult to search for the minimum in all s dimensions, it is useful to consider lattices of so-called Korobov type where $g = (1, q, q^2, ..., q^{s-1})$. In this situation the search is done with respect to only one variable. Since carefully determined Korobov type lattice points are expected to have a good performance, we are content to consider in this paper only this type of lattice points.

Randomization. Randomly shifted lattices play the role of randomized nets. A randomly shifted lattice has the points

$$x_i = \left(\left\{ \frac{ig_1}{n} + u_1 \right\}, ..., \left\{ \frac{ig_s}{n} + u_s \right\} \right) \text{ for } i = 0, 1, ..., n-1, \tag{2.11}$$

where $u_1, ..., u_s$ are independent random uniform numbers on $[0, 1)$. We note that if g_j and n are relative primes then $\left\{\frac{ig_j}{n}\right\}$ for $i = 0, 1, ..., n-1$ take all values of $0, \frac{1}{n}, ..., \frac{n-1}{n}$, though probably in a different order. In other words, a one-dimensional lattice has the points of the regular grid $0, \frac{1}{n}, ..., \frac{n-1}{n}$. When shifted by $u \in [0, 1)$ the lattice preserves its grid structure but it is translated by a number $c \in \left[0, \frac{1}{n}\right)$. That is, the points of the shifted lattice will be $c, \frac{1}{n} + c, ..., \frac{n-1}{n} + c$. This argument shows that a random shift preserves the regular grid structure of lattice points.

The randomly shifted lattice points defined in (2.11) can be regarded as having been drawn from the uniform distribution if we replace i by $\pi(i)$ in (2.11), where π is a random uniform permutation of $0, 1, ..., n-1$. To see that this statement is true consider the j'th component $x_{ij} = \left\{\frac{\pi(i)g_j}{n} + u_j\right\}$ of x_i. Following the ideas from the previous paragraph this can be written as $x_{ij} = \frac{\varpi(i)}{n} + c_j$, where ϖ is a random uniform permutation of $0, 1, ..., n-1$ and $c_j \in \left[0, \frac{1}{n}\right)$ is a random uniform number. Therefore, x_{ij} has the uniform distribution on $[0, 1)$ (there is some similarity to LH samples; see (2.4)). The fact that the lattice point $x_i \in [0, 1)^s$ is uniform random follows from the independence of $u_1, ..., u_s$. The reason that in the definition of lattice points no random permutation is involved is that this only changes the order of the lattice points randomly, and the Monte Carlo sum does not depend on the order of the points. Due to the fact that the randomly shifted lattice points can be regarded as uniformly distributed in $[0, 1)^s$, the integral estimate using these points will be unbiased.

2.6. Properties, existence and construction

In section 2.2 while presenting LH, OA and OA based LH samples we also compared them regarding their equidistribution property. Here we discuss further properties of Halton, LH, OA based LH samples, randomized (t, m, s)-nets and randomly shifted lattice points. We also provide some information on the existence and construction of $(0, m, s)$-nets.

Properties of the samples. As we mention in section 2.4, each one-dimensional component of a Halton sequence is a $(0, 1)$-sequence in the prime number base in which it was generated. This suggests that it is related to $(0, 1, s)$-nets. A one-dimensional base b Halton sequence of size n, if b does not divide n, is a reunion (in the sense of reunion of sets) of $\lfloor n/b \rfloor$ pieces of $(0, 1, 1)$-nets in base b and a number of points that we refer to as truncated net, where $\lfloor x \rfloor$ is the largest integer not greater than x. If the dimension of the Halton sequence is not very large then the prime numbers in which the one-dimensional components are generated will not be very large and there will be many $(0, 1, 1)$-nets of a moderate size in the reunion. Hence the truncated net has a small size and does not play an important role in this component. Then we expect this Halton sequence to have a good equidistribution property, better than $(0, 1, s)$-nets, due to the low correlation between the components. This latter feature is a consequence of the fact that the components are generated using bases that are mutually relative primes. If the dimension of the Halton

sequence is large then the prime bases will be large and it is likely that the prime bases do not divide the sample size. Moreover, the size of the truncated net becomes large relative to the size of the component nets. For example, the 40'th prime number is 173. If $n = 1000$ then there will be $\lfloor 1000/173 \rfloor = 5$ pieces of $(0, 1, 1)$-nets, each of size 173, and a truncated net of size 135. This large truncated net tends to destroy the attractive equidistribution property of nets. Therefore, we expect Halton sequences with large dimension to have a poorer equidistribution property than $(0, 1, s)$-nets.

There is a simple relationship between $(0, m, s)$-nets and OA's. If a_{ij} are the elements of a $(0, m, s)$-net in base b with $m < s$, then $\lfloor ba_{ij} \rfloor$ are the elements of an OA of maximal strength m. The reverse of this is true for $m = 1$ and 2. Any LH given as an $\mathrm{OA}(b, s, b, 1)$ can be transformed into a $(0, 1, s)$-net in base b by dividing its elements by b. This is because all elementary intervals corresponding to these nets have one component of the form $[t_j/b, (t_j + 1)/b)$ and the rest equal to $[0, 1)$. Moreover, if one uses the same permutations for randomizing the $(0, 1, s)$-net and for generating the sample from the corresponding LH, then the two samples will be identical. In other words, randomization of a $(0, 1, s)$-net is equivalent to generating a LH sample. For each $\mathrm{OA}(b^2, s, b, 2)$ a $(0, 2, s)$-net in base b can be constructed. One construction is shown by Owen (1997a). It is easy to show that the method by Tang (1993) of OA based LH sampling offers another way. To see this, take any LH sample based on an $\mathrm{OA}(b^2, s, b, 2)$. For this to be a $(0, 2, s)$-net in base b the criterion is that any elementary interval in base b of volume b^{-2} has exactly one point of the sample. Any such elementary interval has either two components of the form $[t_j/b, (t_j + 1)/b)$ with $0 \le t_j < b$ and the rest identical to $[0, 1)$, called two-dimensional elementary intervals, or one component of the form $[t_j/b^2, (t_j + 1)/b^2)$ with $0 \le t_j < b^2$ and the rest identical to $[0, 1)$, called one-dimensional elementary intervals. The criterion is satisfied for the latter type of elementary intervals since the sample derives from a LH with b^2 symbols. Two-dimensional elementary intervals satisfy the criterion due to the fact that the sample is obtained from an OA of strength 2 ensuring that each two-dimensional component contains each pair of symbols exactly once. Hence the LH sample based on an $\mathrm{OA}(b^2, s, b, 2)$ is a $(0, 2, s)$-net in base b.

From the discussion above we draw the conclusion that from an equidistribution point of view OA's of strength m are equivalent to $(0, m, s)$-nets for $m = 1$ and 2. This property does not remain true for $m \ge 3$. It is easy to see that an OA based LH based on an OA of strength 3 does not have the property implied by an elementary interval with one component of type $[t_j/b^2, (t_j + 1)/b^2)$, another component of type $[t_k/b, (t_k + 1)/b)$ and the rest equal to $[0, 1)$. By definition, a $(0, 3, s)$-net satisfies the property that each such elementary interval contains exactly one point of the net. Hence $(0, 3, s)$-nets have a better equidistribution property than OA based LH based on an OA of strength 3. This feature holds also for $m > 3$.

In this paper we employ two types of (t, m, s)-nets in base b, one developed by Niederre-

iter (1988) and the other by Niederreiter and Xing (1996). Since the nets from the former paper have an explicit equidistribution structure, here we provide some details about them. The construction of the generator matrices by Niederreiter's (1988) method is based on irreducible polynomials having coefficients from the set $\{0, 1, ..., b-1\}$ with the coefficient of the highest degree term equal to 1.[2] Each generator matrix is constructed with a different irreducible polynomial. Niederreiter shows that the (t, s)-sequence in base b constructed with this method has

$$t = \sum_{i=1}^{s} \left[\deg\left(p_i\right) - 1\right], \tag{2.12}$$

where $\deg(p)$ is the degree of the polynomial p. Since lower values of t yield (t, s)-sequences with a better equidistribution property, one should take the lowest degree irreducible polynomials. The above formula implies that the (t, s)-sequence has a better equidistribution property in subcomponents corresponding to the first few dimensions than the last few dimensions. For the $b = 2$ case we present the degrees of some irreducible polynomials in Table 1. In the table the polynomials are ordered according to their degree and their coefficients. The upper row of the table gives the position of the polynomial in this ordering while the lower row presents the degree of the polynomial. For example, the 19'th polynomial has degree 6.

[TABLE 1 ABOUT HERE]

According to this table, the first two polynomials have degree 1 and hence their contribution to t is zero. From the table it is possible to compute the t's corresponding to any (t, m, s)-net in base 2 constructed with Niederreiter's (1988) method. For extensive tables of irreducible polynomials for small prime bases we refer to Lidl and Niederreiter (1983).

Based on Table 1 we find that for $s = 3$ and $s = 4$ we have $t = 1$ and $t = 3$, respectively, while for $s \geq 9$ we have $t \geq 18$. In this latter case we cannot obtain a (t, m, s)-net for $m \leq 10$ because t is too high. Nevertheless, from Table 1 we can also obtain information about the low-dimensional equidistribution property of these nets by computing the t-values corresponding to lower-dimensional components of a net. For example, the t-value of the 2-dimensional component corresponding to dimensions 5 and 19 is $t = 7$. Proceeding in this manner, we can see that the Niederreiter nets have a good low-dimensional equidistribution property. In section 3.2 we explain in detail why the equidistribution property of lower-dimensional components is important.

[2] A polynomial with coefficients from $\{0, 1, ..., b-1\}$ is irreducible if it cannot be written as the product of two polynomials with coefficients from $\{0, 1, ..., b-1\}$, where none of the two polynomials is constant. The addition and multiplication operations of the polynomials' coefficients are done in the special way described in section 2.3. These irreducible polynomials in the set of polynomials with coefficients from $\{0, 1, ..., b-1\}$ are analogous to prime numbers in the set of integers.

In Table 2 the t-values corresponding to the (t, m, s)-nets in base 2 constructed by the Niederreiter-Xing method are presented for values m and s of interest in our simulation study from section 5. These values are not as easy to compute as those corresponding to the Niederreiter nets, but they are available on Gottlieb Pirsic's home page.[3] Apart from those for $s = 3, 4$, these t-values are lower than those corresponding to the Niederreiter nets. For example, for $m = 10$ and $s = 9$ we have $t = 6$, which is much lower than 18, the t-value corresponding to the Niederreiter net. Still, for dimensions ≥ 14 (apart from $m = 10$) the t-values corresponding to the Niederreiter-Xing nets are not useful since they are at least as large as m. Moreover, we are not aware of any property of Niederreiter-Xing nets that would imply that these nets have a good low-dimensional equidistribution property. Hence we conclude that for $m = 6, 8, 10$ the Niederreiter-Xing nets have a better equidistribution property for dimension 9, for dimension 3 the Niederreiter nets have a better equidistribution property, while for the other dimensions from Table 2 it is not possible to make a judgement.

[TABLE 2 ABOUT HERE]

Randomly shifted lattice points, as defined in (2.11), can be viewed as $(0, 1, s)$-nets in base n, provided that g_j and n are relative primes. The arguments for showing this are provided in section 2.5. However, this observation should not be taken as sufficient for comparing lattice points and versions of orthogonal arrays and (t, m, s)-nets. Besides being $(0, 1, s)$-nets in base n, lattice points have a regular grid structure that makes them well suited for integrating periodic functions. Due to this feature they are different from the other samples discussed in this paper, and therefore, the equidistribution property in the sense of elementary intervals does not serve as a tool for judging their performance for integration.

We conclude this subsection by summarizing the hierarchy of the samples presented in section 2 regarding their equidistribution property. We assume that each sample has the same size. LH samples are equivalent to (randomized) $(0, 1, s)$-nets and to some extent to Halton sequences since their one-dimensional components are parts of $(0, 1)$-sequences. Halton sequences are expected to have a better equidistribution property in lower dimensions while in higher dimensions LH samples and $(0, 1, s)$-nets tend to have a better equidistribution property for low sample sizes. OA's of strength 2 have additional equidistribution structure in two dimensions, hence OA samples and especially OA based LH samples and $(0, 2, s)$-nets have a better equidistribution property. The latter two are essentially equivalent. (t, m, s)-nets in base 2, both of the Niederreiter and the Niederreiter-Xing type, have lower t-values for lower dimensions, so they have a better equidistribution property if s is relatively low. In higher dimensions we expect $(0, 2, s)$-nets to have a better equidistribution property than these. Lattice points should not be compared to the other samples on the basis of the equidistribution property.

[3] http://www.dismat.oeaw.ac.at/pirs/niedxing.html

Existence and construction. Here we present an existence and construction result of $(0, m, s)$-nets. As we have shown previously, OA's of strength m can be constructed from $(0, m, s)$-nets. Even more, we have shown that for $m = 2$ the reverse of this is also true. Hence the existence of $\text{OA}(b^2, s, b, 2)$ is equivalent to the existence of $(0, 2, s)$-nets in base b. An existence result on $(0, m, s)$-nets is the following.

PROPOSITION 4. *If $m \geq 2$ and b is a power of a prime number, then a $(0, m, s)$-net in base b exists if and only if $s \leq b + 1$ (proved in Niederreiter, 1992).*

These $(0, m, s)$-nets in a prime power base b can be constructed by Niederreiter's (1988) construction method. There are exactly b linear irreducible polynomials $X, X + 1, ..., X + b - 1$, so these can be used to construct a $(0, m, b)$-net in base b. Then, as Niederreiter shows, if we let the vector $\left(0, \frac{1}{b^m}, \frac{2}{b^m}, ..., \frac{b^m-1}{b^m}\right)'$ be the $(b + 1)$'th column of the net, then the resulting set of points is a $(0, m, b + 1)$-net in base b. In the simulation study we use this construction method for obtaining a $(0, 2, 17)$-net in base 16 and a $(0, 2, 9)$-net in base 8.

We conclude this section with the remark that samples with a good equidistribution property are not obtained easily. The better equidistributed a sequence, the more difficult to construct it. LH samples are easy to construct and Halton sequences are just slightly more difficult to obtain. More equidistributed sequences like samples based on OA's and (t, m, s)-nets are not trivial to construct. Lattice points are relatively easy to construct.

3. Properties of the estimates

In this section we discuss some properties of the integral estimates. For this we introduce the ANOVA (analysis of variance) decomposition in the first subsection. We apply this in the following subsection to show intuitively how the discussed samples work for estimating integrals. Then, based on the Fourier series representation of a function, we provide intuitive arguments that show why lattice points work well for periodic integrands. Here we also describe the two selection criteria for lattice points mentioned in section 2.5. In the last subsection we present some finite sample and asymptotic results on the variance of the integral estimates for different samples.

3.1. ANOVA decomposition

In order to be able to explain how the sampling methods work, we present the ANOVA decomposition used by Owen (1992). ANOVA decomposes a function to a sum of mutually uncorrelated functions that depend on different subsets of the variables. Hence it makes it possible to express the variance of the function as the sum of the variances of these functions.

Formally, denote $S = \{1, 2, .., s\}$ and let $u \subseteq S$ be a subset of the coordinates of $[0, 1)^s$. Let $|u|$ and $\overline{u}$ denote the cardinality of u and its complement $\{1, 2, .., s\} \setminus u$, respectively.

For $x \in [0,1)^s$ let x^u denote the coordinates of x indexed by elements of u and write $[0,1)^u$ for the domain of x^u. For $u \subseteq [0,1)^s$, define the functions α_u recursively by

$$\begin{aligned} \alpha_u &= \int_{[0,1)^{\overline{u}}} \left(f - \sum_{v \subsetneq u} \alpha_v \right) dx^{\overline{u}}, \\ \alpha_\emptyset &= I, \end{aligned} \tag{3.1}$$

where I is defined by (2.1).

The resulting sequence of functions (α_u) satisfies a number of appealing properties of which some were mentioned at the beginning of this subsection, and that help to analyze the variance of the integral estimates. These properties are the following.

1. $f = \sum_{u \subseteq S} \alpha_u$, that is, the integrand function can be written as the sum of all α_u's, where each α_u depends on x_u.
2. $\int_0^1 \alpha_u dx^j = 0$ whenever $j \in u$ (where $x^j \equiv x^{\{j\}}$).
3. $\int_{[0,1)^s} \alpha_u \alpha_v dx = 0$ whenever $u \neq v$, which, together with property 2 implies that two different terms from the sum representing f are uncorrelated.
4. $\int_{[0,1)^s} (f - I)^2 \, dx = \sum_{|u| \geq 1} \int_{[0,1)^s} \alpha_u^2 dx$, that is, the variance of f is the sum of the variances of the terms representing f.

EXAMPLE 5. *In order to illustrate the ANOVA components, we give their expressions in the case $s = 3$. The univariate components are the following.*

$$\begin{aligned} \alpha_1 &= \int_{[0,1)^2} f\left(x^1, x^2, x^3\right) dx^2 dx^3 - I, \\ \alpha_2 &= \int_{[0,1)^2} f\left(x^1, x^2, x^3\right) dx^1 dx^3 - I, \\ \alpha_3 &= \int_{[0,1)^2} f\left(x^1, x^2, x^3\right) dx^1 dx^2 - I, \end{aligned}$$

where $\alpha_j \equiv \alpha_{\{j\}}$ for $j = 1, 2, 3$. The bivariate components are:

$$\begin{aligned} \alpha_{12} &= \int_{[0,1)} f\left(x^1, x^2, x^3\right) dx^3 - \left(I + \alpha_1 + \alpha_2\right), \\ \alpha_{13} &= \int_{[0,1)} f\left(x^1, x^2, x^3\right) dx^2 - \left(I + \alpha_1 + \alpha_3\right), \\ \alpha_{23} &= \int_{[0,1)} f\left(x^1, x^2, x^3\right) dx^1 - \left(I + \alpha_2 + \alpha_3\right). \end{aligned}$$

The three-variable component is

$$\alpha_{123} = f - \left(I + \alpha_1 + \alpha_2 + \alpha_3 + \alpha_{12} + \alpha_{13} + \alpha_{23}\right).$$

This last equality suggests why property 1 holds. Property 2 can be proved easily by induction. Property 3 is a consequence of property 2, while property 4 is an implication of the other three properties.

The ANOVA decomposition provides a fairly abstract way of dealing with the variance of a function since generally it is difficult to give analytical expressions for the components. Nevertheless, it is a useful tool as shown by Owen (1992) and as we also illustrate in the next subsection. The essential fact about the ANOVA decomposition is that a decomposition with such properties exists.

Note that property 4 has interesting implications for the integrand function f. It may happen that some components α_u dominate the variance of f in the sense that the sum of the corresponding variances amounts to a large fraction of the total variance. If it happens that these components have low dimensions, we will say that the integrand has an inherent low-dimensional structure. In the next subsection we present intuitive arguments that in high dimensions the sampling methods presented above are most suitable for functions with significant inherent low-dimensional structure.

3.2. Intuitive arguments using lattice sampling

In section 2.2 when introducing OA based LH sampling we mention that the integrals of functions that are the sum of univariate functions, or more formally, have the ANOVA decomposition

$$f = \sum_{|u|\leq 1} \alpha_u, \tag{3.2}$$

can be estimated very well with a LH sample. This is because integrating f implies integrating each univariate component α_u, i.e.,

$$\int_{[0,1)^s} f dx = \alpha_\emptyset + \sum_{|u|=1} \int_0^1 \alpha_u dx^u$$

so the domain of f is decomposed into its one-dimensional components by the integral. Likewise, the LH sample is also decomposed into its one-dimensional components for estimating the integral since the estimate

$$\widehat{I} = \frac{1}{n}\sum_{i=1}^{n} f\left(x_i\right) = I + \sum_{|u|=1} \frac{1}{n}\sum_{i=1}^{n} \alpha_u\left(x_i^u\right),$$

uses exactly these one-dimensional components x_i^u for $i = 1, ..., n$, where (x_i) is a LH sample of size n. We can see from here that a LH sample does very well in these cases since its one-dimensional components have a very good equidistribution property. In fact one can say that the equidistribution property of the one-dimensional components of a LH sample

are optimal (Niederreiter, 1992, p.23) and hence LH samples are optimal for this type of f. We also note that in higher dimensions samples with a better equidistribution property than LH samples would not do any better. Hence a richer combinatorial structure induced by, for example, an OA based LH sample would not be exploited for this type of function.

Owen (1994) generalizes the intuition for LH samples to OA samples using lattice sampling introduced by Patterson (1954). Lattice samples (although related to lattice points, they should not be confused with them) are similarly generated as OA samples but the uniform random variables are replaced by 1/2. For illustration we present the strength 2 case but the idea applies to OA's of any strength. Let a_{ij} be the elements of an OA$(b^2, s, b, 2)$. Then a lattice sample based on this is defined by

$$x_{ij} = \frac{\pi_j(a_{ij}) + 1/2}{b},\ i = 1, ..., n,\ j = 1, ..., s, \tag{3.3}$$

where π_j is as in (2.3). This way each x_{ij} belongs to the set $B = \{1/2b, 3/2b, ..., (2b-1)/2b\}$, each element of B appears exactly b times in each column of the matrix (x_{ij}) and each pair of elements appears exactly once in any two columns of this matrix. This regularity of the elements make it possible to write any two-dimensional integral as a combination of one-dimensional integrals, as we show next.

Suppose that f can be written as a sum of at most bivariate functions. That is, f has an ANOVA decomposition

$$f = \sum_{|u|\leq 2} \alpha_u = \alpha_\emptyset + \sum_{|u|=1} \alpha_u + \sum_{|u|=2} \alpha_u. \tag{3.4}$$

Hence a sample used for estimating the integral of f in fact estimates univariate and bivariate functions α_u. Take a generic bivariate function $\alpha(y, z)$. Then its integral is estimated by $\widehat{I}_\alpha = 1/n \sum_{i=1}^n \alpha(y_i, z_i)$ where $\{(y_i, z_i)\}_{i=1}^n$ is a sample of size n. If we use a lattice sample as given in (3.3) for estimating the integral of f then $n = b^2$ and $\{(y_i, z_i)\}$ corresponds to two columns of the sample and hence any pair of elements of B is taken exactly once by (y_i, z_i). Therefore

$$\widehat{I}_\alpha = \frac{1}{b^2} \sum_{i=1}^b \sum_{j=1}^b \alpha\left(\frac{2i-1}{2b}, \frac{2j-1}{2b}\right) = \frac{1}{b} \sum_{i=1}^b \left[\frac{1}{b} \sum_{j=1}^b \alpha\left(\frac{2i-1}{2b}, \frac{2j-1}{2b}\right)\right].$$

The expression from the brackets is an estimate of the integral $I_\alpha(y) = \int_0^1 \alpha(y, z)\, dz$ for $y = \frac{2i-1}{2b}$, denoted $\widehat{I}_\alpha(y)$, using a lattice sample of size b. Similarly, the expression of $\widehat{I}_\alpha$ can be interpreted as an estimate of the integral $\int_0^1 \widehat{I}_\alpha(y)\, dy$ using again a lattice sample of size b. So estimation of a bivariate function $\alpha(y, z)$ with a lattice sample of size b^2 implies in fact estimation of univariate functions with lattice samples of size b. Since lattice samples in one dimension are deterministic analogs of LH samples, they are also optimal for univariate functions, in the sense of their equidistribution property. Hence we expect lattice samples obtained from OA's of strength 2 to work well for bivariate functions.

For f defined in (3.4) one needs to estimate also the univariate components with the same lattice sample. Since an OA$(b^2, s, b, 2)$ is not a LH, the lattice sample based on this is not optimal for the univariate components. Tang's (1993) procedure of OA based LH sampling avoids this problem but with this sampling it is difficult to illustrate intuitively how estimation of integrals using samples based on OA's of strength 2 works.

From the above discussion and the similarity of OA samples and lattice samples we draw the conclusion that if a function has the ANOVA representation (3.4) then estimation methods with samples based on OA's of strength 2 (OA samples, OA based LH samples and $(0, 2, s)$-nets) provide a very good way of estimating its integral. This idea generalizes for functions with ANOVA decomposition

$$f = \sum_{|u| \leq m} \alpha_u$$

and samples obtained from OA's with strength m. Similarly as for the $m = 2$ case, it is possible to reduce the integral of an m-variate function to integrals of univariate functions. However, we note that the size of the sample b^m should grow very large as m increases in order to obtain reasonably precise estimates for the one-dimensional integrals estimated with samples of size b.

Now we can argue that the inherent low-dimensional structure of the integrand is that matters for high-dimensional integrals. We extend here the remarks from the previous paragraphs along this idea. So if the integrand has an inherent structure such that ANOVA-components up to m dimensions are dominating and if for estimating the integral we use samples based on OA's of strength m (OA samples, OA based LH samples and $(0, m, s)$-nets) we expect to obtain good estimates of the integral. The sample works well for the components of the ANOVA decomposition that are at most m-dimensional and the rest of the components do not play an important role in evaluating the precision of the integral estimate. This is exactly the idea why these sampling methods may be useful for high-dimensional integrals since they may have substantial inherent low-dimensional structure. If the integrands do not have low-dimensional structure, then the improvement over MC will be marginal. If the integral is not high-dimensional, the above sampling methods are expected to work well generally since the variance of low-dimensional components is likely to represent a substantial part of the total MC variance.

3.3. Fourier series and lattice points

Though the inherent low-dimensional structure of the integrand also facilitates the performance of lattice points in integrating functions, the Fourier series representation of the integrand provides another intuitive argument. Lattice points are especially useful for estimating the integrals of periodic functions, that is, functions f defined on the whole Euclidean space satisfying $f(x) = f(x + z)$ for all integer vectors z and real vectors x. If

the integrand function f defined on the unit hypercube allows for a continuous extension that is periodic, then the Monte Carlo sum defined in (2.2) using the lattice points (2.10) computes exactly most of the Fourier series terms of the function f. This fact is stated in the following result (for a proof see Sloan and Joe, 1994).

PROPOSITION 6. *If f has the absolutely convergent Fourier series*

$$f(x) = \sum_{h \in \mathbb{Z}^s} \widehat{f}(h)\, e^{2\pi i h' x},$$

where

$$\widehat{f}(h) = \int_{[0,1)^s} e^{-2\pi i h' x} f(x)\, dx, \quad h \in \mathbb{Z}^s,$$

then the integration error is

$$\widehat{I} - I = \sum \widehat{f}(h), \tag{3.5}$$

where the summation is over $h \in \mathbb{Z}^s \setminus \{0\}$ for which $h'x \in \mathbb{Z}$ for all lattice points x from (2.10).

Both integration error criteria mentioned in section 2.5 are based on this error expression. The worst-case error criterion P_α considers a class of functions whose mixed partial derivatives of order less than α, for some $\alpha > 1$, are of bounded variation in the sense of Hardy and Krause (e.g., Davis and Rabinowitz, 1984, p.352), and from this class it takes a function that gives the highest integration error in absolute value $P_\alpha = |\widehat{I} - I|$. P_α obviously depends on the lattice points, and through this on g and n defined in section 2.5, so we can put $P_\alpha(g, n)$. Then for a given n, P_α should be minimized with respect to g. Presenting the formula of P_α is beyond the scope of this paper; it can be found in Sloan and Joe (1994).

The other integration error criterion is the spectral test, which is commonly used for assessing uniform random generators with a lattice structure. For n and g defined in section 2.5 the spectral test is defined as

$$\sigma(g, n) = \min_h \|h\|,$$

where $\|\cdot\|$ is the Euclidean distance and the minimum is taken over $h \in \mathbb{Z}^s \setminus \{0\}$ for which $h'x \in \mathbb{Z}$ for all lattice points x from (2.10). Then for a given n, g is determined by maximizing $\sigma(g, n)$. The intuition behind the spectral test is based on the integration error formula (3.5). If f has the absolutely convergent Fourier series, that is, $\sum_{h \in \mathbb{Z}^s} |\widehat{f}(h)| < \infty$, then $|\widehat{f}(h)|$ converges to zero quickly as h gets farther away from the origin. So it is likely that the largest values of $|\widehat{f}(h)|$ occur for those h's that are closest to the origin. In other words, the largest integration errors are likely to occur for those g's for which the distance from the origin of those h's closest to the origin itself is the smallest. This means that

$\sigma(g,n)$ is minimal. On the opposite, those g's, for which $\sigma(g,n)$ is maximal, will yield the lowest integration error.

We have given some explanation above why lattice points are suited for integrating periodic functions or functions that have a continuous periodic extension. Nevertheless, most integrand functions, like those arising from multivariate normal probabilities, do not have a continuous periodic extension. There are, however, ways of transforming the integrand so that it satisfies this property. For this consider the function ψ defined on $[0,1)^s$ by $\psi(t) = (\varphi(t_1), ..., \varphi(t_s))$, where φ maps $[0,1)$ into $[0,1)$, it is sufficiently smooth and $\varphi(0) = \varphi(1)$. In this case it is easy to see that $f(\psi(\cdot))$ is defined on $[0,1)^s$ and it has a continuous periodic extension. By using the transformation of variables $x = \psi(t)$ the integral of f becomes

$$I = \int_{[0,1)^s} f(x)\, dx = \int_{[0,1)^s} f(\psi(t)) \left| \frac{\partial \psi}{\partial t} \right| dt,$$

where $|\partial\psi/\partial t|$ denotes the Jacobian of the variable transformation. Let now φ be the so-called baker's transformation, that is, $\varphi(z) = 1 - |2z - 1|$. Then the Jacobian, apart from a negligible set of points, is equal to 1. Hence for the integral we obtain

$$I = \int_{[0,1)^s} f(x)\, dx = \int_{[0,1)^s} f(\psi(t))\, dt.$$

The integrand expression on the right hand side can be estimated by

$$\widehat{I} = \frac{1}{n} \sum_{i=1}^{n} f(\psi(t_i)), \tag{3.6}$$

and if $(t_i)_{i=1}^n$ are lattice points then the estimator is expected to be efficient because $f(\psi(\cdot))$ has a continuous periodic extension. The above integral estimator can also be seen in a different way, namely, as the estimator of the integral of f using the points $(\psi(t_i))_{i=1}^n$. These points are lattice points transformed by the baker's transformation. Such points have been used successfully in the numerical mathematics literature for estimating integrals (e.g., Genz, 2001). Some theoretical justification for the effectiveness of these points has been given more recently by Hickernell (2001).

3.4. Results on the variances

This subsection reviews theoretical results on the variance of the estimates for the samples presented above. As it is apparent from the results below, rather little is known about the variances. However, some of the results have interesting connections with the intuitive reasoning above. Besides upper bounds on the convergence rate of MC variance, there are also results that compare finite sample variances of MC and different samples. No result is, however, available that states what the variance reduction is as compared to MC. Such

a result should take into account the inherent low-dimensional structure of the integrand function. This is, however, difficult to treat analytically.

We recall the well-known fact that the MC estimator given in (2.2) of the integral (2.1) is unbiased and has the variance $V\left(\widehat{I}\right) = \sigma^2/n$, where σ^2 is the variance of the integrand, i.e.,

$$\sigma^2 = \int_{[0,1)^s} \left(f\left(x\right) - I\right)^2 dx. \tag{3.7}$$

Hence, providing that it exists, the variance converges to 0 at the rate n^{-1}, so we have $V\left(\widehat{I}\right) = O\left(n^{-1}\right)$. According to the results below all the samples used in this paper have a better rate of convergence.

PROPOSITION 7. *If f is of bounded variation in the sense of Hardy and Krause and x_i are the first n points of a randomized Halton sequence then*

$$V\left(\widehat{I}\right) = O\left(n^{-2}\left(\log n\right)^{2s}\right).$$

This result is due to Wang and Hickernell (2000). It implies that, since $n^{-2}\left(\log n\right)^{2s}$ converges faster to 0 than n^{-1}, for sufficiently large samples, estimates using Halton sequences are more precise than MC estimates.

The same result remains true if x_i are the first n points of a randomized (t, s)-sequence. This was shown by Owen (1998). He also shows that if one makes use of more structure offered by nets then the convergence rate of the variance will be lower.

PROPOSITION 8. *If f is of bounded variation in the sense of Hardy and Krause, and x_i are the points of a randomized (t, m, s)-net in base b with $n = b^m$ then*

$$V\left(\widehat{I}\right) = O\left(n^{-2}\left(\log n\right)^{2(s-1)}\right).$$

A result similar to Proposition 7 is available for randomly shifted lattice points. Tuffin (1997) showed that for the same type of integrands the convergence rate of the variance is $n^{-2}\left(\log n\right)^{2s}$. Unfortunately, analogous results for the case when f has the additional property of admitting a continuous periodic extension are only available for integrands with further smoothness properties. The following result is due to Tuffin (1998).

PROPOSITION 9. *If f is from the class of functions with a continuous periodic extension whose mixed partial derivatives of order less than α, for some $\alpha > 1$, are of bounded variation in the sense of Hardy and Krause, and x_i are randomly shifted carefully selected lattice points then*

$$V\left(\widehat{I}\right) = O\left(n^{-2\alpha}\left(\log n\right)^{2\alpha(s-1)+1}\left(\log\log n\right)^{(2\alpha-1)(s-1)}\right).$$

This result implies that the smoother the integrand the more precisely it is integrated by lattice points. If $\alpha = 2$ then the convergence rate of the variance is $n^{-4} (\log n)^{4(s-1)+1} (\log \log n)^{3(s-1)}$. We can compare this with a result obtained by Owen (1997b) for randomized $(0, m, s)$-nets in base b. According to this, if f is smooth (Owen's definition of smoothness is stronger than that from the proposition above for $\alpha = 2$) then the convergence rate of the variance is $O\left(n^{-3} (\log n)^{(s-1)}\right)$. This convergence is slower than the one obtained for lattice points for $\alpha = 2$. Hence for integrands that have a continuous periodic extension and additional smoothness, for sufficiently large samples, estimates using randomly shifted lattice points are more precise than those using randomized $(0, m, s)$-nets.

These results show upper bounds on the convergence rate of the variance. They suggest that in lower dimensions the convergence faster. Intuition from the previous subsection supports this. The results are not able, however, to show how the convergence rate depends on t for (t, m, s)-nets.

Tang (1993) and Owen (1994) proved results that compare the asymptotic variance of OA based LH and OA sampling with MC. The essence of both results is that if x_i are points of an OA based LH sample based on OA(b^m, s, b, m) (Tang, 1993) or of an OA sample based on OA(b^m, s, b, m) (Owen, 1994) with $n = b^m$ then

$$V\left(\widehat{I}\right) \simeq \frac{1}{n} \sum_{|u|>m} \mathrm{Var}\left(\alpha_u\right) \tag{3.8}$$

holds with increasing accuracy as $n \to \infty$, where α_u is from the ANOVA decomposition (3.1). Tang proves the result only for $m = 2$ and mentions that it should remain true for higher m's. Note that the variance from (3.8) is lower than the MC variance, which due to Property 4 from section 3.1 is

$$\frac{\sigma^2}{n} = \frac{1}{n} \sum_{u \in S} \mathrm{Var}\left(\alpha_u\right).$$

Comparing directly the two variances we notice that they have in common variances of the terms with $|u| > m$ but in (3.8) the variances of the terms with $|u| \leq m$ are missing. In other words, OA based LH and OA sampling eliminate the terms with $|u| \leq m$ from the variance. This is in line with the intuition from the previous subsection.

There are also some results available on finite sample variance. Owen (1997a) proved the following

PROPOSITION 10. *If f is square-integrable on $[0,1)^s$ (i.e., f^2 is integrable on $[0,1)^s$) and x_i, $i = 1, ..., n = b^m$, are the points of a randomized $(0, m, s)$-net in base $b \geq \max(s, 2)$ then*

$$V\left(\widehat{I}\right) \leq \frac{\sigma^2}{n} \left(\frac{b}{b-1}\right)^{\min(s-1,m)}.$$

Since a LH sample is a randomized $(0,1,s)$-net, for this $V\left(\widehat{I}\right) \leq \sigma^2/(n-1)$ holds. So the variance when a LH sample is used is less than a slightly higher number than the MC variance. For randomized $(0,2,s)$-nets the same property remains true but the factor by which the MC variance is multiplied, $(b/(b-1))^2$, is higher. Still for a sample size $n = 1024$ this factor is 1.066, which is rather close to 1. Consequently, the variance of the estimate from using LH samples or randomized $(0,2,s)$-nets is essentially smaller than the MC variance. The same property is also expected to hold for OA based LH samples due to their similarity to randomized $(0,2,s)$-nets.

4. Multivariate normal probabilities

We apply the samples presented above to integrals obtained as multivariate normal probabilities. This type of probabilities arise mainly in multinomial probit models as probabilities of choices made from a number of alternatives where the random variables involved are normally distributed (see, e.g., Börsch-Supan and Hajivassiliou, 1993). Such a probability has the form $p = \Pr(x < x_0)$ with x a r-vector having the distribution $N(\mu, \Sigma)$ and x_0 a r-vector of real numbers. Denoting by T the lower triangular square root matrix of Σ and putting $v \equiv x_0 - \mu$ we obtain the formula

$$p = \Pr(Te \leq v) \text{ with } e \sim N(0, I_r). \tag{4.1}$$

We show below that, if we use the appropriate integrand function, this probability can be expressed as an integral of the type (2.1). For this, we need to write out separately the r inequalities from the probability and use a transformation of variables from truncated normal to uniform distribution. Previous work (e.g., Börsch-Supan and Hajivassiliou, 1993, Vijverberg, 1997) employed the same idea for sampling from the truncated normal distribution. While from an estimation point of view the two procedures are equivalent, we prefer presenting the problem as the transformation to an integral of type (2.1), since it shows exactly the nature of the problem.

Let

$$\begin{aligned} T &= \begin{bmatrix} t_{11} & 0 & \dots & 0 \\ t_{21} & t_{22} & \dots & 0 \\ \vdots & \vdots & & \vdots \\ t_{r1} & t_{r2} & \dots & t_{rr} \end{bmatrix}, \text{ and} \\ v &= (v_1, ..., v_r)'. \end{aligned}$$

Then if the domain of events from the probability is denoted

$$D = \left\{ (e_1, ..., e_r) \in \mathbb{R}^r : e_1 \leq \frac{v1}{t_{11}}, e_2 \leq \frac{v_2 - t_{21}e_1}{t_{22}}, ..., e_r \leq \frac{v_r - t_{r1}e_1 - ... - t_{rr-1}e_{r-1}}{t_{rr}} \right\},$$

then

$$p = \int_D \phi(e_1) ... \phi(e_r) \, \mathrm{d}e_1 ... \mathrm{d}e_r \tag{4.2}$$

where ϕ is the standard normal density function. Using the fact that the e's have truncated normal distribution, we can obtain a computationally more tractable version of the probability. Namely, if $e \leq b$ and $e \sim N(0,1)$ then we can sample the e's if we sample from u uniform on $[0,1]$ and let $e = \Phi^{-1}(u \cdot \Phi(b))$, where Φ is the standard normal distribution function.

The integral from (4.2) can be transformed using this idea for the r-dimensional case. For this we employ the following transformation

$$\begin{aligned} e_1 &= \Phi^{-1}\left(u_1 \Phi\left(\frac{v_1}{t_{11}}\right)\right) \\ e_2 &= \Phi^{-1}\left(u_2 \Phi\left(\frac{v_2 - t_{21}e_1}{t_{22}}\right)\right) \\ &\vdots \\ e_r &= \Phi^{-1}\left(u_r \Phi\left(\frac{v_r - t_{r1}e_1 - ... - t_{rr-1}e_{r-1}}{t_{rr}}\right)\right) \end{aligned} \tag{4.3}$$

where $u_1, ..., u_r \in [0,1]$ and e_j are viewed as a functions of $u_1, ..., u_j$ for $j = 1, ..., r-1$. So the probability formula becomes

$$\begin{aligned} p &= \Phi\left(\frac{v_1}{t_{11}}\right) \int_{[0,1]^{r-1}} \Phi\left(\frac{v_2 - t_{21}e_1}{t_{22}}\right) \cdots \\ &\quad \cdots \Phi\left(\frac{v_r - t_{r1}e_1 - ... - t_{rr-1}e_{r-1}}{t_{rr}}\right) \mathrm{d}u_1 ... \mathrm{d}u_{r-1}. \end{aligned} \tag{4.4}$$

The integral in this formula is of the form (2.1) with $s = r - 1$. Hence it can be estimated with Monte Carlo simulation by drawing a sample from the uniform random vector on $[0,1)^{r-1}$. The estimator obtained this way is called the GHK simulator or RIS (i.e., recursive importance sampling) simulator based on truncated normal density (Vijverberg, 1997). Vijverberg also discusses other RIS simulators. A number of other simulators are presented in Hajivassiliou et al. (1996).

5. Simulation study

We have seen in section 3 that no exact results are available on the performance of the samples. In order to analyze their performance, in a simulation study we apply the samples to integrals arising from multivariate normal probabilities with different parameters and of different dimensions. We opted for this class of integrals due to their popularity in the econometric literature of recent years (e.g., Börsch-Supan and Hajivassiliou, 1993, Hajivassiliou et al., 1996, Vijverberg, 1997).

The approximation of integrals arising from multivariate normal probabilities has been studied extensively also in the numerical mathematics literature (e.g., Genz, 1992, 1993). These integrals are often used for testing the performance of new quasi-Monte Carlo samples (e.g., Hong and Hickernell, 2001). These studies, nevertheless, are different from our study since they follow other objectives. For example, they analyze the convergence rate of the integral estimate, and therefore, use larger sample sizes and smaller dimensional integrand functions.

In the simulation study we follow three objectives. The first is to find the improvement from using the different sampling methods over MC, the second is to compare the different methods with each other and determine the best in different cases, and the third is to analyze if the results are in line with the intuition presented above. In the first subsection we describe the simulation design. The second subsection treats some practical issues regarding the samples used. In the third subsection we present the simulation results.

5.1. Simulation design

The simulation design we use is based on that of Vijverberg (1997). First we provide details on the parameters of the probability $p = \Pr(Te \leq v)$ defined in (4.1). The covariance matrix $\Sigma = (\rho_{ij})$, where $\Sigma = TT'$, is of two types, namely, AR(1) defined as $\rho_{ij} = \rho^{|i-j|}$, and one-factor given by $\rho_{ij} = \rho$ if $i \neq j$ and $\rho_{ij} = 1$ for $i = j$. We refer to these covariance matrices as AR and F, respectively. From these matrices we can obtain new covariance matrices by multiplying the last $r/2$ rows and columns by -1 that we refer to as type AR1 and F1. Similarly, by multiplying the even-numbered rows and columns of the AR and F type matrices by -1, we obtain other covariance matrices, referred to as AR2 and F2 type. The parameter ρ takes the values -0.3, -0.2, -0.1, -0.05, 0.1, 0.3, 0.5, 0.7 and 0.9 for the F, F1 and F2 type of covariance matrices, except in the cases when the covariance matrix is not positive definite. For the AR, AR1 and AR2 type of matrices we consider only the five positive values above for ρ since an AR type matrix for a given ρ is identical to the corresponding AR2 matrix for -ρ.

The deterministic vector v, similarly to Vijverberg (1997), takes five different values: $(0, ..., 0)'$, $(1, ..., 1)'$, -$(1, ..., 1)'$, $(0, 2, 0, 2, ...)'$ and -$(0, 2, 0, 2, ...)'$. The dimension of the problem, r, depends on the sample size. We use three different sample sizes: $1024 = 2^{10}$, $256 = 2^8$, $64 = 2^6$. The sample size 1024 is useful for comparing the performance of the samples presented in the paper with previous results from Vijverberg, who considered samples of size 1000. The sample sizes 256 and 64 are useful for practical reasons. For the sample size 1024, in addition to the four dimension values 4, 10, 20 and 50 considered by Vijverberg, takes also 30. We did so because, on the one hand, the OA we consider, $\mathrm{OA}(1024, s, 32, 2)$, exists only for dimensions up to 33, and, on the other hand, generator matrices for the Niederreiter-Xing sequence are available for dimensions between 4 and 32.

For the sample sizes 256 and 64 we considered dimension values 5, 10, 18 and 5, 10, 15, respectively.

When the sample size is 1024, the combination of all these parameters leads to a number of 855 different cases, while for the sample sizes 256 and 64, we obtain 513 different cases for each. For each case and each sampling method we estimated the corresponding probability given in (4.1) 100 times and computed the standard deviation of the log of the probability estimate. For different cases the estimates were computed using the same seed of random numbers. In addition to the samples described in the next subsection, we also computed the standard deviations corresponding to MC and antithetic MC. This latter method was shown by Vijverberg (1997) to work well for the type of integral we deal with. We generated antithetic MC samples of size n by generating a random uniform sample of size $n/2$ with elements x_{ij} and adding a new sample with the elements $1 - x_{ij}$ to it.

The main performance measure of a given sampling method in our analysis is the ratio of standard deviations of the probabilities obtained from the MC and the given method. Since in the case when the sample size is 1024, the probability ratios seem to be less stable for probabilities very close to zero, we used the ratio of the corresponding log-probabilities. The ratio turned out to be similar in the two cases. The similarity of the two and the fact that the probability estimates are unbiased suggests that the bias of the log-probability estimates is not significantly large.

5.2. The samples

We construct the samples in the way described in section 2. Here we discuss practical issues related to the constructions. For the randomized Halton sequence we need to truncate the prime base representation of the starting uniform variables. We did so by taking the number of digits after the decimal point in base b representation equal to $\lfloor 15 \cdot \ln(10) / \ln(b) \rfloor + 1$. This number is roughly equal to the number of digits in base b of a 15-digit integer in base 10. This way the starting uniform variables in a base less than 10 are truncated at more than 15 decimals while those in a base greater than 10 are truncated at less than 15 decimals. This procedure keeps the number of digits in different bases balanced with respect to the 15 digits in base 10.

Construction of LH, OA and OA based LH samples is straightforward according to the description from section 2. For the sample size 1024 the OA employed is an OA$(1024, 33, 32, 2)$.[4] We use this OA to construct $(0, 2, s)$-nets in base 32. For the sample sizes 256 and 64 we construct $(0, 2, s)$-nets in base 16 and base 8 by the Niederreiter method described in section 2.6. We also use Niederreiter nets in base 2. For the construction of both types of Niederreiter nets we use the algorithm from Bratley et al. (1992). For constructing the

[4]Some OA's and software for constructing them are available from the web site of the Statlib in the Designs section (http://lib.stat.cmu.edu/designs/). The OA we use in this paper is taken from there.

Niederreiter-Xing nets in base 2 we used the generator matrices from Gottlieb Pirsic's home page (see footnote 3).

We constructed two types of lattice points, both of Korobov type. For the first type we minimized $P_2(g,n)$. We obtained the parameters $q = 323, 141, 141, 275, 19$ for samples of size 1024 corresponding to dimensions 4, 10, 20, 30, 50, $q = 39, 45, 67$ for samples of size 256 corresponding to dimensions 5, 10, 18, and $q = 3, 11, 19$ for samples of size 64 corresponding to dimensions 5, 10, 15. For the second type of lattice points we chose the parameter value $q = 1571$ from those given by Hickernell et al. (2000). We use this value for all sample sizes and dimension values. These authors construct so-called extensible sequences of lattice points having the property that certain subsets of them are also lattice points, analogously to (t,s)-sequences and (t,m,s)-nets. This way a given parameter can be used for several sample sizes. Hickernell et al. use various criteria for selecting the lattice points; this parameter is based on the spectral test. The parameter value $q = 1571$ is recommended for samples of size 2^m with $m = 6, 7, ..., 12$ and dimensions up to 33.

Gauss-codes for generating the samples are available from the first author.

5.3. Results

The results from the simulation are presented in Tables 3, 4 and 5. The sampling methods are specified in the top rows of the tables. The abbreviations need clarification. AMC stands for antithetic MC sampling, Halt denotes the Halton sequence, OALH stands for OA based LH sampling, (0,2,s) means $(0,2,s)$-net in either base 32 (sample of size 1024), or base 16 (sample of size 256), or base 8 (sample of size 64), Nied and NX stand for Niederreiter and Niederreiter-Xing sequences, respectively. Further, TKor denotes the lattice points of Korobov type based on minimizing $P_2(g,n)$ and transformed by a simpler version of the baker's transformation $\varphi(z) = |2z-1|$. Ext stands for the extensible lattice points (of Korobov type) based on the parameter from Hickernell et al. (2000), and TExt stands for the same lattice points transformed by the above mentioned baker's transformation. For the sample size 1024 we considered all these samples. For the sample sizes 256 and 64 we selected the samples that perform the best for sample size 1024 and in addition we consider AMC and Halt. Due to its simplicity to construct, Halt has been employed in several applications of integral estimation (e.g., Bhat, 2001a, 2001b, Train, 2000, 2002)

Each entry of these tables is the geometric mean of ratios of standard deviations $\sigma_{\widehat{p}_{MC}}/\sigma_{\widehat{p}}$ (for sample sizes 256 and 64) and $\sigma_{\ln(\widehat{p}_{MC})}/\sigma_{\ln(\widehat{p})}$ (for sample size 1024), where $\widehat{p}_{MC}$ and $\widehat{p}$ denote the estimates of the probability corresponding to MC and a certain sample, respectively. Hence the square of each entry gives the average factor by which we have to increase the size of the MC sample in order to achieve the same precision as with the respective sample. The geometric means are computed over different ρ and v values. Some of the probabilities, especially in larger dimensions, are very small, fact that may cause computa-

tional inaccuracies. In the case of sample size 1024, in order to avoid presenting results that may not be accurate, we decided to discard those cases in which the probability estimates are less than e^{-100}. The number of cases that remain for our analysis are given in the rightmost column of Table 3 in the numerator; the denominator represents the number of feasible cases, that is, when the covariance matrix is positive definite. For the sample size 256 we discarded only two cases, which where of F1 type integrals in dimension 18, while for the sample size 64 we did not discard any case.

[TABLE 3 ABOUT HERE]
[TABLE 4 ABOUT HERE]
[TABLE 5 ABOUT HERE]

We recall that in Tables 3, 4, 5, for each dimension the average ratios are computed separately for the different type of covariance matrices. This is because certain type of covariance matrices for certain samples (e.g., F and F2 for Halt and Nied) may show a different pattern. Similarly, since all sampling methods analyzed in this paper are expected to work better in low dimensions as compared to MC (see sections 3.2 and 3.4), we treat each dimension separately.

The general impression from Tables 3, 4, 5 is that the samples presented in the paper improve substantially over MC even in dimensions as high as 50. More exactly, in Table 3 the sample with the best performance for this dimension is TKor, and this has the lowest ratio equal to 2.11 (for AR1 integrals). This implies that on average one needs more than 4 (more precisely, $2.11^2 = 4.45$) times as large MC samples in order to achieve the same precision as the integral estimate using TKor. A similar conclusion can be drawn for dimension 30 (due to 2.18 for F1 type integrals), where OALH and (0,2,s) tend to dominate the rest of the samples. For smaller dimensions, like 4 or 10, the improvement in precision is rather impressive. For example, for dimension 10, for TExt, which is slightly better than Nied (apart from F and F2 type integrals; we will treat these type of integrals below in more detail), MC samples with the same precision need to be 100 times as large (due to 9.99 for AR1 type integrals). For dimension 4 the improvement in precision is even more impressive.

The results in Tables 4 and 5 are similarly remarkable. For samples of size 256, the best performance is achieved by TExt. In order to yield the same precision, MC samples should be at least 10 times as large in dimension 18 (3.21 for F1), 62 times as large in dimension 10 (7.9 for AR1) and 270 times as large in dimension 5 (16.44 for AR1). For samples of size 64, the best performance is achieved by TExt for dimensions 5 and 10, and by TKor for dimension 15. In order to yield the same precision as the best of these, MC samples should be approximately 6 times as large in dimension 15 (2.42 for F1), at least 15 times as large in dimension 10 (3.88 for AR1) and 70 times as large in dimension 5 (8.38 for AR1).

In terms of the hierarchy of the samples, we observe that in the majority of the cases the two samples of lattice points transformed by the baker's transformation, TKor and TExt,

tend to have better performance than the other samples. In a few cases, however, (0,2,s) (and OALH) approximates their performance or even outperforms them. These are the situations when the dimensions of the integrals are close to the base in which the $(0, 2, s)$-net was generated (dimension 30 for samples of size 1024, dimension 18 for samples of size 256, dimension 10 for samples of size 64). Nied and NX perform well in small dimensions, and in addition, Nied performs well in general for integrals of type F and F2 (see the discussion below). The rest of the samples, though better than MC, have a more modest performance. Apart from the largest dimensions, AMC has the poorest performance of all the samples considered.

The results confirm the intuition that the performance of the samples is getting poorer with the increase of the dimension. The results in main lines also confirm the intuition provided by the equidistribution hierarchy of the samples from section 2.6. In this regard, except for a few cases to which we return later, Halt tends to be better than LH in small dimensions (4 and 10) and poorer in high dimensions (30 and 50). (0,2,s) (and OALH) generally dominate Halt, LH and OA. Nied and NX perform well in small dimensions and become poorer in larger dimensions. On the opposite, (0,2,s) are relatively poor in small dimensions but get better as the dimensions get closer to the base in which they were generated. The successive improvements from MC to the samples with one-dimensional equidistribution property LH and Halt, as well as the further improvement to the samples with two-dimensional equidistribution property OA, OALH and (0,2,s) suggest that integrals arising from normal probabilities have substantial inherent low-dimensional structure, in the sense discussed in section 3.2, even in large dimensions.

Below we make some specific remarks on the results from Tables 3, 4, 5. In Table 3 we can notice the similarity between the results obtained from OALH and (0,2,s). This similarity is not surprising since, as we argue in section 2.6, both are in fact $(0, 2, s)$-nets in base 32, the only difference being that they are generated differently.

Another remark concerns comparing the performance of Ext and TExt from Table 3. TExt has a much better performance for dimensions 4 and 10, and still a significantly better performance for higher dimensions. This illustrates the usefulness of the baker's transformation, and, at the same time, the fact that lattice points work significantly better for functions that have a continuous periodic extension.

In section 2.6 we compared Nied and NX with respect to their equidistribution property based on the quality parameter t. According to this, NX has a better equidistribution property in dimension 9. This implies that, if the equidistribution property is the only thing that matters for integration, then NX should have a better performance for dimension 10 (we recall that the dimension of the integral corresponding to a probability is less by one than the dimension of the probability, see the end of section 4). In Tables 3, 4, 5 it turns out that for dimension 10 in the majority of the cases Nied has a better performance than NX. So we can draw the conclusion that the equidistribution property based on the

quality parameter t does not tell the whole story about the performance of a sample for estimating integrals. Rather, what appears to be more important is the low-dimensional equidistribution property. In this regard, we note that Nied has a good equidistribution property in its low-dimensional parts (we refer to section 2.6 for a discussion on this), and we believe this explains its good performance compared to NX.

In some cases for F and F2 type integrals the results appear to be different from the other types of integrals. Halt and Nied tend to have on average a significantly better performance for these integrals than for the others. We recall from section 2.6 that these two samples have the property that their components corresponding to the first few dimensions have a better equidistribution property than the components of the last dimensions. So the results suggest that the integrand functions of type F and F2 have an inherent structure in which the components of the ANOVA decomposition that contain the first few variables dominate the total variance. For the case of dimension 20, we verified whether this is so by reversing the order of the one-dimensional components of the sequences and did the simulations using the sequences obtained this way. The results confirmed our intuition. The reversed Halt for F and F2 type integrals yielded 1.11 and 0.94 as compared to 5.05 and 4.43 in the case of the original Halt. For the reversed Nied the corresponding results are 2.13 and 2.23 as compared to 7.22 and 7.19 for the original Nied. The results for the other types of integrals remained mainly unchanged. These results mean that the reversed sequences have on average significantly poorer performance for these integrals than the original sequences. Since the reversed and original samples have globally the same equidistribution property, the only reason that the results are so different is that the order of the variables in the integrands is crucial.

A similar phenomenon is observed by Vijverberg (2000), who studies two-sided multivariate normal probabilities, that is, of the type $p = \Pr(x_1 < x < x_2)$. This author, using Monte Carlo samples, finds that if in the GHK simulator the component inequalities are reversely ordered according to the length of the interval $[x_1, x_2]$, then the precision of the probability estimates tends to be better. It would be interesting to see whether this ordering of the variables makes the integrands similar to F and F2 type integrands. If it happens so, then further improvements in precision can be obtained if we replace the Monte Carlo samples by samples from the Niederreiter sequence.

The examples above illustrate in practice the intuition from section 3.2. We draw the conclusion that the original ordering of the Halton and Niederreiter sequences is better than the reverse ordering. This, however, raises an interesting problem, namely, that for each integrand there is an optimal ordering of these sequences. Hence with an optimal ordering it may be possible to obtain even better results with the Niederreiter sequence. It is not trivial to find an optimal ordering, hence we postpone it for future research. An alternative to optimal ordering is regrouping the variables of the integrand in a convenient way. Caflish et al. (1997) show how to pack the most important variables of a function into the first few

dimensions for integrals arising from valuation of mortgage backed securities.

An important practical aspect of integral estimation is the time needed for estimating an integral by a certain sample. There are situations when we cannot load the sample into the computer program that estimates the integral, but we have to generate a sample for each integral evaluation. If the amount of time needed for generating a sample with a given size and estimating the integral is larger than doing the same with an MC sample, then we can generate a larger MC sample and obtain more precise estimates for the same amount of time. Denote by t and t_{MC} the amount of generation and estimation time for a given sample and an MC sample, respectively, both of size n. Then the size of the larger MC sample will be roughly nt/t_{MC}, and therefore, the variance of the new MC estimate will be lower by the factor t_{MC}/t. So the gain in precision by using the larger MC sample in terms of the standard deviation will be given by the factor $\sqrt{t_{MC}/t}$. For samples of size 256 and 64 we measured t_{MC} and t for each sample and computed averages of these using the same grouping as in Tables 4 and 5. (For the GHK simulator we used the Gauss code from Vassilis Hajivassiliou's home page[5] tailored to our needs.) Then we multiplied the results from these tables by the corresponding values of $\sqrt{t_{MC}/t}$. What we obtained should be fair estimates of the ratios of standard deviations when we use MC samples of size nt/t_{MC}. Table 6 contains the results for samples of size 256 (the results for samples of size 64 are similar and hence not presented).

[TABLE 6 ABOUT HERE]

We compare Table 6 to Table 4. As expected, the results for AMC are not affected. The results regarding the samples corresponding to the lattice points, TKor and TExt, suffer only minor modifications. This reflects the fact that random shifts do not take much computing time. The results for Nied and NX are deteriorated to a larger extent. We note that these samples are generated by random linear scrambling.[6] The results for (0,2,s) are affected to an even larger extent, probably due to the fact that these samples were generated by Owen's randomization. Random linear scrambling turns out to be even less effective in this case because the special operations involved (see explanation below equation (2.7)) slow down the procedure considerably. Finally, the results for Halt are rather severely deteriorated. More effective randomization techniques are available for Halton sequences (e.g., Tuffin, 1996, Wang and Hickernell, 2000), but we did not pursue this problem further since Halt has anyway a poor performance.

So far we have discussed results based on averages over ρ and v values. In order to show how the performance of the different sampling methods varies with ρ we present some

[5] http://econ.lse.ac.uk/staff/vassilis/

[6] Matoušek (1998) treats some other, more effective randomization methods, but which yield samples with less randomness. It would be interesting to study the performance of these methods with the aim of applying them when fast generation of these samples is important.

graphs. These graphs are presented in the six panels of Figure 2; the upper panels A) and B) are for dimension 10, the middle panels C) and D) for 20 and the bottom panels E) and F) for dimension 30; the left hand side panels A), C), E) are for AR type integrals, while the right hand side panels B), D), F) are for F type integrals. In all cases $v = (0, ..., 0)'$ and the sample size is 1024. In the graphs we present the ratios of the standard deviations relative to MC for the five positive values of ρ and link the obtained points. We selected five samples for this purpose.

[FIGURE 2 ABOUT HERE]

The graphs from Figure 2 show a high dependence of the results on ρ for some of the samples. For example, for AR type integrals the standard deviation ratios for (0,2,s) and TKor decrease strongly as ρ increases. We notice a similar phenomenon for Nied for F type integrals in dimensions 20 and 30, but here the standard deviation ratios increase. TKor shows significant variation for F type integrals especially in dimensions 10 and 30. We notice also that, in general, LH and NX show relatively little variation than the rest of the samples. These features allow us to draw some conclusions regarding the ANOVA decomposition of the integrands for the various ρ values. In this respect we infer that the AR type integrands have dominating 2-dimensional components for a small ρ, but their domination tends to diminish as ρ gets larger. In a similar way we conclude that the for F type integrands the first few-dimensional ANOVA components become more dominating as ρ increases. Due to the little variation of the standard deviation ratios corresponding to LH, we conclude that the variances of the one-dimensional ANOVA components of the AR and F type integrands do not vary much.

Based on the results we draw some conclusions regarding the estimation of multivariate normal probabilities. In doing so we rely on the results for samples of size 256 and 64, since these are more relevant for practical applications in econometrics. These results suggest that extensible lattice points transformed by the baker's transformation (TExt) have the best performance in the majority of cases investigated. They are easy to construct and their generation on computer is fast. Therefore, we recommend them for estimating integrals arising from multivariate normal probabilities.

6. Conclusions and discussion

We have presented several sampling methods developed recently in the statistical literature. We have provided intuition in what circumstances they have good performance for estimating integrals of functions on the unit cube. Then we have shown how the sampling methods work in practice for multivariate normal probabilities. The improvement over Monte Carlo and antithetic Monte Carlo methods is large for smaller dimensions and still significant for dimensions as high as 50.

In our study we found extensible lattice points transformed with the baker's transformation to have the best performance for normal probabilities in cases of practical interest in econometrics, that is, for sample sizes up to about 250. However, the other samples with a good equidistribution property, like the Niederreiter-Xing sequence, or the samples with a good low-dimensional equidistribution property, like the Niederreiter sequence and $(0, m, s)$-nets, can be useful in other situations of integral estimation. We could see examples of this in the case of sample size 1024. So these samples are worthwhile to analyze in future problems of integral estimation.

In the paper we have also provided examples that illustrate in practice the intuition from theory. The results provide much information about how the different methods work but, obviously, they do not show everything. Below we discuss some problems that deserve further attention.

First we discuss some questions on the use of antithetic and importance sampling in combination with the samples used in the paper. Regarding antithetic sampling first we note that an antithetic Monte Carlo estimate of the integral (2.1) is equivalent to the Monte Carlo estimate using a sample of size $n/2$ of the integral $\int_{[0,1)^s} g(x)\,dx$, where

$$g(x) = f(x_1, ..., x_s) + f(1 - x_1, ..., 1 - x_s).$$

Hence antithetic sampling combined with a sampling method is in fact the same as applying the sampling method to the above integral with a sample of size $n/2$. In a simulation study we looked at the performance of the antithetic Halton sequence. Table 7 contains the results. The upper part of the table contains the ratios of standard deviations of the Monte Carlo antithetic to the Halton antithetic estimates. The improvements are about 60-80% of the improvements without antithetic sampling from Table 3. In the lower part of Table 7 we present the absolute improvements, that is, the improvements over Monte Carlo. This way we can compare the antithetic Halton sequence to the rest of the samples. We see substantial improvement over the performance of the Halton sequence from Table 3 for dimensions 4 and 10. For higher dimensions the antithetic Halton sequence is not better than the Halton sequence. We conclude that it is possible to obtain considerable improvement by combining antithetic sampling with a sample used in the paper. The extent to which the improvement occurs is not clear, however, and needs further investigation, which we postpone for the future.

[TABLE 7 ABOUT HERE]

As mentioned in section 4, the simulator we use is the GHK simulator, which is also a RIS simulator based on the truncated normal importance sampling density. One can, obviously, use other importance sampling densities for the estimation as well, as shown by Vijverberg (1997). Since for any importance sampling density the integral can be written

in the form (2.1) for a suitable f, we can employ the sampling methods discussed. Then the main question that arises is whether we can choose the importance sampling density so that the obtained integrand has dominating low-dimensional components in the ANOVA decomposition.

Integral estimation in econometrics is typically employed within iterative optimization procedures for finding extremum estimators. For each iteration the parameters of the integrals differ and for each iteration the integrals need to be estimated. The analysis from this paper, due to the simplified probabilities from which the integrals arise, is not able to show how the different sampling methods work in these situations. We have seen in Figure 2 that the variation of one single parameter may cause huge differences in the relative performance of the different samples. Hence we think that an analysis of how the different sampling methods influence the efficiency of these extremum estimators would be very useful. Such a study could make use of the simulation design from Geweke et al. (1994), where the same problem is investigated using simulators based on Monte Carlo sampling.

The sampling methods discussed in the paper could also be applied in Bayesian models to replace Monte Carlo sampling. One technique used in these models is importance sampling for estimating integrals arising from moments of posterior distributions (Geweke, 1989). Antithetic sampling in these models, similar to the multinomial probit, has been used successfully to improve the precision of the estimates (Geweke, 1988). Shaw (1988) employed lattice points in order to make the estimators of such integrals more precise. Another technique employed in Bayesian models is Markov chain Monte Carlo for sampling from a distribution (for an overview of these methods in econometrics, see Chib and Greenberg, 1996). It appears that quasi-Monte Carlo sampling has the potential of improving Monte Carlo for these problems as well. In this regard we mention the work by Liao (1998), where it is found that carefully selected lattice points speed up the convergence in the Gibbs sampling algorithm. This result suggests that, indeed, quasi-Monte Carlo sampling can be applied with success to these types of problems as well.

The construction of quasi-Monte Carlo samples is an important research problem also currently. For integration problems in econometrics we can expect further developments from the construction of samples that have a better low-dimensional equidistribution property than the samples we discussed. Current research focuses on constructing (t, m, s)-nets in a way similar to the Niederreiter-Xing method that instead of the quality parameter t minimizes another parameter responsible for equidistribution in low-dimensional parts of the sample (see Larcher, 1998 for more details). Along the same lines, lattice points with a good low-dimensional equidistribution property can also be constructed based on a criterion introduced by Hickernell (1996). This approach is even more general, because it allows to construct lattice points that are suited for integrating functions with a certain ANOVA decomposition. For integration problems where the ANOVA decomposition is known, or can be estimated (Lemieux and Owen, 2001), lattice points constructed this way are potentially

useful.

Acknowledgments. We benefitted from discussions with Erik Meijer, Tom Wansbeek and Michel Wedel, as well as from the computing assistance by Wouter Elshof. We thank two anonymous reviewers for very useful remarks. The first version of this paper was written while the first author was a Ph.D. student at the University of Groningen, The Netherlands.

Bibliography

Bakhvalov, N.S. (1959). On approximate computation of multiple integrals, *Vestnik Moskovskogo Universiteta, Seriya Matematiki, Mehaniki,Astronomi, Fiziki, Himii* **4**: 3–18.

Bhat, C. (2001a). Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model, *Transportation Research B* **35**: 677–693.

Bhat, C. (2001b). Simulation estimation of mixed discrete choice models using randomized and scrambled halton sequences, *Transportation Research* . forthcoming.

Börsch-Supan, A. and V.A. Hajivassiliou (1993). Smooth unbiased multivariate probability simulators for maximum likelihood estimation of limited dependent variable models, *Journal of Econometrics* **58**: 347–368.

Boyle, P., M. Broadie and P. Glasserman (1997). Monte Carlo methods for security pricing, *Journal of Economic Dynamics and Control* **21**: 1267–1321.

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

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

Chib, S. and E. Greenberg (1996). Markov chain Monte Carlo simulation methods in econometrics, *Econometric Theory* **12**: 409–431.

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

Davis, P.J. and P. Rabinowitz (1984). *Methods of Numerical Integration*, Academic Press, Orlando [etc.].

Entacher, K., P. Hellekalek and P. L'Ecuyer (2000). Quasi-Monte Carlo node sets from linear congruential generators, *in* H. Niederreiter and J. Spanier (ed.), *Monte Carlo and Quasi-Monte Carlo Methods 1998*, Berlin: Springer, pp. 171–260.

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

Genz, A. (1992). Numerical computation of the multivariate normal probabilities, *Journal of Computational and Graphical Statistics* **1**: 141–150.

Genz, A. (1993). A comparison of methods for numerical computation of multivariate normal probabilities, *Computing Science and Statistics* **25**: 400–405.

Genz, A. and F. Bretz (2001). Comparison of methods for the computation of multivariate t-probabilities, Submitted for publication.

Geweke, J. (1988). Antithetic acceleration of Monte Carlo integration in Bayesian inference, *Journal of Econometrics* **38**: 73–89.

Geweke, J. (1989). Bayesian inference in econometric models using Monte Carlo integration, *Econometrica* **57**: 1317–1339.

Geweke, J. (1996). Monte Carlo simulation and numerical integration, *in* H.M. Amman, D.A. Kendrick, and J. Rust (ed.), *Handbook of Computational Economics*, Vol. 1, Elsevier Science, chapter 15, pp. 731–800.

Geweke, J., M. Keane and D. Runkle (1994). Alternative computational approaches to inference in the multinomial probit model, *Review of Economics and Statistics* **76**: 609–632.

Hajivassiliou, V., D. McFadden, and P. Ruud (1996). Simulation of multivariate normal rectangle probabilities and their derivatives: Theoretical and computational results, *Journal of Econometrics* **72**: 85–134.

Halton, J.H. (1960). On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals, *Numerische Mathematik* **2**: 84–90.

Hedayat, A.S., N.J.A. Sloane and J. Stufken (1999). *Orthogonal Arrays: Theory and Applications*, Springer, New York.

Hickernell, F.H. (2001). Obtaining $O\left(n^{-2+\epsilon}\right)$ convergence for lattice quadrature rules, Submitted for publication.

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

Hickernell, F.J., H.S. Hong, P. L'Ecuyer and C. Lemieux (2000). Extensible lattice sequences for quasi-Monte Carlo quadrature, *SIAM Journal on Scientific Computing* **22**: 1117–1138.

Hlawka, E. (1962). Zur angenäherten berechnung mehrfacher integrale, *Monatshefte für Mathematik* **66**: 140–151.

Hong, H.S. and F.H. Hickernell (2001). Implementing scrambled digital sequences, Submitted for publication.

Judd, K.L. (1998). *Numerical Methods in Economics*, MIT Press, Cambridge and London.

Korobov, N.M. (1959). Approximate evaluation of repeated integrals, *Doklady Akademii Nauk SSSR* **124**: 1207–1210.

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

Lemieux, C. and A.B. Owen (2001). Quasi-regression and the relative importance of the anova components of a function, *in* K.-T. Fang, F.J. Hickernell and H. Niederreiter (ed.), *Monte Carlo and Quasi-Monte Carlo Methods 2000*, Springer. forthcoming.

Lerman, S. and C. Manski (1981). On the use of simulated frequencies to approximate choice probabilities, *in* C. Manski and D. McFadden (ed.), *Structural Analysis of Discrete Data with Econometric Applications*, Vol. 10, MIT Press, Cambridge, MA, pp. 305–319.

Liao, J.G. (1998). Variance reduction in Gibbs sampler using quasi random numbers, *Journal of Computational and Graphical Statistics* **7**: 253–266.

Lidl, R. and H. Niederreiter (1983). *Finite Fields*, Addison-Wesley.

Matoušek, J. (1998). On the L2-discrepancy for anchored boxes, *Journal of Complexity* **14**: 527–556.

McFadden, D. (1989). A method of simulated moments for estimation of discrete response models without numerical integration, *Econometrica* **57**: 995–1026.

McKay, M.D., W.J. Conover and R.J. Beckman (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, *Technometrics* **21**: 239–245.

Niederreiter, H. (1987). Point sets and sequences with small discrepancy, *Monatshefte für Mathematik* **104**: 273–337.

Niederreiter, H. (1988). Low-discrepancy and low-dispersion sequences, *Journal of Number Theory* **30**: 51–70.

Niederreiter, H. (1992). *Random Number Generation and Quasi-Monte Carlo Methods*, SIAM, Philadelphia.

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

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

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

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

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

Owen, A.B. (1997b). Scrambled net variance for integrals of smooth functions, *Annals of Statistics* **25**: 1541–1562.

Owen, A.B. (1998). Scrambling Sobol' and Niederreiter-Xing points, *Journal of Complexity* **4**: 466–489.

Patterson, H.D. (1954). The errors of lattice sampling, *Journal of the Royal Statistical Society B* **16**: 140–149.

Revelt, D. and K. Train (1998). Mixed logit with repeated choices: Households' choices of appliance efficiency level, *Review of Economics and Statistics* **80**: 647–57.

Sándor, Z. and K. Train (2002). Quasi-random simulation of discrete choice models, Working Paper, Department of Economics, University of California, Berkeley.

Shaw, J.E.H. (1988). A quasirandom approach to integration in Bayesian statistics, *Annals of Statistics* **16**: 895–914.

Sloan, I.H. and S. Joe (1994). *Lattice methods for multiple integration*, Clarendon Press, Oxford.

Sobol', I.M. (1967). On the distribution of points in a cube and the approximate evaluation of integrals, *USSR Computational Mathematics and Mathematical Physics* **7**: 86–112.

Stern, S. (1992). A method for smoothing simulated moments of discrete probabilities in multinomial probit models, *Econometrica* **60**: 943–952.

Tan, K.S. and P.P. Boyle (2000). Applications of randomized low discrepancy sequences to the valuation of complex securities, *Journal of Economic Dynamics and Control* **24**: 1747–1782.

Tang, B. (1993). Orthogonal array-based Latin hypercubes, *Journal of the American Statistical Association* **88**: 1392–1397.

Train, K. (2000). Halton sequences for mixed logit, Working Paper No. E00-278, Department of Economics, University of California, Berkeley.

Train, K. (2002). *Discrete Choice Methods with Simulation*, Cambridge University Press, New York.

Tuffin, B. (1996). On the use of low-discrepancy sequences in Monte Carlo methods, *Monte Carlo Methods and Applications* **2**: 295–320.

Tuffin, B. (1997). Variance reduction applied to product-form multiclass queuing networks, *ACM Transactions on Modeling and Computer Simulation* **7**: 478–500.

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

Vijverberg, W.P.M. (1997). Monte Carlo evaluation of multivariate normal probabilities, *Journal of Econometrics* **76**: 281–307.

Vijverberg, W.P.M. (2000). Rectangular and wedge-shaped multivariate normal probabilities, *Economics Letters* **68**: 13–20.

Wang, X. and F.J. Hickernell (2000). Randomized Halton sequences, *Mathematical and Computer Modelling* **32**: 887–99.

TABLE 1 Degrees of irreducible polynomials

Polynomial	1	2	3	4	5	6...8	9...14	15...23	24...41	42...71
Degree	1	1	2	3	3	4	5	6	7	8

TABLE 2 t-values for Niederreiter-Xing *(t,m,s)*-nets in base 2

$m\backslash s$	3	4	9	14	17	19	29
6	2	2	5	6	6	6	6
8	2	3	6	8	9	9	9
10	3	4	6	9	10	10	11

TABLE 3 Performance in terms of standard deviation ratios of the samples of size 1024

	AMC	Halt	LH	OA	OALH	(0,2,s)	Nied	NX	TKor	Ext	TExt	
dim = 4												
AR	2.78	12.49	6.43	9.92	23.56	23.59	53.26	-	56.79	12.83	53.64	25/25
AR1	2.11	9.80	5.17	8.66	20.70	19.65	40.26	-	42.62	11.56	46.38	25/25
AR2	1.89	14.80	4.94	9.33	19.31	19.77	71.81	-	104.76	11.01	93.00	25/25
F	2.57	17.57	8.86	10.20	31.10	29.92	77.16	-	99.32	14.17	97.64	45/45
F1	2.20	12.34	7.27	10.12	26.81	25.76	55.60	-	61.87	13.42	67.01	45/45
F2	1.98	14.98	6.31	10.59	29.83	29.25	60.48	-	66.38	12.79	75.43	45/45
dim = 10												
AR	2.05	4.77	3.11	5.26	8.16	7.94	9.98	8.46	9.27	5.52	12.26	25/25
AR1	1.55	3.87	2.72	4.52	6.61	6.25	7.35	6.82	7.64	4.48	9.99	25/25
AR2	1.50	4.46	2.77	5.02	7.64	6.96	9.39	9.62	7.45	4.99	13.61	23/25
F	1.91	8.08	3.12	5.41	8.33	8.38	17.78	10.89	12.82	6.68	15.37	35/35
F1	1.63	4.98	2.61	4.58	6.04	6.21	10.27	7.89	8.44	5.08	11.82	34/35
F2	1.56	8.72	2.82	7.44	10.45	10.04	22.37	13.37	11.98	7.09	17.30	35/35
dim = 20												
AR	1.62	1.82	2.26	3.50	4.80	5.08	2.71	2.00	5.33	2.17	2.60	25/25
AR1	1.54	1.59	1.97	3.03	3.93	4.25	2.69	1.97	4.69	1.98	2.33	25/25
AR2	1.47	1.54	2.38	3.93	5.39	5.64	3.25	2.13	5.93	2.55	3.67	21/25
F	1.61	5.05	2.18	3.33	4.82	4.47	7.22	3.60	5.45	3.76	3.83	28/30
F1	1.21	1.92	1.47	2.20	2.42	2.35	2.09	1.59	2.44	1.80	2.11	28/30
F2	1.33	4.32	2.05	4.45	5.23	5.12	7.19	2.77	4.84	2.79	4.17	28/30
dim = 30												
AR	1.63	1.24	1.99	3.26	3.96	3.99	0.77	1.15	3.01	2.15	2.44	25/25
AR1	1.46	1.14	1.78	2.98	3.63	3.61	0.79	1.11	2.89	2.19	2.44	25/25
AR2	1.38	1.01	2.01	3.24	4.61	4.28	0.56	0.93	3.31	2.13	3.01	19/25
F	1.59	4.29	2.04	2.92	3.60	3.59	5.52	1.67	3.06	3.21	3.07	25/25
F1	1.28	1.50	1.37	1.83	2.18	2.06	1.52	1.10	2.13	1.52	1.84	20/25
F2	1.30	4.06	2.07	4.35	4.78	4.56	4.77	1.36	3.66	2.38	3.53	21/25
dim = 50												
AR	1.26	0.77	1.56	-	-	-	0.51	-	2.58	1.62	1.98	24/25
AR1	1.24	0.72	1.33	-	-	-	0.42	-	2.11	1.45	1.81	24/25
AR2	1.23	0.58	1.60	-	-	-	0.37	-	2.98	1.77	2.27	15/25
F	1.47	3.26	1.72	-	-	-	3.54	-	3.19	2.55	2.38	25/25
F1	1.13	1.17	1.34	-	-	-	0.66	-	2.35	1.44	1.56	16/25
F2	1.10	2.30	1.43	-	-	-	2.08	-	3.51	1.63	2.29	16/25

TABLE 4 Performance in terms of standard deviation ratios of the samples of size 256

	AMC	Halt	(0,2,s)	Nied	NX	TKor	TExt
dim = 5							
AR	2.61	7.35	12.60	18.02	18.13	12.81	19.22
AR1	2.10	5.84	11.80	17.02	18.11	13.11	16.44
AR2	1.99	6.81	10.46	18.91	20.87	13.06	26.25
F	2.41	9.73	17.18	27.73	21.61	23.49	31.58
F1	2.11	6.13	13.12	18.23	17.36	15.20	19.62
F2	2.09	8.75	17.92	25.44	24.09	16.95	28.29
dim = 10							
AR	1.82	2.89	6.19	5.29	3.72	3.90	9.79
AR1	1.79	2.87	5.46	4.69	3.87	3.25	7.90
AR2	1.74	2.81	5.50	4.22	3.72	4.30	8.97
F	1.72	4.84	6.60	9.40	3.58	4.67	11.63
F1	1.45	2.94	4.65	5.17	4.14	3.69	8.07
F2	1.63	5.52	8.55	10.12	5.57	6.59	14.13
dim = 18							
AR	1.75	1.50	4.17	1.36	1.99	4.28	3.73
AR1	1.61	1.27	3.86	1.25	1.91	4.51	4.13
AR2	1.55	1.34	4.01	1.15	2.08	4.41	3.87
F	1.59	2.94	3.70	4.40	2.30	4.16	4.80
F1	1.19	1.33	2.44	1.45	1.62	3.09	3.21
F2	1.50	3.65	5.88	4.40	3.31	6.14	7.10

TABLE 5 Performance in terms of standard deviation ratios of the samples of size 64

	AMC	Halt	(0,2,s)	Nied	NX	TKor	TExt
dim = 5							
AR	2.25	4.57	8.34	8.90	9.51	6.76	9.96
AR1	1.90	3.11	7.04	7.80	7.28	5.01	8.38
AR2	2.09	4.74	8.12	8.15	9.95	6.18	11.22
F	2.12	5.95	10.76	12.12	12.52	8.17	15.47
F1	2.09	4.07	9.39	10.79	10.23	6.20	10.42
F2	1.88	5.18	10.37	11.48	11.21	7.40	12.92
dim = 10							
AR	1.94	1.73	4.47	2.81	3.16	5.15	5.45
AR1	1.81	1.83	4.20	2.66	3.39	4.25	3.88
AR2	1.61	1.78	4.00	2.28	3.11	4.74	5.45
F	2.00	3.05	4.73	5.27	3.84	5.25	6.28
F1	1.57	1.94	3.91	3.25	3.40	4.45	5.18
F2	1.65	2.98	5.33	4.80	3.94	5.70	5.97
dim = 15							
AR	1.61	1.32	-	1.34	2.25	4.57	4.21
AR1	1.49	1.06	-	1.30	2.03	3.68	3.31
AR2	1.49	1.26	-	1.13	2.13	4.45	4.24
F	1.56	2.25	-	3.27	1.49	4.23	4.24
F1	1.22	1.31	-	1.38	1.50	2.42	2.31
F2	1.50	2.77	-	3.38	1.91	5.22	4.99

TABLE 6 Performance of the samples of size 256 corrected for computing time

	AMC	Halt	(0,2,s)	Nied	NX	TKor	TExt
dim = 5							
AR	2.61	3.93	9.67	15.20	15.86	12.68	18.81
AR1	2.09	3.22	9.30	14.55	16.02	13.04	16.10
AR2	1.99	3.88	8.43	16.45	18.80	13.16	26.09
F	2.40	5.25	13.31	23.50	19.01	23.36	30.69
F1	2.11	3.35	10.39	15.63	15.42	15.17	19.31
F2	2.08	4.83	14.14	21.68	21.32	16.80	27.54
dim = 10							
AR	1.82	1.66	4.74	4.53	3.25	3.87	9.60
AR1	1.79	1.66	4.23	4.03	3.39	3.22	7.74
AR2	1.74	1.74	4.48	3.76	3.38	4.34	8.95
F	1.72	2.79	5.08	8.07	3.14	4.64	11.41
F1	1.45	1.74	3.66	4.48	3.66	3.67	7.91
F2	1.62	3.31	6.78	8.79	4.95	6.53	13.79
dim = 18							
AR	1.75	0.90	3.19	1.17	1.74	4.24	3.65
AR1	1.60	0.76	2.98	1.08	1.67	4.47	4.05
AR2	1.55	0.86	3.25	1.03	1.88	4.43	3.84
F	1.59	1.76	2.83	3.81	2.01	4.15	4.72
F1	1.19	0.83	1.92	1.26	1.43	3.07	3.14
F2	1.50	2.30	4.67	3.86	2.94	6.08	6.95

TABLE 7 Performance in terms of standard deviation ratios of the antithetic Halton sequence

Antithetic MC vs. antithetic Halton					
	dim = 4	dim = 10	dim = 20	dim = 30	dim = 50
AR	7.09	3.49	1.25	0.81	0.55
AR1	7.17	3.37	1.15	0.84	0.51
AR2	9.31	3.61	1.07	0.62	0.42
F	9.59	4.84	3.18	2.83	2.31
F1	7.75	3.70	1.48	1.16	0.90
F2	9.64	5.92	3.14	2.72	1.65
MC vs. antithetic Halton					
AR	19.71	7.13	2.03	1.32	0.69
AR1	15.15	5.21	1.78	1.24	0.62
AR2	17.62	4.89	1.47	0.82	0.55
F	24.63	9.26	4.56	4.49	3.39
F1	17.09	5.71	1.71	1.36	1.00
F2	19.07	9.24	4.35	3.96	2.53

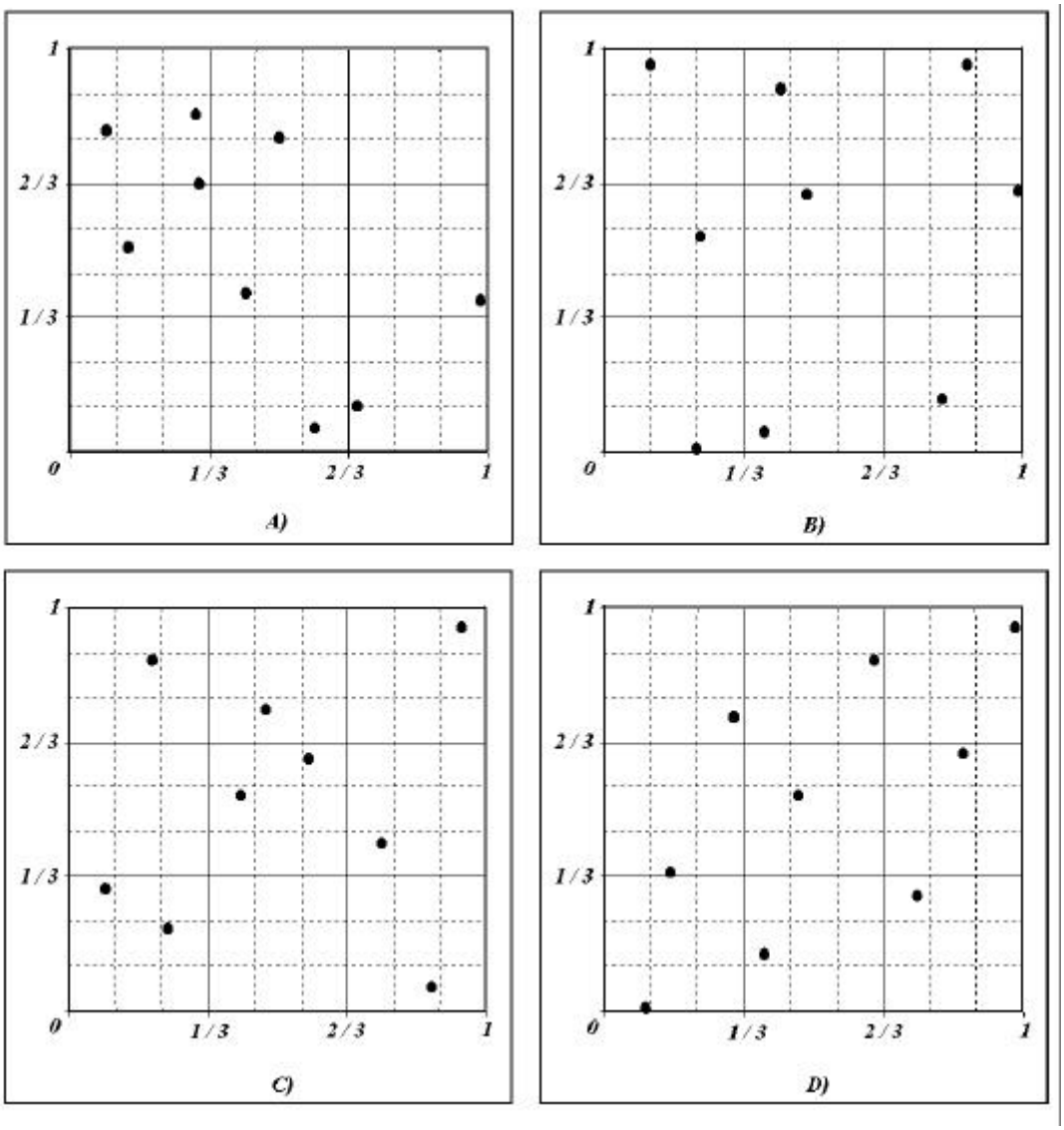

Figure 1: Samples of 9 points in the unit square. A) random, B) orthogonal array, C) Latin hypercube, D) orthogonal array based Latin hypercube.

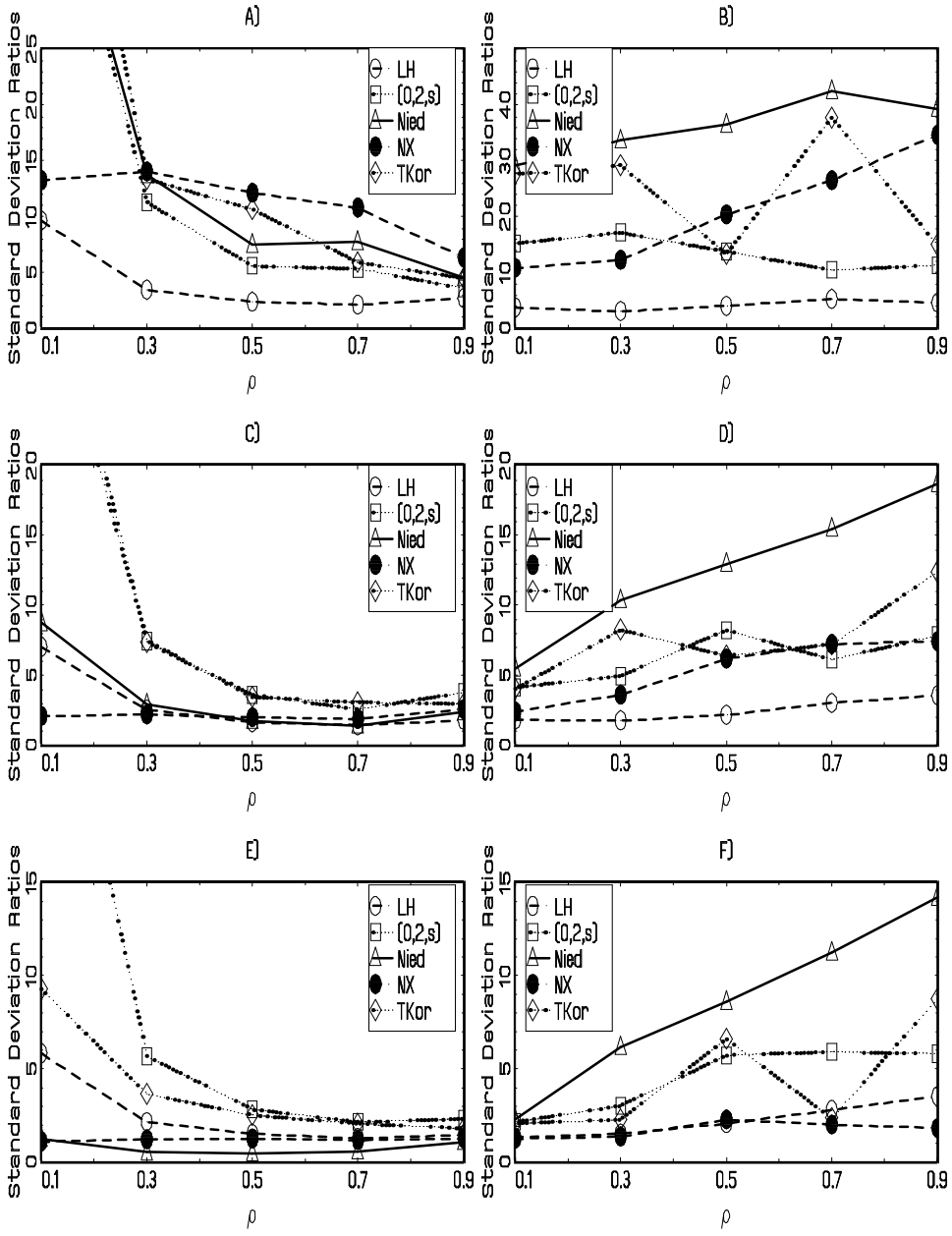

Figure 2: Ratios of MC standard deviations and standard deviations of different sampling methods as a function of ρ. A) AR type integrals, dimension 10; B) F type integrals, dimension 10; C) AR type integrals, dimension 20; D) F type integrals, dimension 20; E) AR type integrals, dimension 30; F) F type integrals, dimension 30.