Construction and Comparison of High-Dimensional Sobol' Generators

Ilya M. Sobol', Danil Asotsky

Institute for Mathematical Modelling of the Russian Academy of Sciences

Alexander Kreinin

Algorithmics LLC

Sergei Kucherenko

CPSE, Imperial College London, e-mail: s.kucherenko@imperial.ac.uk

Abstract

Sobol' sequence generators are used actively in financial applications. In this paper, we explore the effect of the uniformity Properties A and A' on the generator performance in high-dimensional problems. It is shown that these properties provide an additional guarantee of uniformity for high-dimensional problems even at a small number of sampled points. By imposing additional uniformity properties on low-dimensional projections of the sequence in addition to the uniformity properties of the d-dimensional sequence itself, the efficiency of the Sobol' sequence can be increased. The SobolSeq16384 generator, which satisfies additional uniformity properties (Property A for all 16,384 dimensions and Property A' for adjacent dimensions), is constructed. A comparison of known Sobol' sequence generators for a set of tests shows that for the majority of tests, the SobolSeq16384 generator performs better than other generators.

Keywords

effective dimensions, quasi-Monte Carlo, Sobol' sequences

1 Introduction

Financial institutions implement pricing and risk management engines for accurate and efficient valuation of their complex portfolios, having several hundred thousand positions in financial derivatives. Monte Carlo (MC) simulation is at the heart of these systems: it represents a unique universal method for pricing and risk management, working even when the dimensionality of the problem is very high.

Convergence of MC methods does not depend on the dimension of the risk factor space. However, the rate of convergence, $O(N^{-1/2})$, where N is the number of generated scenarios, is rather slow. A higher rate of convergence can be obtained by using quasi-Monte Carlo (QMC) methods based on low-discrepancy sequences (LDS). Asymptotically, QMC can provide the rate of convergence $O(N^{-1+\varepsilon})$ with an arbitrarily small $\varepsilon > 0$.

There are a few well-known and commonly used LDS such as Halton, Faure, Sobol', Niederreiter, and some others. Many practical studies have proven that the Sobol' sequences are in many aspects superior to other LDS (see, e.g., Paskov and Traub, 1995; Kreinin *et al.*, 1998; Jaeckel, 2002; L'Ecuyer and Lemieux, 2002), and this is why they are so widely used in finance. Paul Glasserman, in his highly acclaimed book *Monte Carlo Methods in Financial Engineering* (2004) says: "Preponderance of the experimental evidence amassed to date points to Sobol' sequences as the most effective quasi-Monte Carlo method for application in financial engineering."

The Sobol' LDS were constructed by following the three main requirements introduced in Sobol' (1967):

1. Best uniformity of distribution as $N \to \infty$.
2. Good distribution for fairly small initial sets.
3. A very fast computational algorithm.

The properties of the Sobol' LDS are determined by the design of direction numbers. Different criteria were used for their construction by various authors (see, e.g., Jaeckel, 2002; Lemieux *et al.*, 2004; Silva and Barbe, 2005; Joe and Kuo, 2008).

The present work contains an attempt to further increase the efficiency of the Sobol' sequence by imposing additional uniformity properties on low-dimensional projections of the sequence in addition to the uniformity properties of the d-dimensional sequence itself.

The QMC valuation efficiency of a portfolio of financial derivatives depends on the properties of the portfolio pricing function, $f(x)$, that we assume to be defined and square integrable in the d-dimensional unit hypercube H^d.[1]

It is not uncommon that the portfolio pricing function depends mainly on small groups of neighboring variables. Such models can be represented (or approximated) by equations of the following type:

$$f(\mathbf{x}) = f_0 + \sum_{i=0}^{d-m} g_i(x_{i+1}, \ldots, x_{i+m}), \tag{1.1}$$

where $m \ll d$. Obviously, the ANOVA decomposition of (1.1) contains low-dimensional terms only: their dimensions cannot exceed m. Therefore, the average dimension of (1.1) also does not exceed m. The concept of average dimension was introduced by A. Owen in Liu and Owen (2006) and independently by Asotsky *et al.* (2006). These papers contain an important suggestion: for integrands with small average dimension, QMC integrations are superior to MC integrations. Similar conclusions were drawn in Caflisch *et al.* (1997). This paper introduced the notion of effective dimensions. In more recent papers (see, e.g., Kucherenko and Shah, 2007; Wang and Sloan, 2005), it was shown that the majority of practical high-dimensional problems in finance have low effective dimensions. This explains the success of QMC methods for solving such high-dimensional problems.

Models that can be presented in the form (1.1) may benefit from using Sobol' LDS, which has additional uniformity properties on low-dimensional projections (adjacent dimensions).

This paper is organized as follows. Section 2 describes the design principles of Sobol' sequence generators. Section 3 gives definitions of Properties *A* and *A'* and shows that for dimensions higher than 3, it is possible to construct a Sobol' sequence satisfying Property *A'* but not Property *A*. In Section 4 we show that Properties *A* and *A'* provide an additional guarantee of uniformity even with a small number of sampled points. New additional uniformity properties for low-dimensional projections are introduced in Section 5. Comparison of various Sobol' sequence generators is presented in Section 6. Some examples of the application of numerical strategies based on QMC scenario generation in financial modeling are considered in Section 7. Finally, conclusions are given in Section 8.

2 Construction of Sobol' sequences

Definition 1. *Dyadic intervals*. Let m be a non-negative integer. Dyadic intervals with length 2^{-m} are obtained by dividing the unit interval $0 \le z \le 1$ into 2^m equal intervals. We assume that the left end of a dyadic interval is closed and the right end is open, with one exception: if the right end is $z = 1$ then the right end is closed also.

Thus, the unit interval is a sum of 2^m dyadic intervals with length 2^{-m}.

Definition 2. *Dyadic boxes*. A dyadic box Π in the d-dimensional unit hypercube H^d is a product of d dyadic intervals.

Thus, a given set of d non-negative integers $(m_1,..., m_d)$ defines a decomposition of H^d into a sum of 2^m dyadic boxes with volumes 2^{-m}, where $m = m_1 +... + m_d$.

Definition 3. *P_τ-nets*. Consider two integers $\nu > \tau \ge 0$. A point set consisting of $N = 2^\nu$ points in H^d is called a P_τ-net if each dyadic box with volume $2^\tau/N$ contains exactly 2^τ points of the net.

Obviously, smaller values of τ imply a more uniform distribution of the points of the P_τ-net.

Definition 4. *A dyadic section of an infinite sequence of points*. Let $x_0, x_1, x_2,...$ be an infinite sequence of points in H^d. A subset of points x_i with indices i satisfying inequalities $(k-1)\,2^p \le i < k2^p$ with arbitrary positive integers k and p is called a dyadic section of the sequence. 2^p is the length of the section. For example, sections (x_0, x_1, x_2, x_3), (x_4, x_5, x_6, x_7) are dyadic sections of length 4. But sections (x_1, x_2, x_3, x_4), (x_2, x_3, x_4, x_5) are not dyadic.

Definition 5. *LP_τ-sequences*. An infinite sequence of points $x_0, x_1, x_2,...$ in H^d is called an LP_τ-sequence if all its dyadic sections with length exceeding 2^τ are P_τ-nets.

LP_0-sequences exist only in H^1 and H^2. In higher dimensions, as d increases, the smallest possible values of τ increase also.

Definition 6. *Sobol' sequences*. LP_τ-sequences are widely used in computational mathematics. But the term 'LP_τ-sequences' is not popular. In Niederreiter (1988), P_τ-nets and LP_τ-sequences are called '(t, m, s)-nets in base 2' and '(t, s)-sequences in base 2'; here, s is the dimension and $t \equiv \tau$, $m \equiv \nu$. However, most mathematicians and practitioners prefer a more brief terminology: LP_τ-sequences are called 'Sobol' sequences'.

All Sobol' sequences are uniformly distributed in H^d. Furthermore, they are well distributed.

An efficient algorithm for generating such sequences has been introduced in Sobol' (1967). It is based on independence properties of primitive polynomials in the field GF(2). The coordinates of all these Sobol' points are binary rational numbers. For these Sobol' sequences, direction points that allow a fast generation of the sequence can be defined.

Definition 7. *Direction points*. The points $V_k = x_{2^{k-1}}$ at $k = 1, 2, 3,...$ are called direction points for the sequence $x_0, x_1, x_2,...$ Given m direction points $V_1,..., V_m$, one can easily obtain 2^m points x_i of the sequence with $i < 2^m$. In fact, if the number i in the binary system is $i = e_m...e_2e_1$, then $x_i = e_1V_1 \oplus e_2V_2 \oplus \cdots \oplus e_m V_m$ where $\oplus$ is the XOR operation applied to each coordinate. Coordinates of the direction points are called direction numbers.

The Sobol' numbers $x_n = x^1_n, x^2_n,...x^d_n$ are generated from a set of binary fractions of length b bits, $v^j = (0.v^j_2 v^j_3 ... v^j_b)_2$, $v^j_i \in 0,1, j = 1,......d$ known as direction numbers. Direction numbers are defined below. Consider the nth number in the sequence given in binary form, $n = (b_b ... b_3 b_2 b_1)$. To produce the Sobol' integer number x^j_n, the following formula is used:

$$x^j_n = b_1 v^j_1 \oplus b_2 v^j_2 \oplus \cdots \oplus b_b v^j_b,$$

where $\oplus$ is an addition modulo 2 operator: $0 \oplus 0 = 0$, $1 \oplus 1 = 0$, $0 \oplus 1 = 1$, $1 \oplus 0 = 1$. $\oplus$ can also be seen as bitwise XOR. This result is obtained by performing the bitwise exclusive XOR of the direction numbers v^j_i for which $b_i \ne 0$.

The final conversion to a uniform variate y^j_n is performed by dividing x^j_n by 2^{b_j}:

$$y^j_n = x^j_n / 2^{b_j}.$$

Direction numbers are selected by the following rules (see, e.g., Bratley and Fox, 1988).

Let $P_j = x^{s_j} + a_{1,j} x^{s_j - 1} + \cdots + a_{s_j-1,j} x + 1$, $j = 1,..., d$ be a set of different primitive polynomials. Primitive polynomials are irreducible (can't be factored) and the smallest power p for which a polynomial divides $x^p + 1$ is $p = 2^q - 1$. Let $m_{i,j} = 2^i \cdot v_{i,j}$. For numbers $m_{i,j}$ at any $j \ge 2$, $i > s_j$, where s_j is an order of a primitive polynomial P_j corresponding to dimension j, the following relationship is satisfied:

$$m_{i,j} = 2a_{1,j}m_{i-1,j} \oplus 2^2 a_{2,j} m_{i-2,j} \oplus \cdots \oplus 2^{s_j-1} a_{s_j-1,j} m_{i-s_j+1,j} \oplus 2^{s_j} m_{i-s_j,j} \oplus m_{i-s_j,j}, \tag{2.1}$$

where $a_{k,j}$ are the coefficients of a primitive polynomial P_j. Initial values $m_{1,j}$, $m_{2,j}$,..., $m_{s_j,j}$ can be chosen arbitrarily provided that conditions $m_{k,j} < 2^k$ and $m_{k,j}$ is odd are satisfied. Therefore, it is possible to construct different Sobol' sequences for the fixed dimension d.

64-79_WILM_Nov_2011_TP_Kucherenk65 65 12/23/11 10:28:42 AM

3 Properties A and A'

Sobol' (1976) introduced additional uniformity conditions known as Properties A and A'.

Definition 8. *Property A*. A low-discrepancy sequence is said to satisfy *Property A* if for any binary segment[2] of the d-dimensional sequence of length 2^d there is exactly one draw in each 2^d hypercube that results from subdividing in half the unit hypercube along each of its length extensions.

Definition 9. *Property A'*. A low-discrepancy sequence is said to satisfy *Property A'* if for any binary segment (not an arbitrary subset) of the d-dimensional sequence of length 4^d there is exactly one draw in each 4^d hypercube that results from subdividing into four equal parts the unit hypercube along each of its length extensions.

Theorem 1. The d-dimensional Sobol' sequence possesses Property A if and only if

$$det(\mathbf{V}_d) = 1(mod\ 2), \tag{3.1}$$

where V_d is the $d\times d$ binary matrix defined by

$$\mathbf{V}_d = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & v_{2,2,1} & v_{3,2,1} & \cdots & v_{d,2,1} \\ 1 & v_{2,3,1} & v_{3,3,1} & \cdots & v_{d,3,1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & v_{2,d,1} & v_{3,d,1} & \cdots & v_{d,d,1} \end{pmatrix} \tag{3.2}$$

with $v_{k,j,1}$ denoting the first digit after the binary point of the kth direction number for dimension $j (v_{k,j} = (0.v_{k,j,1}v_{k,j,2}\ldots)_2)$.

Theorem 2. The d-dimensional Sobol' sequence possesses Property A' if and only if

$$det(\mathbf{U}_d) = 1(mod\ 2), \tag{3.3}$$

where U_d is the $2d\times 2d$ binary matrix defined by

$$\mathbf{U}_d = \begin{pmatrix} v_{1,1,1} & v_{1,1,2} & v_{2,1,1} & v_{2,1,2} & \cdots & v_{d,1,1} & v_{d,1,2} \\ v_{1,2,1} & v_{1,2,2} & v_{2,2,1} & v_{2,2,2} & \cdots & v_{d,2,1} & v_{d,2,2} \\ v_{1,3,1} & v_{1,3,2} & v_{2,3,1} & v_{2,3,2} & \cdots & v_{d,3,1} & v_{d,3,2} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ v_{1,2d,1} & v_{1,2d,2} & v_{2,2d,1} & v_{2,2d,2} & \cdots & v_{d,2d,1} & v_{d,2d,2} \end{pmatrix} \tag{3.4}$$

with $v_{k,j,i}$ denoting the ith digit after the binary point of the kth direction number for dimension j $(v_{k,j} = (0.v_{k,j,1}v_{k,j,2}\ldots)_2)$.

Proof of these theorems can be found in Sobol' (1976).

Tests for Properties A and A' are independent. Thus it is possible to construct the Sobol' sequence that satisfies both Properties A and A' or only one of them (see Section 3.1).

To find determinants of matrices (3.2) and (3.4) it is necessary to calculate elements $v_{i,j,k}$ for any (including high values) i and j. At the same time it is desirable to avoid full calculation of $v_{i,j}$ numbers. Recurrent relationships for calculation of $v_{i,j,1}$ are given in Joe and Kuo (2003). We derive a generalized variant of these relationships for an arbitrary k.

From (2.1) it follows that direction numbers $v_{i,j}$ satisfy a relationship

$$v_{i,j} = a_{1,j}v_{i-1,j} \oplus a_{2,j}v_{i-2,j} \oplus \cdots \oplus a_{s_j-1,j}v_{i-s_j+1,j} \oplus v_{i-s_j,j} \oplus \frac{v_{i-s_j,j}}{2^{s_j}} \tag{3.5}$$

for all $i > s_j$

Let $v_{k,j} = 0.v_{k,j,1}v_{k,j,2}v_{k,j,3}\ldots$ be the binary representation of a number $v_{k,j}$. Then $\frac{v_{i,j}}{2^{s_j}} = 0.\underbrace{0\ldots0}_{s_j\ times}v_{i,j,1}v_{i,j,2}v_{i,j,3}\ldots$. Representation (3.5) leads to the following lemma.

Lemma 1. Let $P_j = x^{s_j} + a_{1,j}x^{s_j-1} + \ldots + a_{s_j-1}x + 1$ is a primitive polynomial of some degree s_j in the field GF(2). For $j \geq 2$ and $i > s_j$, we have

$$v_{i,j,k} = a_{1,j}v_{i-1,j,k} \oplus a_{2,j}v_{i-2,j,k} \oplus \cdots \oplus a_{s_j-1,k}v_{i-s_j+1,j,k} \oplus v_{i-s_j,j,k}, k \leq s_j, \tag{3.6}$$

$$v_{i,j,k} = a_{1,j}v_{i-1,j,k} \oplus a_{2,j}v_{i-2,j,k} \oplus \cdots \oplus a_{s_j-1,k}v_{i-s_j+1,j,k} \oplus v_{i-s_j,j,k} \oplus v_{i-s_j,j,k-s_j}, k > s_j \tag{3.7}$$

with $v_{i,j,k}$ denoting the kth digit after the binary point of the direction number $v_{i,j}$.

We will use this lemma later, in Sections 3.1 and 5.1 to investigate relations between uniformity properties without construction of the complete Sobol' sequence.

3.1 Sobol′ sequence satisfying only Property A'

The objective of this section is to establish a relationship between Properties A and A' for Sobol' sequences. It was known from early original papers by Sobol' that it is possible to construct Sobol' sequences which satisfy Property A without satisfying Property A'. However, Property A' previously was not studied in detail. Specifically, we would like to find out if the existence of Property A' automatically guarantees the existence of Property A or whether it is possible to construct a sequence satisfying only Property A' without satisfying Property A.

Consider possible variants of Sobol' sequences for low dimensions and determine how Properties A and A' depend on the choice of direction numbers. For construction direction numbers we will use the algorithm presented in Section 2.

The 1-dimensional Sobol' sequence has a special form. It is known as the van der Corput sequence. For this sequence, all direction numbers $m_{i,j} = 2^i v_{i,j}$ are equal to 1. It is easy to verify that it satisfies both Properties A and A'.

Consider the Sobol' sequence for dimension 2. We will employ commonly used variants of primitive polynomials in ascending degree order. To obtain direction numbers for dimension 2, the polynomial $P_2 = x + 1$ is used. In this case we have only one free parameter $m_{1,2}$ which can be equal only to 1. Thus, the following sequence $m_{i,2} = 1, 3, 7, 15,\ldots$ is obtained. For dimension 2, $V_2 = 1$ and $U_2 = 1$. This means that the Sobol' sequence for dimension 2 satisfies both Properties A and A'.

Consider the Sobol' sequence for dimension 3. To obtain direction numbers for dimension 3, the following polynomial is used: $P_3 = x^2 + x + 1$. In this case we have two free parameters $m_{1,3}$ and $m_{2,3}$. The value of $m_{1,3}$ is always equal to 1. For the parameter $m_{2,3}$ there are two options: $m_{2,3} = 1$ or $m_{2,3} = 3$. Thus, it is possible to construct two variants of the Sobol' sequence.

In a binary representation, let $m_{2,3} = (x1)_2$. We calculate $m_{k,3}$, $k \geq 3$ using formula (2.1) and substitute in the matrices (3.2) and (3.4). The result is

$$det(\mathbf{V}_3) = \begin{vmatrix} 1 & 1 & 1 \\ 0 & 1 & x \\ 0 & 1 & \overline{x} \end{vmatrix} = x \oplus \overline{x} = 1,$$

$$det(\mathbf{U}_3) = \begin{vmatrix} 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & x & 1 \\ 0 & 0 & 1 & 0 & \overline{x} & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & x & 1 \\ 0 & 0 & 1 & 1 & \overline{x} & 1 \end{vmatrix} = 1.$$

The super line denotes the binary operation of negation ($\overline{x} = x \oplus 1$).

Hence, both Properties A and A' are satisfied independently of the choice of x and, respectively, $m_{2,3}$.

Consider the Sobol' sequence for dimension 4. To obtain direction numbers, the following polynomial is used: $P_4 = x_3 + x + 1$. In this case there are three free parameters: $m_{1,4} = 1$, $m_{2,4}$, and $m_{3,4}$. Parameter $m_{2,4}$ can take two possible values (1 or 3), while parameter $m_{3,4}$ can take four possible values (1, 3, 5, or 7). Parameter $m_{2,3}$ can take two possible values, hence there are 16 different variants of the Sobol' sequence for dimension 4.

In a binary representation, let $m_{2,3} = (x1)_2$, $m_{2,4} = (y1)_2$, and $m_{3,4} = (zt1)_2$. Using relationships (3.2), (3.4), and (3.6)–(3.7), we obtain:

$$det(\mathbf{V}_4) = \begin{vmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & x & y \\ 0 & 1 & \overline{x} & z \\ 0 & 1 & 1 & \overline{y} \end{vmatrix} = \overline{x}\,(y \oplus z) \oplus 1,$$

$$det(\mathbf{U}_4) = \begin{vmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & x & 1 & y & 1 \\ 0 & 0 & 1 & 0 & \overline{x} & 1 & z & t \\ 0 & 0 & 1 & 1 & 1 & 0 & \overline{y} & 1 \\ 0 & 0 & 1 & 0 & x & 1 & z \oplus y & \overline{t} \\ 0 & 0 & 1 & 1 & \overline{x} & 1 & z \oplus \overline{y} & \overline{t} \\ 0 & 0 & 1 & 0 & 1 & 0 & \overline{z} & t \\ 0 & 0 & 1 & 1 & x & 1 & 1 & 0 \end{vmatrix} = t.$$

As a result, from 16 possible variants of the Sobol' sequence, 12 variants satisfy Property A (8 variants for which $m_{2,3} = 3$ as well as 4 variants for which $m_{2,3} = 1$ and $y = z$, while t can attain any value). 8 variants satisfy Property A', for which $t = 1$. The complete list of all variants with values of free parameters and conditions for Properties A and A' is given in Table 1.

We can conclude that it is possible to construct a Sobol' sequence for which Property A' is satisfied while Property A is not satisfied. Examples of such sequences are sequences 4 and 6 (Table 1). It can be assumed that such a construction is also possible for sequences of dimensional orders higher than 4.

4 About additional uniformity properties

Originally, Sobol' sequences were known as LP_τ-sequences, therefore in this section we will keep their original name.

Consider an arbitrary LP_τ-sequence $x_0, x_1, x_2,...$ in H^d. It is known that all LP_τ-sequences are uniformly distributed in H^d, moreover all Sobol' sequences are well distributed. The lower the value of τ, the better the uniformity properties of the distribution. However, all these properties are asymptotic: they are fulfilled when the number of points is sufficiently large. At the same time, the initial sets of points can be distributed non-uniformly.

Table 1: Sobol′ sequences for dimension 4

No.	$m_{2,3}$	$m_{2,4}$	$m_{3,4}$	A	A'
1	1	1	1	Yes	No
2	1	1	3	Yes	Yes
3	1	1	5	No	No
4	1	1	7	No	Yes
5	1	3	1	No	No
6	1	3	3	No	Yes
7	1	3	5	Yes	No
8	1	3	7	Yes	Yes
9	3	1	1	Yes	No
10	3	1	3	Yes	Yes
11	3	1	5	Yes	No
12	3	1	7	Yes	Yes
13	3	3	1	Yes	No
14	3	3	3	Yes	Yes
15	3	3	5	Yes	No
16	3	3	7	Yes	Yes

Consider the initial part of the LP_τ-sequence

$$x_0, x_1, x_2, \ldots, x_{N_0-1}, \tag{4.1}$$

the length of which is equal to $N_0 = 2^{\tau+1}$.

It follows from the definition of the LP_τ-sequence that the sequence (4.1) is a P_τ-net, that is any dyadic box with a volume $2^\tau/N_0$ contains exactly 2^τ points of sequence (4.1). (A dyadic box is a product of dyadic intervals.) In the considered case this volume is equal to $2^\tau/N_0 = 1/2$. There are $2d$ dyadic boxes with a volume equal to 1/2.

It is easy to see that such uniformity in 'halves' is quite a weak criterion from the point of view of uniformity. For example, if 2^τ points belong to the hyperoctant $[0, 1/2)^d$ and the remaining 2^τ points belong to the hyperoctant $[1/2, 1]^d$, then the condition of uniformity with respect to 'halves' of the unit hypercube H^d is fulfilled.

4.1 LP_τ-sequence satisfying Property A

Consider an LP_τ-sequence satisfying Property A for the case $\tau \geq d$. Then the length of the sequence (4.1) can be presented as $N_0 = 2^{d+p}$, where $p = \tau - d + 1 \geq 1$. Therefore the sequence (4.1) consists of 2^p sections containing 2^d points that belong to various hyperoctants. Hence, there are exactly 2^p points of the sequence (4.1) in each of the 2d hyperoctants.

This condition is much stronger than uniformity with respect to 'halves.' Indeed, each of the 'halves' consists of 2^{d-1} hyperoctants, therefore the number of points belonging to each of the 'halves' is equal to $2^{d-1} * 2^p = 2^\tau$.

Consider now the case when $\tau < n$. Then $N_0 \leq 2^d$ and all points of the sequence (4.1) belong to different hyperoctants.

4.2 LP_τ-sequence satisfying Property A'

Assume that the LP_τ-sequence satisfies Property A'. Consider first the case when $\tau \geq 2d$. Then the length of the sequence (4.1) can be presented as N_0 =

2^{2d+q}, where $q = \tau - 2d + 1 \geq 1$. Hence the sequence (4.1) is a sum of 2^q sets consisting of 4^d points, each of which belongs to one of the 4^d hypercubes with a volume $1/4^d$ (which appear in the definition of Property *A*'). We denote such hypercubes as $H_{1/4}$. Therefore, each of these 4^d hypercubes contains 2^q points of the sequence (4.1). This property is stronger than uniformity in 'halves.' Indeed, each 'half' consists of $1/2*(4^d)$ hypercubes and contains $1/2 * (4^d) * 2^q = 2^\tau$ points.

In the case when $\tau < 2d$, the length of the sequence is $N_0 \leq 4^d$ and all points of (4.1) belong to different $H_{1/4}$ hypercubes.

4.3 The case of small sets of points

Consider the sequence $x_0, x_1, x_2, \ldots, x_{M-1}$ when $M < N_0$. In this case, from the definition of an LP_τ-sequence, it is not possible to extract any additional information on the distribution of these points.

It follows from Property *A* that the number of points in various hyperoctants differs in not more than one point. Property *A*' implies that the number of points in various $H_{1/4}$ hypercubes differs in not more than one point.

5 Properties *A* and *A*′ for adjacent dimensions

We introduce two additional uniformity conditions that expand Properties *A* and *A*'. Any low-dimensional projection of the Sobol' sequence can be considered as a low-discrepancy Sobol' sequence. Therefore, we can impose additional uniformity properties for some projections of the sequence.

Definition 10. A low-discrepancy d-dimensional sequence is said to satisfy *Property* A_k for some fixed $k \leq d$ if for any set of k adjacent dimensions a projection of the sequence on this set satisfies Property *A*.

Definition 11. A low-discrepancy d-dimensional sequence is said to satisfy *Property* A'_k for some fixed $k \leq d$ if for any set of k adjacent dimensions a projection of the sequence on this set satisfies Property *A*'.

Properties *A* and *A*' are special cases of properties A_k and A'_k, respectively under condition $k = d$.

We can formulate algebraic equations that guarantee Properties A_k and A'_k.

Theorem 3. The d-dimensional Sobol' sequence satisfies Property A_k ($k \leq d$) if and only if

$$det(\mathbf{V}_{j,k}) = 1(mod\ 2), \tag{5.1}$$

for any $1 \leq j \leq (d - k + 1)$, where $V_{j,k}$ is the $k \times k$ binary matrix defined by

$$\mathbf{V}_{j,k} = \begin{pmatrix} v_{j,1,1} & v_{j+1,1,1} & v_{j+2,1,1} & \cdots & v_{j+k-1,1,1} \\ v_{j,2,1} & v_{j+1,2,1} & v_{j+2,2,1} & \cdots & v_{j+k-1,2,1} \\ v_{j,3,1} & v_{j+1,3,1} & v_{j+2,3,1} & \cdots & v_{j+k-1,3,1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ v_{j,k,1} & v_{j+1,k,1} & v_{j+2,k,1} & \cdots & v_{j+k-1,k,1} \end{pmatrix} \tag{5.2}$$

with $v_{m,n,1}$ denoting the first digit after the binary point of the mth direction-number for dimension n ($v_{m,n} = (0.v_{m,n,1}v_{m,n,2}\ldots)_2$).

Theorem 4. The d-dimensional Sobol' sequence satisfies Property A'_k if and only if

$$det(\mathbf{U}_{j,k}) = 1(mod\ 2), \tag{5.3}$$

for any $1 \leq j \leq (d - k + 1)$, where $U_{j,k}$ is the $2k \times 2k$ binary matrix defined by

$$\mathbf{U}_{j,k} = \begin{pmatrix} v_{j,1,1} & v_{j,1,2} & v_{j+1,1,1} & v_{j+1,1,2} & \cdots & v_{j+k-1,1,1} & v_{j+k-1,1,2} \\ v_{j,2,1} & v_{j,2,2} & v_{j+1,2,1} & v_{j+1,2,2} & \cdots & v_{j+k-1,2,1} & v_{j+k-1,2,2} \\ v_{j,3,1} & v_{j,3,2} & v_{j+1,3,1} & v_{j+1,3,2} & \cdots & v_{j+k-1,3,1} & v_{j+k-1,3,2} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ v_{j,2k,1} & v_{j,2k,2} & v_{j+1,2k,1} & v_{j+1,2k,2} & \cdots & v_{j+k-1,2k,1} & v_{j+k-1,2k,2} \end{pmatrix} \tag{5.4}$$

with $v_{m,n,i}$ denoting the ith digit after the binary point of the mth direction number for dimension $n(v_{m,n} = (0.v_{m,n,1}v_{m,n,2}\ldots)_2)$.

These two theorems can easily be obtained from Theorems 1 and 2 given in Section 3. Let's consider some arbitrary set of k adjacent dimensions $\{j, j+1, \ldots, j+k\}$. The projection of the Sobol' sequence on this set is the k-dimensional Sobol' sequence given by a set of direction vectors $(v_{m,j}, v_{m,j+1}, \ldots, v_{m,j+k})$, $m = 1, 2, \ldots$

We obtain relationships (5.2) and (5.4), substituting these direction vectors into formulas (3.2) and (3.4), respectively.

According to the definitions of Properties A_k and A'_k relationships (5.2) and (5.4) could be fulfilled for all possible values of j.

5.1 Sobol′ sequence satisfying Property A_k for all $k \leq d$

Our first objective is to investigate the possibility of constructing a Sobol' sequence that satisfies Property A_k for all $k \leq d$. First, we consider Sobol' sequences at low dimension. If it is possible to construct a sequence satisfying the required criteria at low dimensions, then it can be generalized for sequences at high dimensions.

We will use an algorithm described in Section 2 to construct the Sobol' sequence. The construction of all possible Sobol' sequences for dimensions up to 4 was considered in Section 3.1. We note that $m_{1,m} = 1$ for any dimension $m = 1, 2, \ldots$ This result can easily be obtained from the constraints $m_{m,1} \leq 2^1$ and $m_{m,1}$ is odd, therefore all Sobol' sequences satisfy Property A_1.

The 1-dimensional Sobol' sequence has a special form. It is known as the van der Corput sequence. For this sequence, all direction numbers $m_{i,j} = 2^i v_{i,j}$ are equal to 1. Consider the Sobol' sequence for dimension 2. To obtain direction numbers for dimension 2, a polynomial $P_2 = x + 1$ is used. In this case we have only one free parameter $m_{1,2}$, which can only be equal to 1. Thus, the following sequence $m_2 = \{1, 3, 7, 15, \ldots\}$ is obtained. Property A_2 is the same as Property *A* for dimension $d = 2$. Therefore, Properties *A*, A_1, and A_2 are satisfied for this sequence.

Consider the Sobol' sequence for dimension 3. To obtain direction numbers for dimension 3 the following polynomial is used: $P_3 = x_2 + x + 1$. In this case we have two free parameters $m_{1,3}$ and $m_{2,3}$. The value of $m_{1,3}$ is always equal to 1. For the parameter $m_{2,3}$ there are two options: $m_{2,3} = 1$ or $m_{2,3} = 3$.

Let's denote $m_{2,3} = (x1)_2$. If we impose Property A_2, we obtain

$$det(\mathbf{V}_{1,2}) = \begin{vmatrix} 1 & 1 \\ 0 & 1 \end{vmatrix} = 1,$$

$$det(\mathbf{V}_{2,2}) = \begin{vmatrix} 1 & 1 \\ 1 & x \end{vmatrix} = x \oplus 1 = \overline{x} = 1,$$

$$x = 0 \Rightarrow m_{3,2} = 1.$$

We find that only one 3-dimensional Sobol' sequence with $m_3 = \{1, 1, 7, 11,...\}$ satisfies Property A_2. As already checked in Section 3.1, this sequence satisfies Property A and Property A_3. Therefore, we constructed the 3-dimensional sequence that satisfies Properties A, A_1, A_2, and A_3.

Consider the Sobol' sequence for dimension 4. To obtain direction numbers, the following polynomial is used: $P_4 = x^3 + x + 1$. In this case there are three free parameters $m_{1,4} = 1$, $m_{2,4} \in \{1, 3\}$, and $m_{3,4} \in \{1, 3, 5, 7\}$. Value $m_{2,4}$ can be found from a condition required by Property A_2. We have already imposed conditions $det(V_{1,2}) = 1$ and $det(V_{2,2}) = 1$. Therefore, we check only the determinant $V_{3,2}$, which is equal to

$$det(\mathbf{V}_{3,2}) = \begin{vmatrix} 1 & 1 \\ 0 & x \end{vmatrix} = x = 1.$$

This gives $m_{2,4} = 3$. A requirement for Property A_3 leads to the following relation:

$$det(\mathbf{V}_{2,3}) = \begin{vmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & y \end{vmatrix} = y \oplus 1 = \bar{y} = 1,$$

$$y = 0 \Rightarrow m_{3,4} \in \{1, 3\}.$$

Therefore, we have two possible sets of initials numbers for the fourth dimension: $m_4^{(1)} = \{1,3,1,5,...\}$ and $m_4^{(2)} = \{1,3,1,5,...\}$. As already checked in Section 3.1, both these sequences do not satisfy Properties A and A_4. Therefore, we can formulate the following lemma.

Lemma 2. It is not possible to construct a Sobol' sequence for dimension $d \geq 4$ which satisfies Property A_k for all $k \leq d$.

5.2 Sobol′ sequence satisfying Property A_k for arbitrary k

If we impose Property A_k only for one fixed k, then we can construct the Sobol' sequence under this condition. For example, the requirements of Property A_2 lead to the following conditions:

$$m_{1,n} = 1, n = 1, 2, \ldots$$

$$m_{2,n} = \begin{cases} 1, & \text{if } n \text{ is odd,} \\ 3, & \text{otherwise,} \end{cases}$$

It is easy to see that the sequence with these properties can be constructed for any dimension d. Similar conditions can be derived for other $k > 2$, but it requires analysis in each case. Therefore, we need to select some fixed value k to construct the Sobol' sequence. The Sobol' sequence that satisfies Property A_k for all k does not exist.

6 Comparison of Sobol′ sequence generators

In the present work we developed the Sobol' sequence generator which satisfies additional uniformity properties: Property A for all 16,384 dimensions and Property A' for 5 adjacent dimensions (see BRODA, 2011). It is called the SobolSeq16384 generator. The performance of the SobolSeq16384 generator is compared with that of other known generators on a set of test problems. All other generators were taken from the Quantlib library, popular among practitioners in finance (see Quantlib, 2011).

It contains the following sets of direction numbers for Sobol' sequence generators:

1. The unit direction numbers suggested in Press *et al*. (2007). This generator fails the test for Properties A and A' even for low dimensions. It is the worst choice of direction numbers and was not considered in our tests.
2. The direction numbers for dimensions up to 40 constructed by Sobol' and Levitan (see Sobol' and Levitan, 1976). The same set of direction numbers was used later by Bratley and Fox (1988). This generator was not considered in this report.
3. Direction numbers up to 32 dimensions provided by Jaeckel (2002). In our results, this generator is referred to as *Jaeckel*.
4. The implementation of Lemieux, Cieslak, and Luttmer (further referred to as *Lemieux*) includes coefficients of the free direction integers up to dimension 360. Coefficients for $d \leq 40$ are the same as in Bratley and Fox (1988). For dimensions $40 < d \leq 360$ the coefficients were calculated as optimal values based on the 'resolution' criterion (see Lemieux *et al*., 2004 for details).
5. Two series of direction numbers provided by Joe and Kuo (Joe and Kuo, 2011). The first series contains three generators marked *Kuo1*, *Kuo2*, and *Kuo3*. Generators from the second series are denoted *JoeKuoD5*, *JoeKuoD6*, and *JoeKuoD7*. The extended set of direction numbers corresponding to the same search criteria is available on the Web (Joe and Kuo, 2011). In this work we only considered the number sets used in the Quantlib library (Quantlib, 2011).

We also included the generator *Sobol'-370*, which can be seen as an extension of the generator presented in Sobol' *et al*. (1992). A summary of all the generators used for comparison is given in Table 2. The dimensions shown for the generators from the Quantlib library are the maximum tabulated dimensions.

6.1 Discrepancy test

In the first test, L_2-discrepancy (often denoted T_N) was calculated for dimensions 8 and 50. The maximum number of sampled points used for calculations was 65,536. The test results are presented in Figures 1 and 2. The behavior of various generators at high dimensions ($d > 50$) is almost identical due to some limitations of the computational algorithm for discrepancy, and it is not presented in this report.

Table 2. List of generators used in tests

Generator	Maximum dimension
Sobol′-370	370
SobolSeq16384	16384
Jaeckel	32
Lemieux	360
Kuo1	4925
Kuo2	3946
Kuo3	4586
JoeKuoD5	1999
JoeKuoD6	1799
JoeKuoD7	1899

For low dimensions (up to 20), the generators Sobol'-370 and SobolSeq16384 show the best performance (lowest values of the discrepancy), while for higher dimensions these two generators are at least as good as the best-performing Quantlib generators.

6.2 Subcube volume test

In this test the following test integral was considered:

$$V(d) = \int_0^1 \cdots \int_0^1 \left(\prod_{i=1}^{d} H(a - x_i) \right) dx_1 \cdots dx_d, \quad H(t) = \begin{cases} 0, & t < 0, \\ 1, & t \geq 0. \end{cases}$$

The value of the integral is equal to the volume of a d-dimensional cube with side length $a < 1$. The approximation error δV was defined as the difference between numerical and exact values of an integral. Integral calculations were performed by using the QMC method. Values d varied from 1 to a maximum dimension for a particular generator d_{max} (see Table 2).

Two values of the parameter a, $a = 0.5$ and 0.75, were chosen for the test. For these values, the parameter approximation error will be lower if the generator satisfies Properties A and A'. For the case $a = 0.5$, only Property A is important while for the case $a = 0.75$, both properties are important for achieving high convergence of the integral.

Figures 3 and 4 show the dependence of the approximation error versus dimension at the number of sampled points equal to 2^{14}. It is known that due to properties of Sobol' sequences, the approximation error is minimal at $N = 2^k$, with k an integer. It can be shown that for the case of low dimension, the approximation error will be equal to zero if a generator satisfies both Properties A and A'.

The worst-case scenario as far as the approximation error δV is concerned is shown in Figures 5 and 6. The number of quasi-random points in

Figure 1: Discrepancy test results for dimension $d = 8$

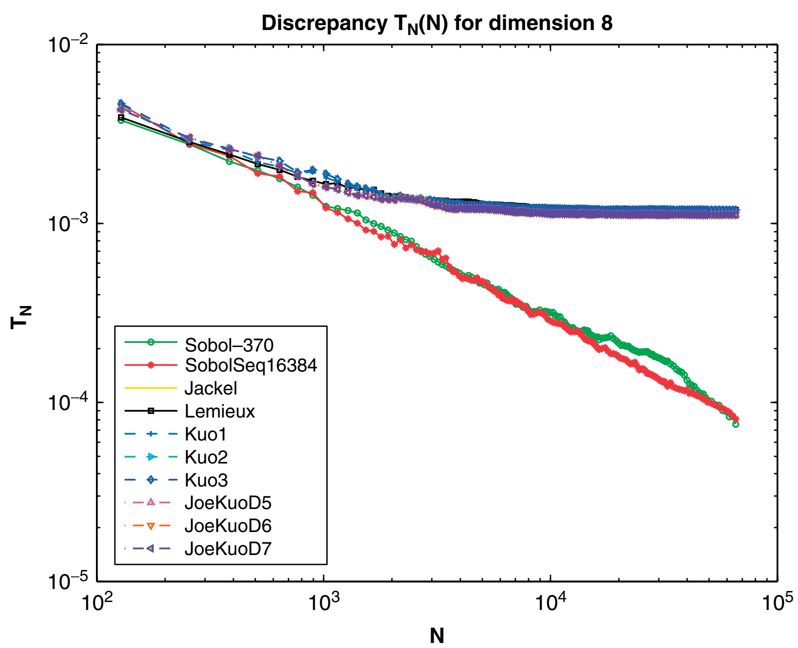


Figure 2: Discrepancy test results for dimension $d = 50$

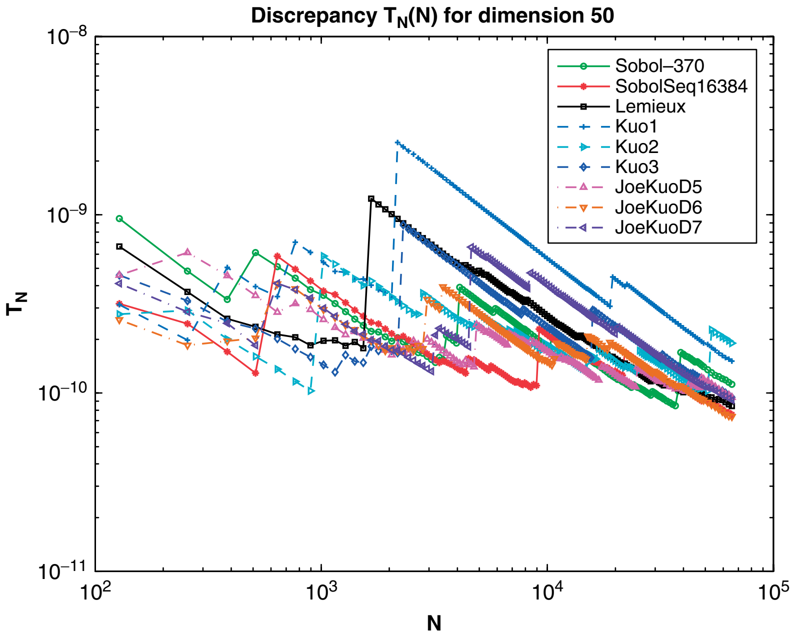
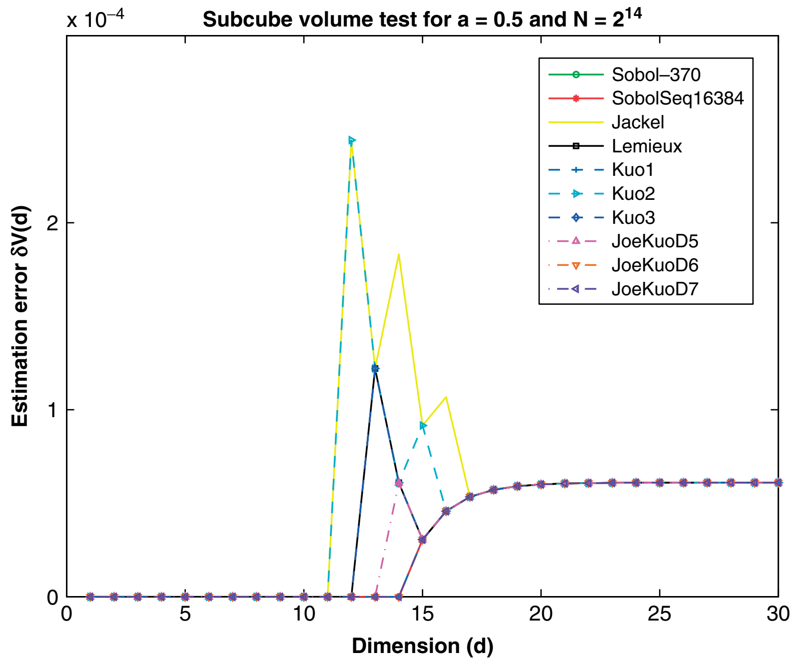


Figure 3: Subcube volume test results for $a = 0.5$ and $N = 2^{14}$



this case is a prime number, $N = 30{,}031$. The Lemieux generator shows the worst performance in this test. We also note high values of an approximation error for the Jaeckel generator and the Kuo1–Kuo3 family at $N = 2^{14}$ (see Figures 3 and 5). The same generators show results comparable with other generators in the worst case ($N = 30{,}031$). SobolSeq16384 on average shows the best performance.

Figure 4: Subcube volume test results for $a = 0.75$ and $N = 2^{14}$

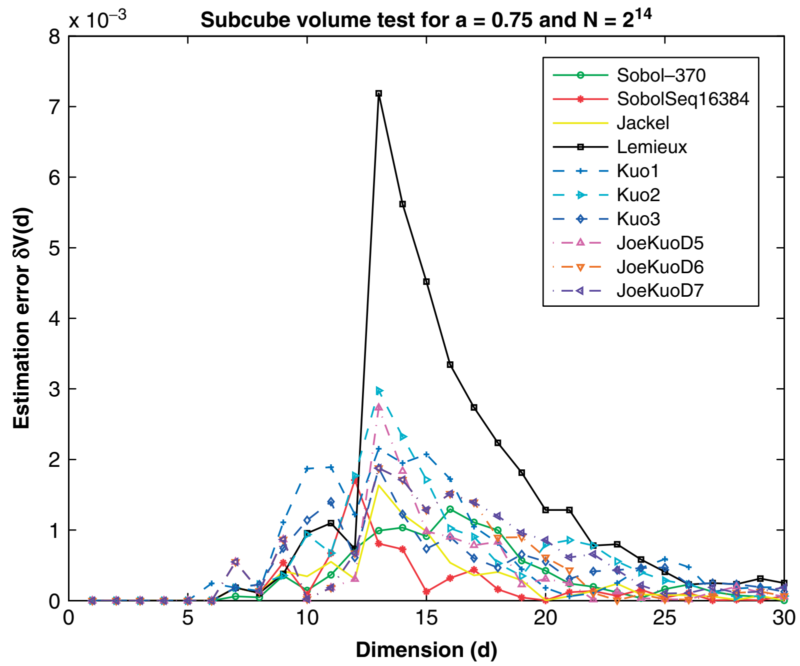


Figure 6: Subcube volume test results for $a = 0.75$ and $N = 30031$

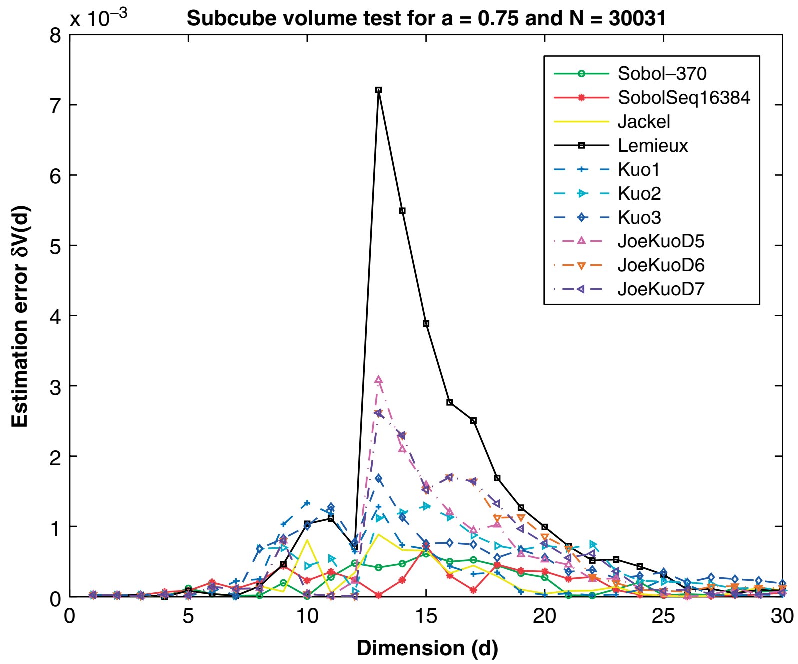


Figure 5: Subcube volume test results for $a = 0.5$ and $N = 30031$

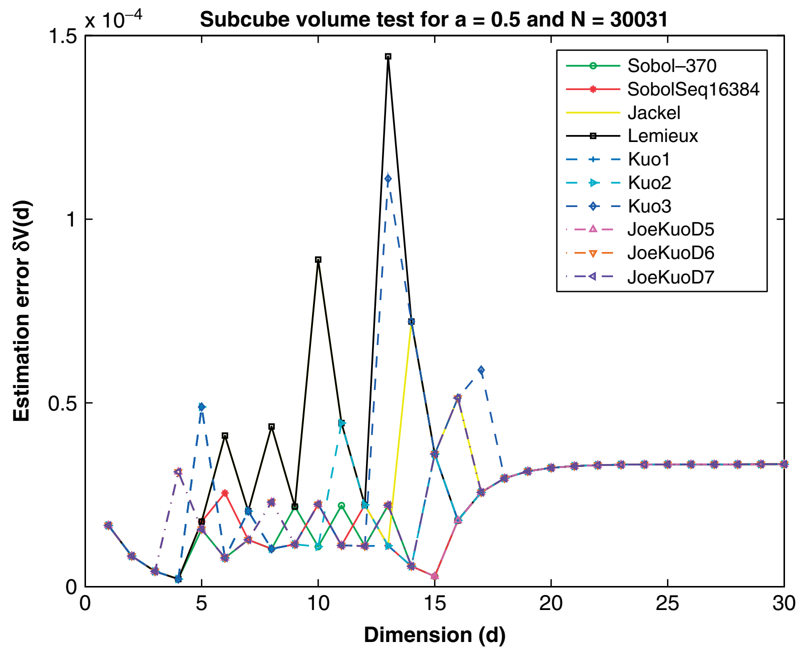
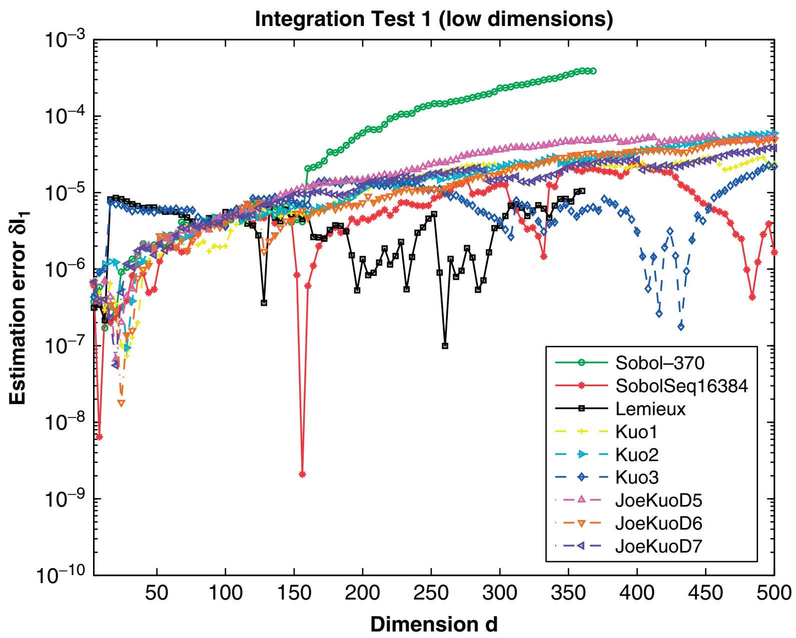


Figure 7: Integration test 1 results for low dimensions



6.3 Evaluation of high-dimensional integrals

The following two test integrals are considered in this section:

$$I_1 = \int_{[0,1]^d} \prod_{i=1}^{d} (1 + c_i (x_i - 0.5))\, dx_i \tag{6.1}$$

and

$$I_2 = \sqrt{\frac{1}{d+1}} \int_{[0,1]^d} \prod_{i=1}^{d} x_i^{\lambda_i - 1}\, dx_i, \tag{6.2}$$

where d is the problem dimension and $\lambda_i = \sqrt{\frac{i}{i+1}}$.

64-79_WILM_Nov_2011_TP_Kucherenk71 71 12/23/11 10:29:05 AM

Figure 8: Integration test 1 results for average dimensions

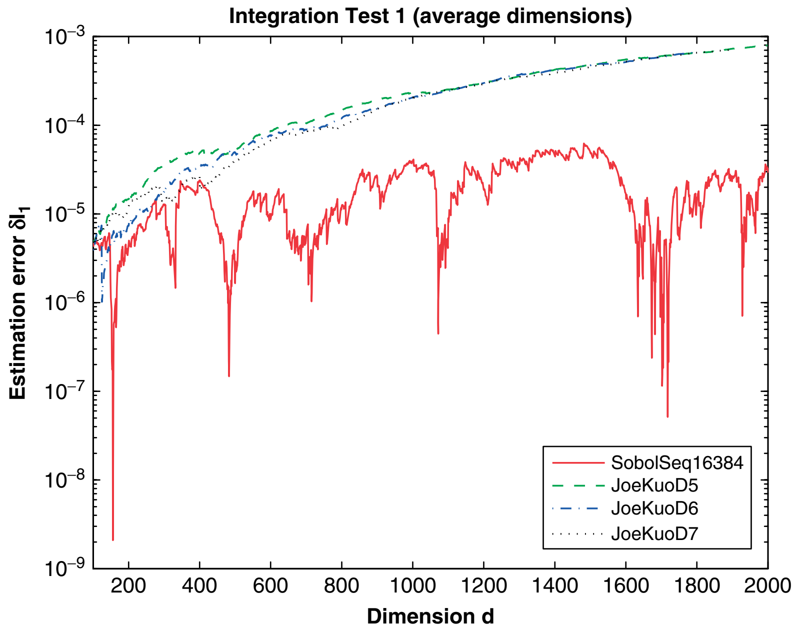
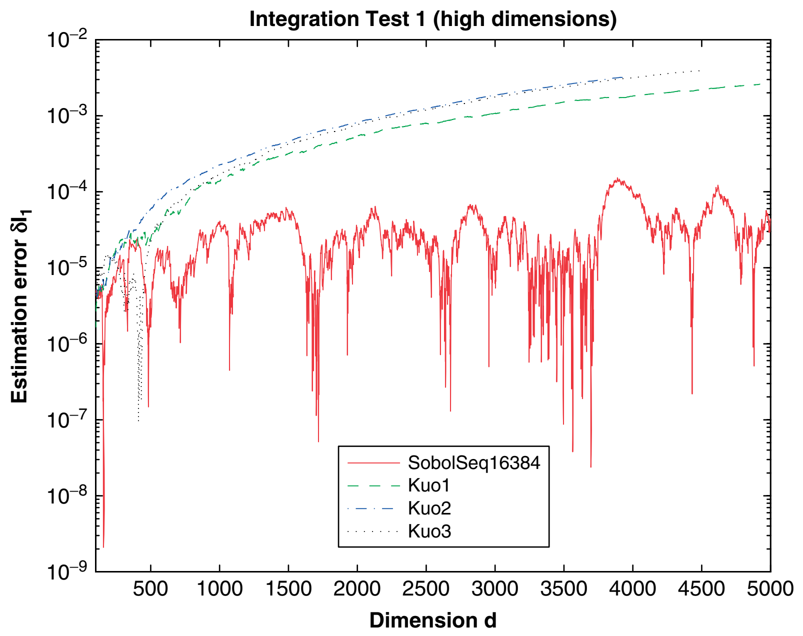


Figure 9: Integration test 1 results for high dimensions



The first integral is used for comparison of the generators; the second integral is used to compare the rate of convergence as a function of the index of the first coordinate J of the Sobol' point used for computation of the integrand in (6.2). Namely, if $(x_1, x_2,..., x_d)$ is a Sobol' point, d^* is the dimensionality of the generator, and the dimension of the integral $d < d^*$, then the integrand in (6.2) can be computed analytically:

$$f(x_1, x_2, \ldots, x_{d_*}) = \prod_{i=J}^{d+J-1} x_i^{\lambda_i - 1},$$

where $d + J - 1 < d^*$. Obviously, the results of numerical integration depend on J.

The approximation error ΔI_1 was compared for different Sobol' sequence generators. Two sets of values for parameter c_i were used. The first series of calculations (Integration Test 1) were carried out at $c_i = 0.01$, $i = 1, d$. In this case, all coordinates are equivalent. In the second series (Integration Test 2) the value of the parameter c_i was taken to be $c_i = 0.01/i$. In this case coordinates with lower index number have greater weights in the integrand. The exact value of the integral I_1, in both cases is equal to 1.

The results are shown in Figures 7–15. Values of an approximation error δI versus dimension d are presented in Figures 7–9 (Test 1). There are three plots corresponding to different parameters c: the first one is for low-order (less than 500) dimensions, the second one is for average-order (less than 2000) dimensions, and the third one is for high dimensions. The number of sampled points is equal to 30,031 in all cases.

In Test 1 for low-order (less than 500) dimensions, three generators (Lemieux, Kuo3, and SobolSeq16384) are superior to other generators, while for average-order (less than 2000) dimensions and for high dimensions clearly the SobolSeq16384 generator shows the best performance. Similar calculations of the approximation error ΔI_1 in Test 2 show that all considered generators are comparable in performance with the difference between the best and worst-performing generators being only $\approx 4\times10^{-7}$.

In Figures 10–15 the convergence of an approximation error δI for Tests 1 and 2 is shown for three different dimensions d ($d = 10$, $d = 100$, and $d = 1000$). The best performance is shown by the SobolSeq16384 generator in all tests.

Figure 10: Integration test 1 results for dimension $d = 10$

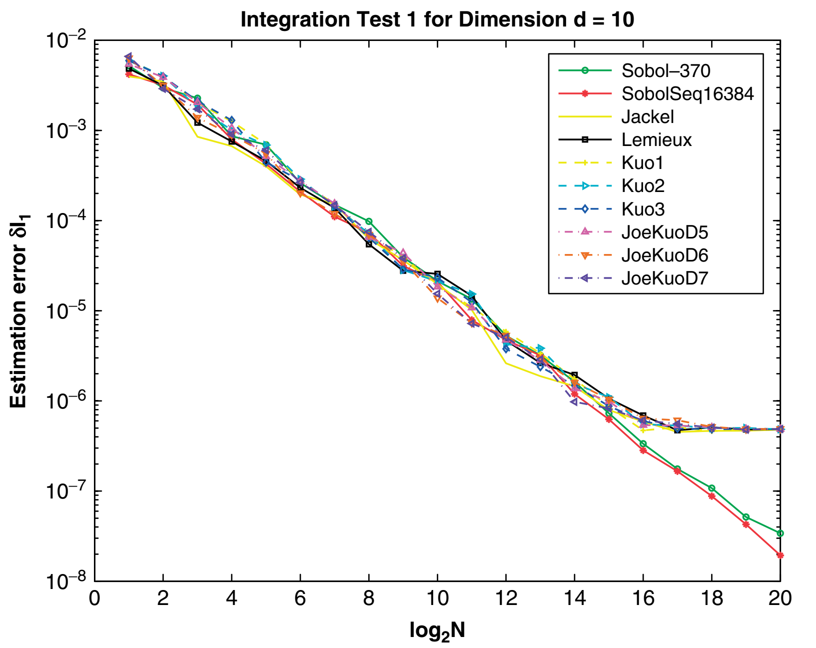


Figure 11: Integration test 1 results for dimension $d = 100$

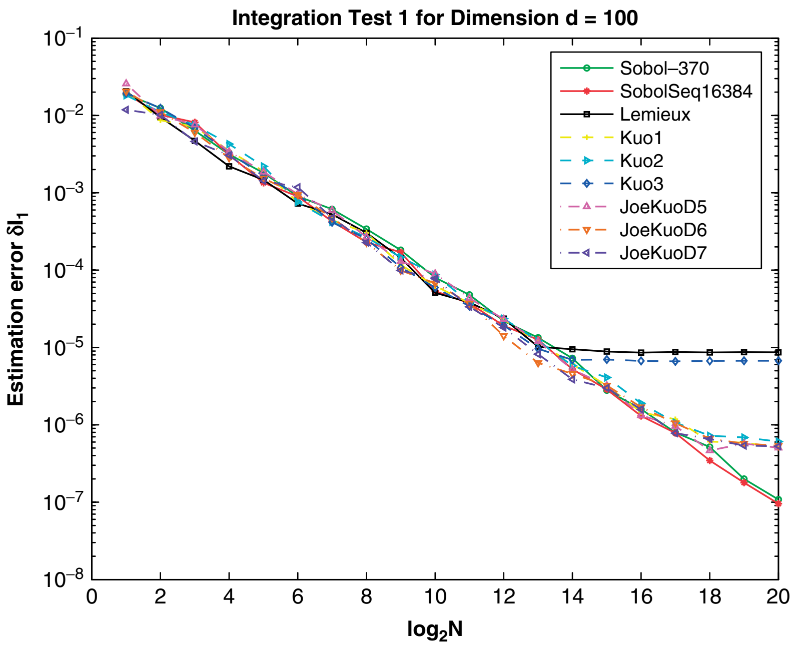


Figure 12: Integration test 1 results for dimension $d = 1000$

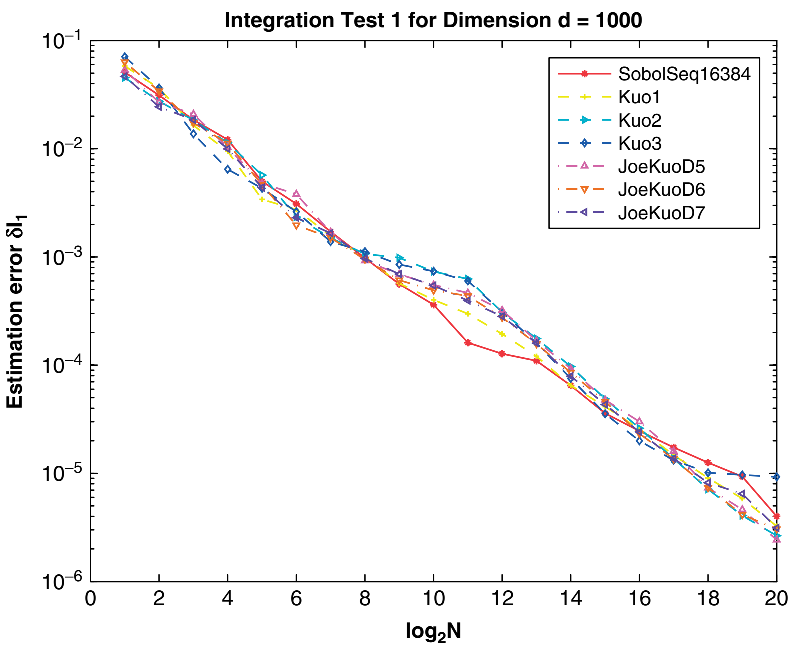


Figure 13: Integration test 2 results for dimension $d = 10$

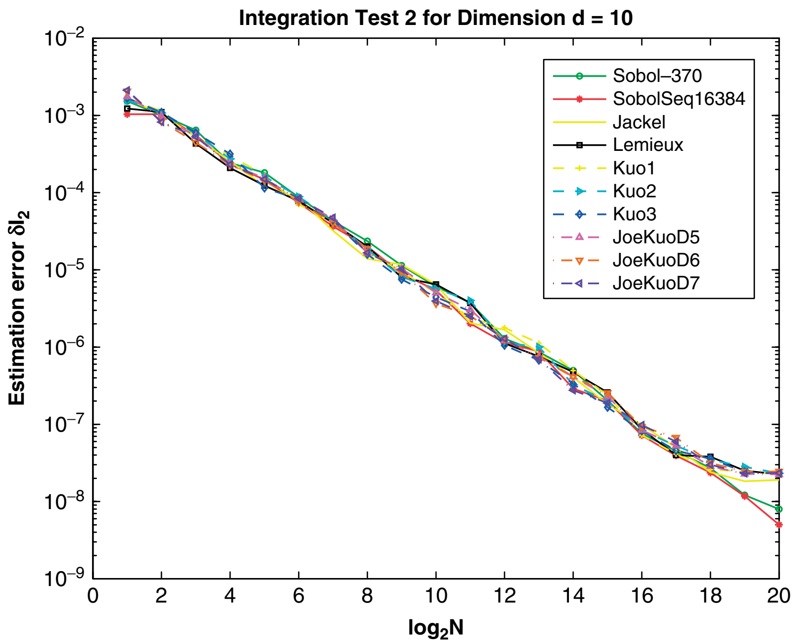
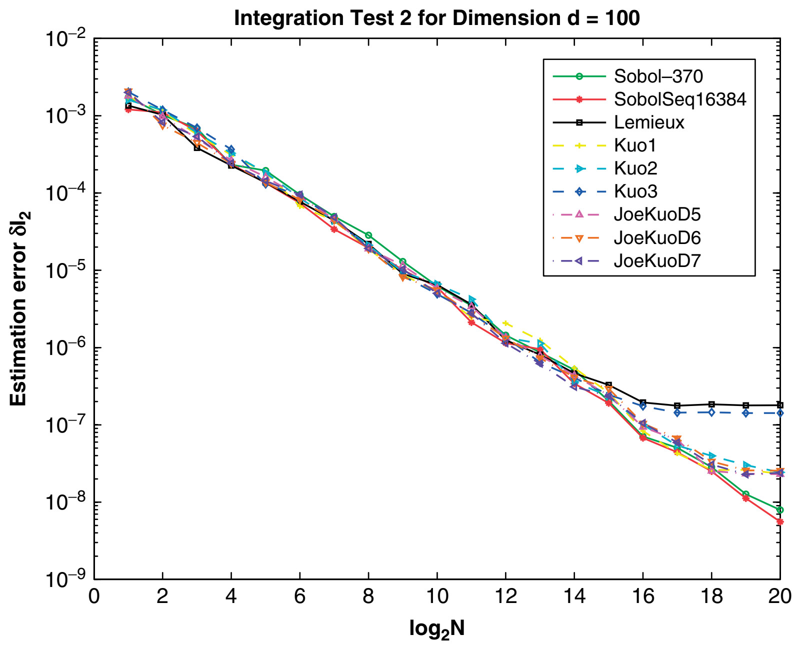


Figure 14: Integration test 2 results for dimension $d = 100$



Analysis of the relative error for the integral I_2 in (6.2) as a function of the first coordinate index J of the Sobol' points is presented in Figures 16 and 17. Notice that the value of this integral, $I_2 = 1$. For this particular test we considered only the SobolSeq16384 generator.

One could expect a steady deterioration of the approximation quality with increasing J. In fact, this deterioration is not observed in our numerical experiments with the integral I_2. The results presented in these figures demonstrate that the choice of the initial index J of the Sobol' points has limited effect on the approximation error.

6.4 Evaluation of improper integrals

Consider the improper multidimensional integral in the unit hypercube. It is possible to use low-discrepancy sequences for the evaluation of an integral

Figure 15: Integration test 2 results for dimension d = 1000

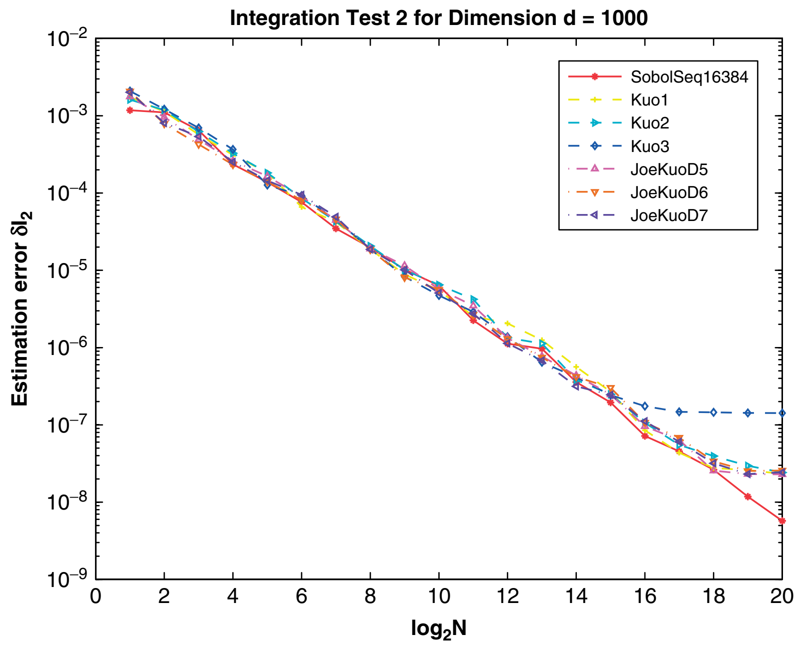


Figure 16: Relative error of computation of the integral I_2; d = 1024

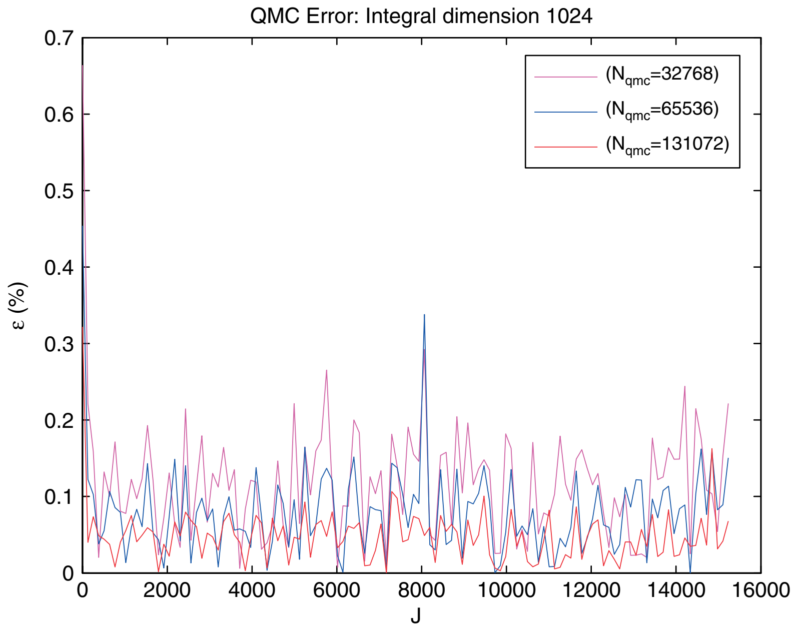


Figure 17: Relative error of computation of the integral I_2; d = 2048

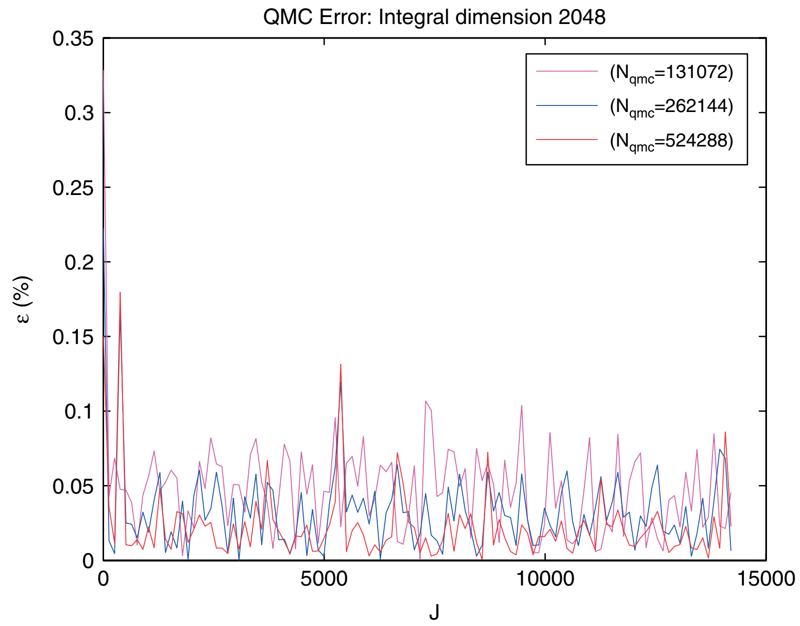
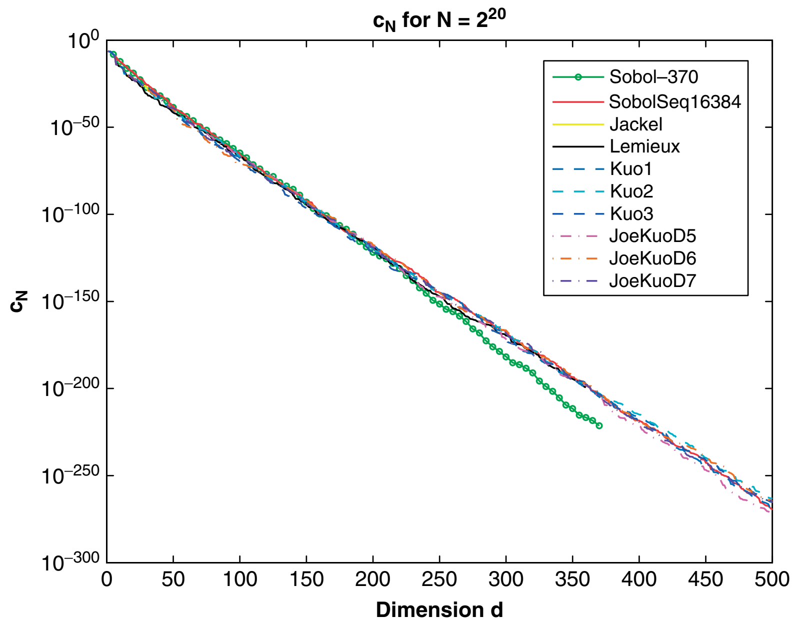


Figure 18: Improper integral test results: value c_N for $N = 2^{20}$ vs. dimension of integral



$I = \int_{[0,1]^d} \phi\left(a^{-1}\ln x_i\right) dx_i$ which has a singularity at the origin (this was first-considered in Sobol', 1967).

Denote by P_k the points of an equidistributed sequence that we will use as integration points. Let $P_k = (x_{k,1},\ldots, x_{k,d})$. By definition, c_N is the minimum of the product

$$c_N = \min_{1\le k\le N}(x_1\cdots x_d)$$

A set of sufficient conditions for the validity of the equality

$$\lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N} f\left(P_k\right) = \int_0^1\ldots\int_0^1 f(P)dP$$

is given by Theorem 2 in Sobol' (1967). One of these conditions requires that

$$N\cdot D_N\int_{G(c_N)}\left|\frac{\partial^d f}{\partial x_1\cdots\partial x_d}\right| dx_1\cdots dx_d = o(N)$$

Figure 19: Improper integral test results: value c_N for dimension $d = 10$

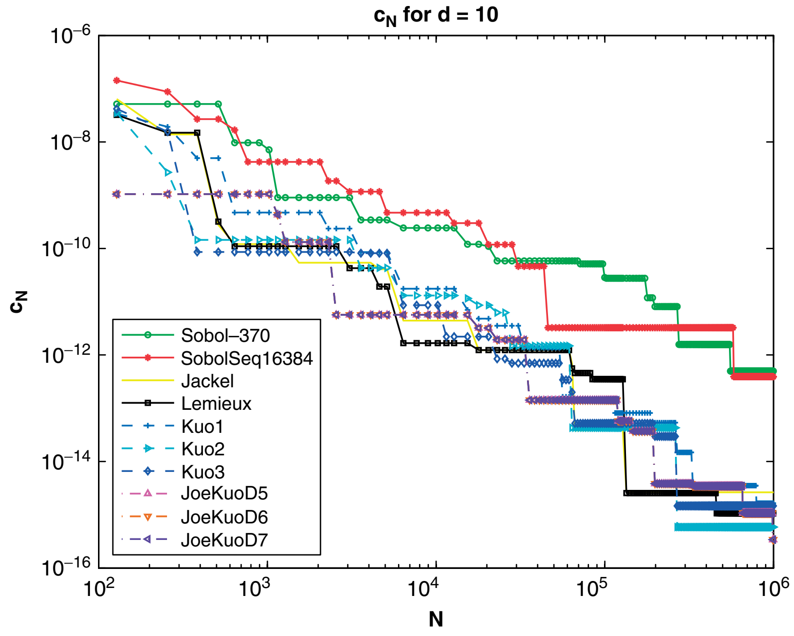
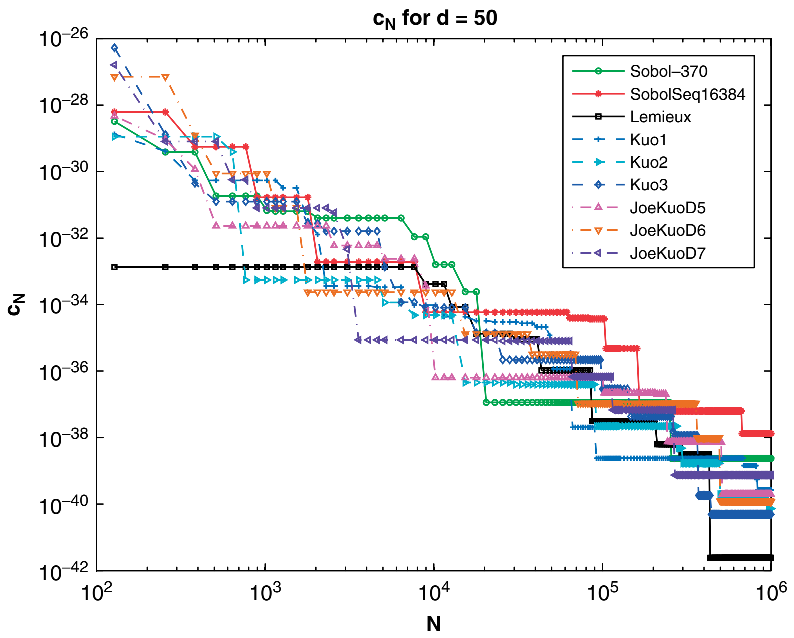


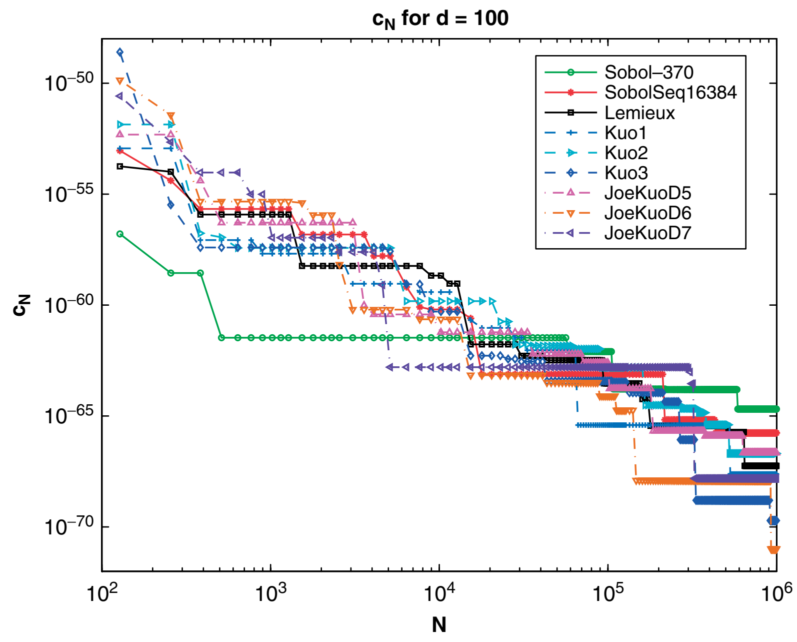
Figure 20: Improper integral test results: value c_N for dimension $d = 50$



at $N \to \inf$, where $G(c_N)$ is the part of the unit cube in which the inequality $x_1 \cdots x_d \geq c_N$ holds, D_N is discrepancy, d is the dimension of the integral.

In the following test, dependence of the value c_N versus dimension d and the number of sampled points N was monitored. The point with zero coordinates was discarded from the calculations. It is obvious that an approximation error will be smaller for higher values of c_N. For random numbers the value c_N can not be estimated, therefore the Monte Carlo method can not be used for an estimation of such integrals. Comparison of Sobol' sequences

Figure 21: Improper integral test results: value c_N for dimension $d = 100$



with other low-discrepancy sequences can be found in Asotsky and Sobol' (2005).

We compared the values of c_N for different Sobol' sequence generators. Values of c_N versus dimension d are shown at the fixed number of quasi-random points $N = 2^{20}$ in Figure 18. Values of c_N versus the number of computational points N at three different dimensions d ($d = 10$, $d = 50$, and $d = 100$) are shown in Figures 19–21. Sobol'-370 and SobolSeq16384 generators show the best performance for low dimensions. All generators show equally good performance for average and high dimensions. The efficiency of the Sobol'-370 generator deteriorates as the dimension d increases.

7 Some Applications in Financial Modeling

In this section we consider examples of the application of numerical strategies based on QMC scenario generation. Computation of sensitivities of financial derivatives by MC simulation represents a well-known problem (see Glasserman, 2004): a straightforward approximation by finite differences usually leads to a very poor approximation of sensitivities. Alternative methods evaluate sensitivities directly. These methods are based on the following ideas: differentiation of the paths and differentiation of the measures (see Glasserman, 2004; Chen and Glasserman, 2007). The second group of methods is also often called the likelihood ratio method (LLR).

The LLR method appeared to be an efficient tool for computation of the sensitivities of the path-dependent options. This method allows us to find sensitivities without re-evaluation of the financial derivatives.

Consider an arithmetic Asian option on the underlying stock whose dynamics under a risk-neutral measure $\mathbb{Q}$ are described by the standard Black–Scholes model

$$S_t = S_0 \exp(\mu t + \sigma W_t), \quad \mu = r - \frac{1}{2}\sigma^2,$$

where r is a risk-free rate, σ is the stock volatility, W_t is a standard Brownian motion, and $S_0 > 0$ is the initial value of the stock at time 0.

Let $0 < t_1 < t_2 < \cdots < t_d = T$. Denote $S_k = S(t_k)$. We consider a discrete-time model for the Asian option with discounted payoff

$$f(S_0; S_1, \ldots, S_d) = \exp(-rT)\left(\frac{1}{d}\sum_{k=1}^{d} S_k - K\right)^+, \tag{7.1}$$

where $K > 0$ is the strike of the option. (7.1)

The option value $V_0 = V(S_0) = \mathbb{E}[f(S_0; S_1, \ldots, S_d)]$, and the option sensitivity,

$$\delta = \frac{\partial V(S_0)}{\partial S_0} \approx \frac{V(S_0 + \Delta) - V(S_0)}{\Delta}, \tag{7.2}$$

can be estimated using QMC path generation. As usual, the increments of the underlying process, $\Delta W_k = W_{tk} - W_{tk}-1$, are computed as

$$\Delta W_k = \sqrt{t_k - t_{k-1}} \cdot \Phi^{-1}(\xi_k), \quad k = 1, 2, \ldots, d, \tag{7.3}$$

where $\xi = (\xi_1, \xi_2, \ldots, \xi_d)$ is a random vector with independent components uniformly distributed in the cube $[0\ 1]^d$. Thus, using different implementations of the uniform vector generators one can easily build the generators of independent normal vectors.

Consider the Asian call option with parameters $S_0 = K = 1.45$, $r = 0.0376$, $\sigma = 0.3$ and assume that $T = 5$ (yr) and the number of time steps, $d = 120$. In this example, the sample paths of the underlying process form vectors in the space $\mathbb{R}^d_+$. Thus, the discrete-time model dimension, dim = d in our example.

We compare three numerical strategies: MC, QMC_M (MATLAB implementation of the Sobol' generator), and QMC_B (BRODA's generator SobolSeq16384). There is no analytical formula for the price of the Asian option. The benchmark value is computed using $N_* = 2^{22}$ paths. Both QMC strategies estimate the value of the Asian option to be $V_* = 10.4907$. The sensitivity value is estimated as follows. At first, we perturb the initial value of the underlying equity, $S^*_0 = S_0 \cdot (1 + \varepsilon)$, and compute the values of the payoff function in the new set of risk-neutral scenarios of size $N = 2^{n_0}$ ($n_0 = 10, 11, \ldots, \log_2(N_*)$) to estimate $V_* = V(S^*)_0$. This computation allows us to find the 'benchmark' value of the sensitivity, for $N = N_*$:

$$\delta_* = \frac{V_* - V_0}{S^*_0 - S_0}.$$

The LLR method allows one to compute V^{LLR} without computation of the payoff function in the new set of scenarios (see Glasserman, 2004).

7.1 Simulation results

The simulation process was organized as follows. The group of $N_0 = 1024$ scenarios was generated and then used for pricing of the Asian option. Then the payoff function is computed in each risk-neutral scenario. The number of groups is 4096 in our experiment.[3]

Table 3 represents the simulation results for the Asian option. The first column is the type of the generator, the second column represents the range of the results, the third column represents the standard deviation of the results in the groups, the fourth column is the discounted option value, the fifth and sixth columns represent estimators of V_*. The last column represents the values of the sensitivity, δ.

Table 3: Comparison of the simulation results for Asian option

Generator	Range	σ	V_0	V_*	V_*^{LLR}	δ
SobolSeq16384	[10.0137 11.0366]	0.021	10.4907	10.6943	10.6942	81.44
MATLAB	[9.96 11.04]	0.027	10.4906	10.694	10.6941	81.44
MC	[8.68 12.52]	0.27	10.4997	10.7035	10.7034	81.43

Figure 22: Rate of Convergence

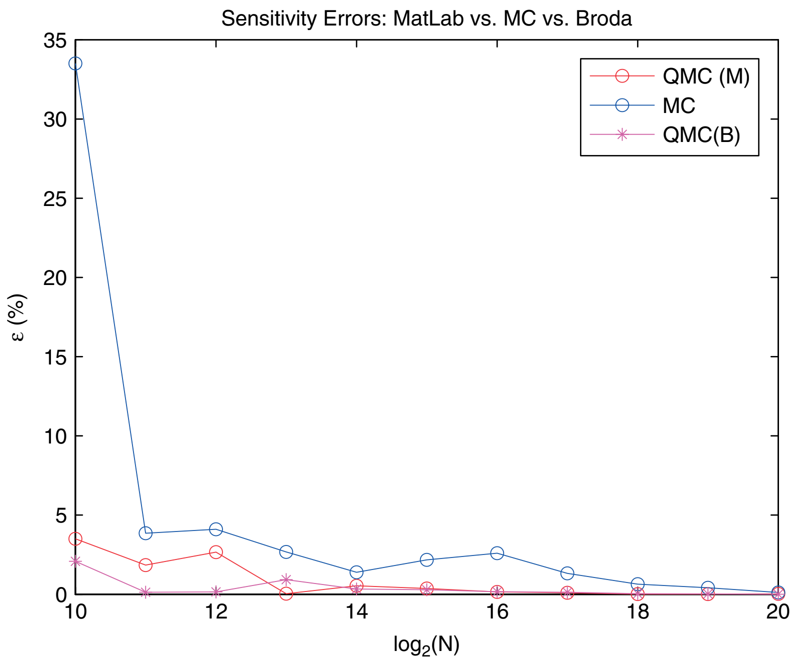


Figure 22 displays the convergence rate of the MC and QMC methods. The QMC methods outperform the MC method, with the SobolSeq16384 generator being the most efficient among two compared generators.

7.2 Computation of credit value adjustment

Our next example is computation of the credit value adjustment (CVA) – an important quantity that recently became a focus of research in the area of financial modeling among practitioners and academics. As a pricing concept, CVA represents risk of the counterparty default. The simplest approach to CVA computation is the so-called unilateral approach considered in Brigo and Masetti (2005, 2006). It neglects the effect of the possibility that the financial institution defaults before the counterparty defaults. An alternative approach to CVA computation, called bilateral CVA, is not considered in the present paper.

Denote by τ the counterparty's default time. The parameter L_c denotes the loss-given-default of the counterparty. Let T denote the final maturity of the instrument or portfolio under consideration. In the case of a portfolio,

64-79_WILM_Nov_2011_TP_Kucherenk76 76 12/23/11 10:29:15 AM

Figure 23: Accuracy of LLR method

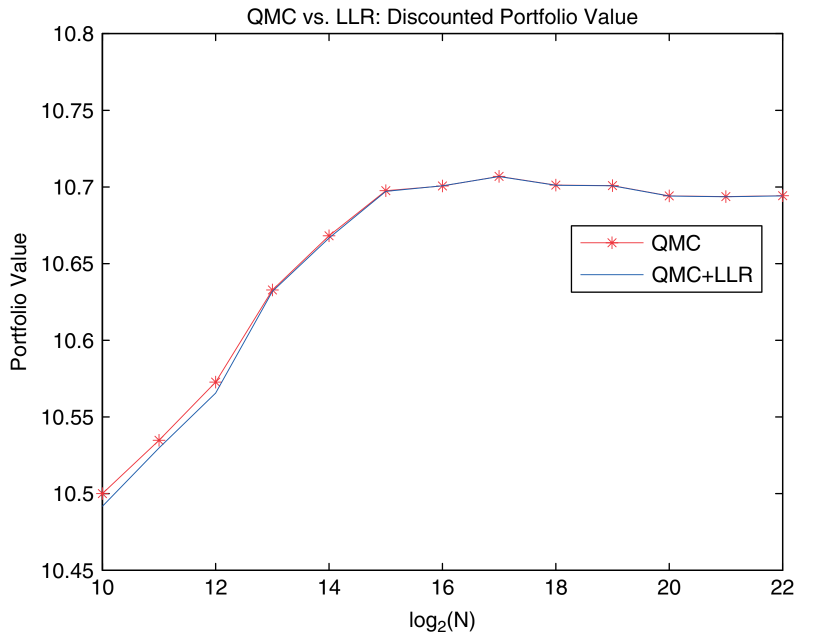


Table 4: Equity parameters

Stock index	$S_0^{(i)}$	σ_i	$Q_0^{(i)}$	σ_i^Q	r_i
1	100	0.75	1.0	0	0.045
2	100	0.5	1.5	0.35	0.055
3	100	0.5	0.4	0.45	0.025
4	100	0.45	2.4	0.35	0.045
5	100	0.45	1.6	0.5	0.045

Table 5. Parameters of put options

Instrument	Underlying	Strike	Maturity
1	1	100	16
2	1	100	15
3	2	100	14
4	2	100	13
5	3	100	12
6	3	100	11
7	4	100	10
8	4	100	9
9	5	100	13
10	5	100	14

T is the maximum of the maturities of the individual instruments which comprise the portfolio. According to Brigo and Masetti (2005, 2006), the unilateral adjustment, $A(t)$, is

$$A(t) = \mathbb{E}_t\Big[\mathbb{I}\,(t < \tau \leq T)\, L_c D(t, \tau) V_\tau^+\Big], \tag{7.4}$$

where $D(t, t')$ is the discount factor between times t and t', τ is the counterparty default time, $\mathbb{I}(\cdot)$ is the indicator function, V_t^+ is the exposure of the t to the counterparty[4] and $\mathbb{E}_t[\mathcal{A}]$ denotes the conditional institution at time expectation $\mathbb{E}[\mathcal{A}\,|\,\mathfrak{F}_t]$ with respect to a risk-neutral measure under the natural filtration $\{\mathfrak{F}_t\}_{t\geq 0}$.

Notice that V_t represents the default-free present value of the portfolio at time t $(0 < t < T)$ for the institution. Since CVA is defined in (7.4) as an integral with respect to the risk-neutral measure, it is quite natural to use QMC for its computation.

The CVA computation of large portfolios based on QMC scenario generation appeared to be an efficient risk management solution. We illustrate the results of application of this approach to a small counterparty portfolio of financial derivatives in several currencies containing stocks and put options. We assume, for simplicity, that the interest rates are deterministic in this example. There are two types of risk factor in this model: stock indices, $S^{(j)}_t$ $(j = 1, 2,..., m_S)$ and the foreign exchange rates, $Q^{(j)}_t$ $(j = 2,..., m_S)$. All risk factors are governed by the system of stochastic differential equations

$$\frac{dS_t^{(j)}}{S_t^{(j)}} = \mu_j^S\,dt + \sigma_j^S\,dW_t^{(j,S)}, \quad j = 1, \ldots, m_S, \tag{7.5}$$

$$\frac{dQ_t^{(j)}}{Q_t^{(j)}} = \mu_j^Q\,dt + \sigma_j^Q\,dW_t^{(j,Q)}, \quad j = 2, \ldots, m_S.$$

The underlying processes $W^{(j,S)}_t$ and $W^{(j,Q)}_t$ are correlated Brownian motions. The total number of risk factors is thus $m = 2 \cdot m_S - 1$. We assume that the portfolio currency has index $j = 1$. Under the risk-neutral measure, the drift coefficients μ_j, satisfy the relationships

$$\mu_j^Q = r_1 - r_j, \quad j = 2, 3, \ldots, m_S,$$

$$\mu_j^S = r_j - \sigma_j^Q \cdot \sigma_j^S \cdot \rho_j, \quad j = 1, 2, \ldots, m_S,$$

where ρ_j is the instantaneous correlation of the underlying processes $W^{(j,S)}_t$ and $W^{(j,Q)}_t$; r_j is the risk-free rate in the jth currency.

The QMC scenarios[5] are generated in accordance with (7.5). The ith scenario is represented by the matrix $\mathcal{R}(i) = \|\mathcal{R}_{kl}(i)\|$, where $\mathcal{R}_{kl}(i)$ is the value of the kth risk factor at time t_l in the scenario i $(k = 1, 2,..., m; l = 1, 2,..., n)$. Therefore, the total dimension of the scenario space is $D = m \cdot n$.

In the case of a portfolio of path-dependent options, the portfolio value, $V^{\mathcal{P}}(t_{l^0}) = V^{\mathcal{P}}(\mathcal{R}^{l^0})$ at time t_{l^0}, is a function of the sub-matrix $\mathcal{R}^{l^0} = \|\mathcal{R}_{kl}\|$, $l = l^0$, $l^0 + 1,..., n$, at time t, is a function of the sub-matrix $R = R$, $l = l0, l0 + 1,..., n$. However, in the case of the portfolio of the plain equity derivatives, the portfolio value process is Markovian and the function $V^{\mathcal{P}}(t)$ depends only on the vector $(\mathcal{R}_{1l^0}, \mathcal{R}_{2l^0},..., \mathcal{R}_{ml^0})^t$ describing the current state of the world. The effective dimension in this case should be relatively small and application of the QMC methods should be very efficient.

The portfolio, $\mathcal{P}$, contains shares of five underlying stock indices (one index per currency) and 10 at-the-money put options (two per currency). The time horizon $\mathcal{T} = 16$ months. The initial stock values, $S_0^{(i)}$, the volatility parameters, σ_i, the initial values of the FX rates, $Q_0^{(i)}$, the volatility, σ_i^Q, of the FX rates and the risk-free interest rates are shown in Table 4. Notice that the first FX rate, $Q_t^{(1)} = 1$. For this reason, we choose $\sigma_1^Q = 0$.

The portfolio simulation problem may include sub-portfolios of many counterparties. The instruments of these counterparties may require a

Table 6. Portfolio value and CVA

$\log_2 N$	CVA	$V^P(T)$	$\Delta V^P(T)$	Δ(CVA)
11	1311.06	13147.88	34.94	0.471
12	1318.43	13619.98	37.21	0.482
13	1321.5	14111.4	39.76	0.491
14	1322.84	14253.71	40.41	0.492
15	1323.54	14200.36	40.15	0.491
16	1324.17	14232.0	40.35	0.492
17	1324.11	14251.23	40.51	0.491
18	1324.57	14236.23	40.47	0.493
19	1324.61	14214.17	40.37	0.493

smaller time step for scenario generation, say weekly time steps, $\Delta t = 1/52$. The total dimension of the risk factor space, in this case, is $D = 52 \times 4 \times 9 = 1872$. However, the effective dimension of the problem is significantly smaller in our case: $D_* < 40$.

Parameters of the put options are displayed in Table 5; maturities of the options are given in months.

The risk-neutral default time distribution is defined by the equation

$$\mathbb{P}(\tau \leq t) = 1 - \exp(-\lambda t), \quad t > 0.$$

The default intensity, $\lambda = 0.04$, in our numerical example. We assume that default events are independent of the market risk factors.

The results of the CVA computation are shown in Table 6. The second column displays the credit value adjustment of the portfolio, the third column represents the portfolio value at maturity $t = T$. The fourth and last columns represent the sensitivity to the perturbation of the initial value of the first equity index by 0.5% of the portfolio value and CVA, respectively. The SobolSeq16384 generator was used in all tests.

The results demonstrate that an accurate estimation of the CVA can be obtained with $N_* = 8192$ scenarios. The portfolio value can be estimated with the relative error $\varepsilon < 0.01$. Estimation of the sensitivity of the portfolio value at time $t = 0$ with the relative error $\varepsilon < 0.01$ requires $N_* = 16{,}384$ scenarios. The CVA sensitivity estimation with N_* scenarios is quite accurate in this case.

The CVA can be a substantial part of the portfolio value. In our example, the CVA comprises about 10% of V^P. The relation between these quantities depends on the risk-neutral default probability.

8 Conclusions

The SobolSeq16384 generator allows us to solve high-dimensional problems without loss of accuracy as the dimension of the problem increases. This behavior is explained by additional uniformity properties studied in this paper, which can increase the efficiency of Sobol' sequences. Properties *A* and *A'* provide the additional guarantee of uniformity for high-dimensional problems even at a small number of sampled points. A comparison of several known Sobol' sequence generators, using a number of high-dimensional tests, shows a definite advantage of using the SobolSeq16384 generator.

Ilya M. Sobol' received his Master of Science in Maths with Honors from the Moscow State University in 1948. He received a PhD in 1959 and a Doctor of Science in 1977. He became a full professor in 1980. He has more than 150 publications on differential equations, Monte Carlo methods (including applications in nuclear physics and astrophysics), uniform distributions, quasi-Monte Carlo methods, multiple criteria decision making, global sensitivity analysis, etc. His citation index exceeds 2000. He has delivered lecture courses in many countries, including Russia, Germany, France, Italy, and Austria. He is a member of the Moscow Mathematical Society and the New York Academic Society. In 2004, *Wilmott magazine* awarded Professor Sobol' the first Wilmott fellowship. Currently he holds the position of Principal Researcher at the Keldysh Institute of Applied Mathematics, Russian Academy of Science, Moscow.

Danil Asotsky received an MSc degree in computational mathematics from Moscow Institute of Physics and Technology in 2001. He held a research position at the Institute of Mathematical Modelling, Russian Academy of Science. Currently he is a lecturer at the National Research University Moscow Institute of Electronic Technology, Russia. His research interests include Monte Carlo and quasi-Monte Carlo methods, low discrepancy sequences, and global sensitivity analysis. He is an author of several papers on Sobol' sequences and global sensitivity analysis.

Alexander Kreinin has been with Algorithmics since 1995, currently as the Senior Director of Quantitative Research. He has a PhD in Probability and Statistics from the University of Vilnius (Lithuania). He has published over 50 papers and two monographs. His research areas include market and credit risk modeling, numerical methods for risk management, Monte Carlo methods, calibration of stochastic models, semi-analytical methods of portfolio valuation, design of numerical algorithms and their software implementation. His current research projects are focused on pricing and optimization of credit portfolios. He is also an Adjunct Professor in the Computer Science Department of the University of Toronto and has been affiliated with the 'Masters of Mathematical Finance' program.

Sergei Kucherenko received his MSc degree with Honors and PhD from the Moscow Engineering Physics Institute in Russia. He has held a number of research and faculty positions in various universities in Russia, the United States, Italy, and the UK. He also worked in an investment bank. Currently, he holds the position of Senior Research Associate at Imperial College London. He is also a director of BRODA Ltd. This company provides consultancy services to investment banks and financial companies in the area of MC and quasi-MC simulation and other advanced numerical techniques used in financial mathematics.

FOOTNOTES

1 We assume that the variables are independent and the dimension *d* is very high.
2 Not an arbitrary subset.
3 Thus, the *k*th group contains scenarios with indices $j = 1024 \cdot (k-1)+1, \ldots, 1024 \cdot k$.
4 We use the standard notation $x^+ = \max(0, x)$ and $x^- = \min(0, x)$, for any real number *x*.
5 These scenarios describe the future state of the risk factors in a risk-neutral world.

REFERENCES

Asotsky, D. and Sobol', I. M. 2005. On sequences of points for the evaluation of improper integrals by quasi-Monte Carlo methods. *Comput. Math. Math. Phys.* **45**(3): 394–398.
Asotsky, D., Myshetskaya, E., and Sobol', I. M. 2006. The average dimension of a multidimensional function for quasi-Monte Carlo estimates of an integral. *Comput. Math. Math. Phys.* **46**(12): 2061–2067.
Bratley, P. and Fox, B. 1988. Algorithm 659: Implementing Sobol's quasi-random sequence generator. *ACM Trans. Math. Softw.* **14**: 88–100.
Brigo, D. and Masetti, M. 2005. A formula for interest rate swaps valuation under counterparty risk in the presence of netting agreements. Working Paper, Banca IMI, Milan, Italy.

Brigo, D. and Mercurio, F. 2006. *Interest Rate Models – Theory and Practice. With Smile, Inflation and Credit*, 2nd edn. Springer, Berlin.
BRODA Ltd. 2011. http://www.broda.co.uk
Caflisch, R., Morokoff, W., and Owen, A. 1997. Valuation of mortgage backed securities using Brownian bridges to reduce effective dimension. *J. Comput. Fin.* **1**(1): 27–46.
Chen, N. and Glasserman, P. 2007. Malliavin Greeks without Malliavin calculus. *Stochast. Proc. Appl.* **117**: 1689–1723.
Glasserman, P. 2004. *Monte Carlo Methods in Financial Engineering*. Springer, New York.
Jaeckel, P. 2002. *Monte Carlo Methods in Finance.* John Wiley & Sons, New York.
Joe, S. and Kuo, F. 2003. Remark on Algorithm 659: Implementing Sobol's quasirandom sequence generator. *ACM Trans. Math. Softw.* **1**(29): 49–57.
Joe, S. and Kuo, F. 2008. Constructing Sobol' sequences with better two- dimensional projections. *SIAM J. Sci. Comput.* **30**: 2635–2654.
Joe, S. and Kuo, F. 2011. http://web.maths.unsw.edu.au/fkuo/Sobol'/
Kreinin, A., Merkoulovitch, L., Rosen, D., and Zerbs, M. 1998. Measuring portfolio risk using quasi Monte Carlo methods. *Alg. Res. Quart.* **1**(1): 17–25.
Kucherenko, S. and Shah, N. 2007. The importance of being global. Application of global sensitivity analysis in Monte Carlo option pricing. *Wilmott* 82–91.
Kucherenko, S., Feil, B., Shah, N., and Mauntz, W. 2011. The identification of model effective dimensions using global sensitivity analysis. *Reliab. Eng. Syst. Saf.* **96**: 440–449.
L'Ecuyer, P. and Lemieux, C. 2002. Recent advances in randomized quasi-Monte Carlo methods. In *Modeling Uncertainty: An Examination of Stochastic Theory, Methods, and Applications*, M. Dror, F. L'Ecuyer, and E. Szidarovszki (eds). Kluwer Academic Publishers, Boston: 419–474.
Lemieux, C., Cieslak, M., and Luttmer, K. 2004. *RandQMC User's Guide – A Package for Randomized Quasi-Monte Carlo Methods in C*, version January 13.
Liu, R. and Owen, A. 2006. Estimating mean dimensionality of analysis of variance decompositions. *Am. Stat. Assoc. J.* **101**(474): 712–721.
Niederreiter, H. 1988. Low-discrepancy and low-dispersion sequences. *J. Num. Theor.* **30**: 51–70.
Paskov, S. and Traub, J. 1995. Faster evaluation of financial derivatives. *J. Portf. Manag.* **22**(1): 113–120.
Press, W., Teukolsky, S., Vetterling, W., and Flannery, B. 2007. *Numerical Recipes in C*, 3rd edn. Cambridge University Press, Cambridge.
Quantlib. 2011. A Free/Open-Source Library for Quantitative Finance, http://www.quantlib.org
Silva, M. and Barbe, T. 2005. Quasi-Monte Carlo in finance: Extending for problems of high effective dimension. *Econ. Apl.* **9**: 577–594.
Sobol', I. M. 1967. On the distribution of points in a cube and the approximate evaluation of integrals. *Comput. Math. Math. Phys.* **7**: 86–112 (in Russian).
Sobol', I. M. 1973. Evaluation of improper integrals using equidistributed sequences. *Dokl. Akad. Nauk SSSR* **210**(2): 278–281 (in Russian).
Sobol', I. M. 1976. Uniformly distributed sequences with an additional uniform property. *Zh. Vich Math. Math. Phys. Fiz.* 16: 1332–1337. [English translation 1976. *U.S.S.R. Comput. Math. Math. Phys.* **16**: 236–242.]
Sobol', I. M. and Levitan, Y. 1976. The production of points uniformly distributed in a multidimensional cube. *Technical Report Preprint 40.* Institute of Applied Mathematics, USSR Academy of Sciences (in Russian).
Sobol', I. M., Turchaninov, V., Levitan, Yu., and Shukhman, B. 1992. Quasirandom sequence generators. *Preprint. Keldysh Inst. Appl. Math*., Moscow.
Sobol', I. M. 1998. On quasi-Monte-Carlo integrations. *Math. Comput. Simul.* **47**: 103–112.
Wang, X. and Sloan, I. 2005. Why are high-dimensional finance problems often of low effective dimension? *SIAM J. Sci. Comput.* **27**(1): 159–183.