ELSEVIER

Mathematics and Computers in Simulation 62 (2003) 255–263

MATHEMATICS AND COMPUTERS IN SIMULATION

www.elsevier.com/locate/matcom

One more experiment on estimating high-dimensional integrals by quasi-Monte Carlo methods

I.M. Sobol'*, D.I. Asotsky

Institute for Mathematical Modelling of the Russian Academy of Sciences, 4 Miusskaya Square, 125047 Moscow, Russia

Abstract

Integrands that depend on a large number of equally important variables are considered and conditions that make expedient quasi-Monte Carlo integrations are investigated for dimensions $n \leq 300$.

Keywords: Monte Carlo method; Quasi-Monte Carlo method; Numerical integration; Quasi-random sequence

1. Introduction

The multidimensional integral

$$J[f] = \int_0^1 \cdots \int_0^1 f(x)\,\mathrm{d}x, \tag{1}$$

where $x = (x_1, \ldots, x_n)$, is often estimated by simple mean value formulas. We consider the approximation error

$$\delta_N[f] = \frac{1}{N}\sum_{k=0}^{N-1} f(x^{(k)}) - J[f]. \tag{2}$$

1.1. Monte Carlo approach

In the crude Monte Carlo (MC) estimate, the nodes $x^{(k)}$ are independent realizations of the random point ξ uniformly distributed in the unit hypercube I^n. If the integral (1) is absolutely convergent, the expectation $Ef(\xi) = J[f]$ exists, and it follows from the Law of large numbers that the averages in (2) stochastically converge. Hence, $\delta_N[f] \xrightarrow{p} 0$.

* Corresponding author. Fax: +7-95-9720723.
E-mail address: hq@imamod.ru (I.M. Sobol').

PII: S0378-4754(02)00228-8

If the integrand $f(x)$ is a square integrable, the variance $Df(\xi)$ is finite and according to the central limit theorem, the random variable $\delta_N[f]$ is asymptotically normal. Hence, for sufficiently large N the probable error r_N is

$$r_N = 0.6745\sqrt{\frac{Df(\xi)}{N}}, \tag{3}$$

which means that $P\{|\delta_N| < r_N\} \approx P\{|\delta_N| > r_N\} \approx 0.50$. According to (3), the errors of MC decrease as $1/\sqrt{N}$.

1.2. Quasi-Monte Carlo approach

In quasi-Monte Carlo (QMC) estimates the nodes $x^{(k)}$ are non-random points belonging to sequences uniformly distributed in I^n in the sense of Weyl. Such sequences having best asymptotic properties are called quasi-random (a somewhat misleading term). If the total variation of the integrand $f(x)$ is bounded, it follows from the Koksma–Hlawka inequality [1] that as $N \to \infty$

$$\delta_N = \mathrm{O}(N^{-1}\log^n N). \tag{4}$$

Comparing (3) with (4) one can see that

$$\lim_{N\to\infty}\left(\frac{\delta_N}{r_N}\right) = 0. \tag{5}$$

Relation (5) shows that QMC is more efficient than MC at all sufficiently large N. However, in practice the amount of points N is always restricted and the situation is not so clear. The most important problems where QMC is more efficient than MC include integrands $f(x)$ whose dependence on x_i decreases as the number i is increased [8]. Weighted MC algorithms are as a rule of this type.

1.3. Equally important variables

The situation is more obscure if all variables $x_1, \dots, x_n$ in $f(x)$ are equally important. On the one hand, simple examples demonstrate the advantages of QMC at relatively small dimensions only. Most authors suggest $n \le 12$–15. On the other hand, problems from financial mathematics are known where QMC outperforms MC at $n = 360$ [4].

Recently Rabitz has suggested that quite often in mathematical models low order interactions of input variables have the main impact upon the output [5]. In the ANOVA representation of such models [7]

$$f(x) = f_0 + \sum_i f_i(x_i) + \sum_{i<j} f_{ij}(x_i, x_j) + \cdots + f_{1,2,\dots,n}(x_1, x_2, \dots, x_n),$$

the main terms are the low order ones. Clearly, QMC integration of such functions will be rather efficient even in the case when n is large and all x_i are equally important. The present experiment's aim was to look closer at this situation.

2. The present experiment

2.1. Test functions

We have selected a set of test functions that depends on one parameter c, $0 < c \le 1$

$$f(x) = \prod_{i=1}^{n} \left[1 + c\left(x_i - \frac{1}{2}\right)\right]. \tag{6}$$

The case $c = 1$ was investigated in [8], cf. [2,9]. In these experiments, the dimensions were $n \le 50$, while here $n \le 300$.

Clearly, (6) is a polynomial in c. Its power expansion is identical with its ANOVA representation

$$f(x) = 1 + \sum_{s=1}^{n} c^s \sum_{i_1<\cdots<i_s} \left(x_{i_1} - \frac{1}{2}\right) \cdots \left(x_{i_s} - \frac{1}{2}\right); \tag{7}$$

in (7) the interior sum is extended over all sets of integers $i_1, \dots, i_s$ satisfying $1 \le i_1 < \cdots < i_s \le n$. The exact value of the integral (1) is $J[f] = 1$ for all n and c.

Obviously, $f(x) \to 1$ as $c \to 0$. But the convergence is non-uniform with respect to n. Indeed, for small c the range

$$\sup f(x) - \inf f(x) = (1 + \tfrac{1}{2}c)^n - (1 - \tfrac{1}{2}c)^n \approx \exp(\tfrac{1}{2}nc).$$

Therefore, if the product nc is large the function $f(x)$ has a sharp peak at the point $(1, \dots, 1)$. Thus, the integrand (6) is very "bad" when nc is very large.

2.2. Integration nodes

For $x^{(k)}$ in (2) we have selected points belonging to LP_τ-sequences with additional uniformity properties. The LP_τ-sequences were defined and constructed by Sobol' in 1966. They are often called (t, s)-sequences in base 2 (here $s \equiv n$-dimension, $t \equiv \tau$-parameter) [3] or Sobol' sequences [1]. In our experiment, $N = 2^m$ with integer m. For such N the general estimate (4) was somewhat improved: $\delta_N = O(N^{-1}\log^{n-1} N)$.

For generating $x^{(k)}$ we applied a "*C*" subroutine published in the brochure [10]. Subroutines from [10] allow both: a fast generation of $x^{(k)}$ from its number k, as well as a superfast generation of $x^{(k)}$ from $x^{(k-1)}$ if the preceding point is available. This peculiarity is rather convenient for multiprocessor computations. The modified direction numbers in [10] are for $n \le 51$, however, they are available for $n \le 370$.

3. Small *c*

3.1. Asymptotic formula

The operator $\delta_N[f]$ is linear. Therefore,

$$\delta_N[f] = \sum_{s=1}^{n} c^s \sum_{i_1<\cdots<i_s}^{n} \delta_N\left[\left(x_{i_1} - \frac{1}{2}\right) \cdots \left(x_{i_s} - \frac{1}{2}\right)\right]. \tag{8}$$

If the Sobol sequence $x^{(k)}$ is used and $N = 2^m$, the first term in (8) can be easily computed

$$\delta_N\left[x_i - \frac{1}{2}\right] = \frac{1}{N}\sum_{k=0}^{N-1}\left(x_i^{(k)} - \frac{1}{2}\right) = \frac{1}{N}\sum_{k=0}^{N-1} x_i^{(k)} - \frac{1}{2}.$$

One-dimensional projections $x_i^{(k)}$ of $x^{(k)}$ are dyadic rational LP_0-sequences [6]. This means that for any fixed i the set $x_i^{(k)}$, $0 \le k \le N-1$, coincides with the set l/N, $0 \le l \le N-1$. So,

$$\delta_N\left[x_i - \frac{1}{2}\right] = \frac{1}{N}\sum_{l=0}^{N-1}\frac{l}{N} - \frac{1}{2} = -\frac{1}{2N}.$$

It follows from (8) that as $c \to 0$

$$\delta_N[f] = -\frac{nc}{2N} + O(c^2). \tag{9}$$

3.2. Numerical results

Fig. 1 shows computed level lines $|\delta_N| = \text{const.}$ in the (n, c) plane for three fixed values $N = 2^{15}, 2^{20}, 2^{25}$. In fact, the $\delta_N[f]$ were computed for all $c = 0.01, 0.02, \ldots, 1.00$ and all dimensions n from 1 to 60. Then an interpolation program was applied.

One can verify that at the left of Fig. 1 the curves are hyperbolas $nc = 2N|\delta_N|$ but when we are nearing to the upper right corner the curves are distorted (errors increase) and finally the interpolation fails.

3.3. On the second term in (8)

The second term in (8) depends on sums of the type

$$\delta_N\left[\left(x_i - \frac{1}{2}\right)\left(x_j - \frac{1}{2}\right)\right] = \frac{1}{N}\sum_{k=0}^{N-1}\left(x_i^{(k)} - \frac{1}{2}\right)\left(x_j^{(k)} - \frac{1}{2}\right).$$

Such sums appeared in numerical investigations of correlation properties for LP_0-sequences [8] but these experiments remained unpublished.

In [8], numbers $\Theta_{ij}(m)$ were introduced

$$\Theta_{ij}(m) = N\sum_{k=0}^{N-1}\left(x_i^{(k)} - \frac{1}{2}\right)\left(x_j^{(k)} - \frac{1}{2}\right) - \frac{1}{2}. \tag{10}$$

Numerous experiments suggest that all $\Theta_{ij}(m)$ are non-negative integers. However, this assertion is not proved.

Since all $x_i^{(0)} = 0$, $x_i^{(1)} = 1/2$, $x_i^{(2)}$ and $x_i^{(3)}$ are 1/4 or 3/4, one can easily find that all $\Theta_{ij}(1) = 0$ and $\Theta_{ij}(2)$ are either 0 or 1. The behavior of other $\Theta_{ij}(m)$ is rather unusual. For example, the numbers $\Theta_{2,3}(m)$ for $2 \le m \le 20$ are 0, 0, 1, 1, 0, 0, 1, 5, 20, 16, 1, 1, 0, 0, 1, 5, 20, 80, 321.

From (8) and (10) an improved version of asymptotic formula (9) can be obtained

$$\delta_N[f] = -\frac{nc}{2N} + \frac{n(n-1)c^2}{4N^2}\left[1 + 2\hat{\Theta}_n(m)\right] + O(c^3),$$

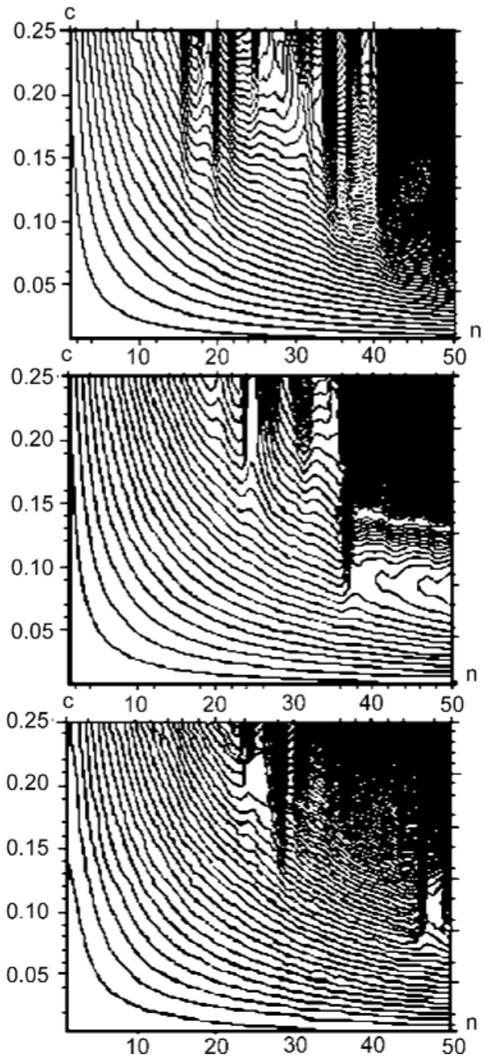

Fig. 1. Level lines $|\delta_N| = \text{const.}$ for $N = 2^{15}, 2^{20}, 2^{25}$ (downwards). Distances between the lines: 2×10^{-6}, 1×10^{-7}, 3×10^{-9}.

where $\hat{\Theta}_n(m)$ is the average of all $\Theta_{ij}(m)$, $1 \le i < j \le n$

$$\hat{\Theta}_n(m) = \frac{2}{n(n-1)} \sum_{i<j} \Theta_{ij}(m).$$

4. Large *c*

4.1. Absolute errors $|\delta_N|$ for fixed c and n

Fig. 2 contains computed values δ_N for increasing N. The scales are logarithmic; the numbers on the x-axis are $\log_2 N = m$ while the numbers on the y-axis are δ_N. A more-or-less linear disposition of points

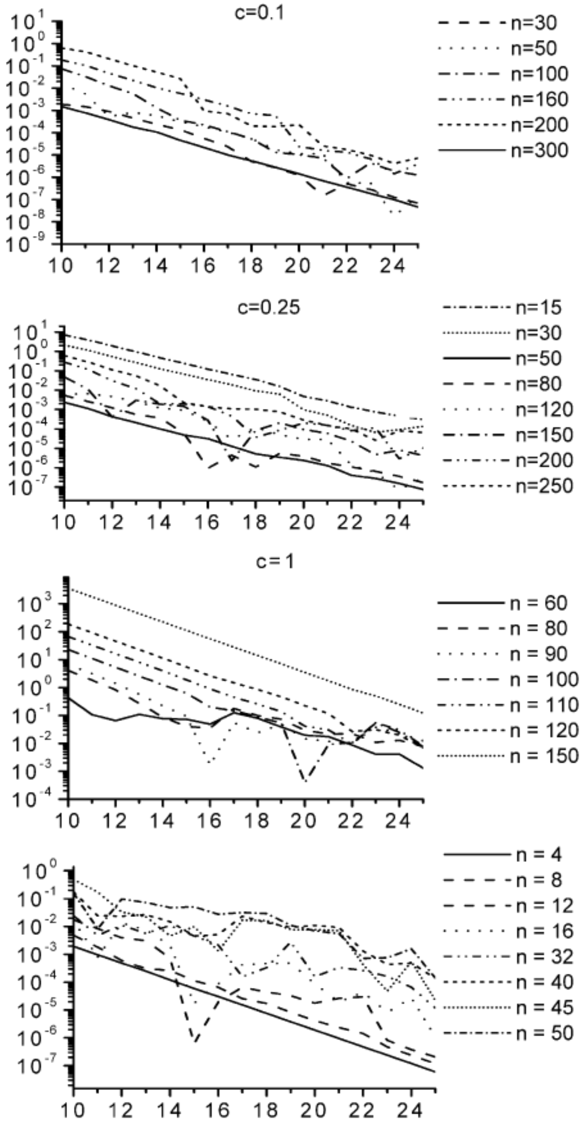

Fig. 2. Absolute errors $|\delta_N|$ for various n at c =0.1 (a), 0.25 (b) and 1.0 (c).

in these pictures means that $|\delta_N|$ behave like $N^{-\alpha}$ with some efficient α. Isolated abnormal small values of $|\delta_N|$ (e.g. at $N = 2^{16}$ and 2^{20} in c) are flukes that occur due to change of sign of δ_N.

In Fig. 2a at $c = 0.1$ all the δ_N decrease as $N^{-\alpha}$. For all dimensions up to 50 the index $\alpha \approx 1.0$, for higher dimensions $\alpha \approx 0.8$.

In Fig. 2b at $c = 0.25$ the situation is somewhat worse: $\alpha \approx 1.0$ seems true only for $n \leq 25$, while for the highest dimensions very crudely $\alpha \approx 0.6$. And another phenomenon can be seen: $|\delta_N|$ values that exceed 1 (upper left). Here the errors exceed 100% so that we can hardly talk of convergence. This phenomenon has already been explained [2,9]: one of the points $x^{(k)}$ with a small number k falls very near to the peak $(1, \ldots, 1)$ of $f(x)$.

Fig. 2c is for $c = 1$ and consists of two parts. The part containing lower dimensions $n \leq 50$ is similar to Fig. 2a and b: for dimensions $n \leq 12$ the index $\alpha \approx 1.0$; as n increases, α decreases up to $\alpha \approx 0.5$ (very crudely). For dimensions $n > 50$ the situation is much worse: there are many values of $|\delta_N|$ that exceed 1, and at $n = 150$ one needs more than a million points $x^{(k)}$ to reduce the error to less than 100%.

4.2. The role of *nc*

According to (5) QMC always outperforms MC as $N \to \infty$. But for restricted N the relations between δ_N and r_N are different.

The taxonomy below has one disadvantage: the bounds for nc depend on N (they increase together with the bounds for N). We shall assume that $N \leq 2^{25}$. Then,

(i) at $nc \leq 15$ the errors $|\delta_N|$ decrease as $1/N$;
(ii) at $15 < nc \leq 35$ the errors $|\delta_N|$ decrease faster than r_N;
(iii) at $35 < nc \leq 60$ the errors $|\delta_N|$ and r_N are of the same order of magnitude;
(iv) at $60 < nc$ the errors $|\delta_N|$ are much larger than r_N.

Examples of all these cases are given in Table 1 where the last line contains theoretical values for comparison. The situations (i)–(iii) generalize the corresponding situations for $c = 1$ in [8]. The situation (iv) is new and shows that if nc is too large QMC can be worse than MC!

One should be careful with the last assertion because a priori one cannot be sure that at these N-s the δ_N are sufficiently normal and r_N is a real probable error. Therefore, we carried out a MC computation of $J[f]$ at $c = 1$, $n = 150$ using ordinary pseudorandom numbers. Instead of the last column in Table 1 we obtained $\sqrt{N}|\delta_N| = 60, 82, 191, 167, 55, 365, 528$. These values are worse than $\sqrt{N} r_N$ but much smaller than the ones in Table 1.

Table 1
Examples of computed errors

c	0.25	0.1	0.25	0.25	1.0
n	24	120	96	200	150
nc	6	12	24	50	150
N	$N\|\delta_N\|$	$N\|\delta_N\|$	$\sqrt{N}\|\delta_N\|$	$\sqrt{N}\|\delta_N\|$	$\sqrt{N}\|\delta_N\|$
2^{16}	1.43	17.8	0.196	8.90	14,200
2^{18}	2.61	15.3	0.072	4.65	7,210
2^{20}	1.52	25.5	0.108	0.94	3,560
2^{22}	3.04	28.2	0.030	0.22	1,740
2^{24}	2.59	3.4	0.007	0.28	1,020
2^{26}	2.78	8.8	0.015	0.59	292
2^{28}	2.87	18.7	0.007	0.12	929
	$nc/2 = 3$	$nc/2 = 6$	$\sqrt{N} r_N = 0.42$	$\sqrt{N} r_N = 0.7$	$\sqrt{N} r_N = 8.4$

Table 2
N_ϵ values for various n and c at $\epsilon = 10^{-2}$ (above) and $\epsilon = 10^{-3}$

$c \setminus n$	4	8	16	32	64	128	256
1	542	1,593	15,513	35,760	8,048,559		
2^{-1}	348	589	1,654	2,802	52,090	755,476	
2^{-2}	122	292	593	1,380	2,832	33,126	1,392,346
2^{-3}	51	112	156	468	748	8,558	148,439
2^{-4}	24	50	111	156	155	3,536	33,175
2^{-5}	12	27	50	94	121	114	8,018
1	8270	23,423	369,914	3,896,430			
2^{-1}	5190	8,074	12,443	65,246	223,8931	16,812,290	
2^{-2}	1909	3,418	7,118	17,294	81,502	435,910	26,735,007
2^{-3}	866	1,702	3,038	6,962	21,995	91,354	1,075,315
2^{-4}	394	632	1,654	2,612	1,630	23,095	547,582
2^{-5}	218	318	592	1,022	950	5,946	72,488

Computations suggest that for small nc, in fact for $nc \leq 6$, limit $\lim_{N\to\infty} N\delta_N[f]$ exists. We have no proof for this assertion but it is very likely that if this limit exists it is equal to $-nc/2$; cf. (9).

4.2.1. Remark

In our text the characteristic parameter nc appeared in the range of the integrand at small c: $\sup f(x) - \inf f(x) \approx \exp(nc/2)$. The total variation of $f(x)$ could be considered instead: $V(f) = (1 + (3/2)c)^n - 1 \approx \exp((3/2)nc)$.

4.3. Number of points required

Consider a fixed positive ϵ and let N_ϵ be the smallest value N' having the property that $|\delta_N| \leq \epsilon$ for all $N \geq N'$. Table 2 contains N_ϵ values for different n and c. The upper table is for $\epsilon = 10^{-2}$, the lower one—for $\epsilon = 10^{-3}$. Empty boxes indicate that N_ϵ has not been reached in our experiments.

One can notice that N_ϵ values that correspond to n and c satisfying $nc =$ const. are of the same order if this constant is sufficiently small. This result is related to the hyperbolas in Fig. 1.

5. Conclusions

The widespread belief that QMC integrations are expedient at moderate dimensions n is one-sided: in reality, the efficiency of QMC depends on both the dimensions n and the structure of integrand, in our example, parameter c. If the product nc is small, QMC integrations are efficient even at very high dimensions. For large nc, QMC approximations converge but the required amount N of trials is impractically large.

We are sure that the convergence of QMC approximations must be monitored while $N = 2^m, 2^{m+1}, 2^{m+2}, \ldots$. Sometimes additional control methods can be applied: MC estimation of the same problem and/or computation of similar problems with known exact solutions.

References

[1] M. Drmota, R.F. Tichy, Sequences, Discrepancies and Applications, Lecture Notes on Mathematics, Springer, Berlin, 1997, N. 1651.

[2] Yu.L. Levitan, N.I. Markovich, S.G. Rosin, I.M. Sobol’, On quasi-random sequences for numerical computations, USSR Comput. Maths Math. Phys. 28 (1988) 86–92.

[3] H. Niederreiter, Random Number Generation and Quasi-Monte Carlo Methods, SIAM, Philadelphia, 1992.

[4] S.H. Paskov, J.F. Traub, Faster valuation of financial derivatives, J. Portfolio Manage. (1995) 113–120.

[5] H. Rabitz, O.E. Alis, J. Shorter, K. Shim, Efficient input–output model representation, Comput. Phys. Commun. 117 (1–2) (1999) 11–20.

[6] I.M. Sobol’, On the distribution of points in a cube and the approximate evaluation of integrals, USSR Comput. Maths Math. Phys. 7 (1967) 86–112.

[7] I.M. Sobol’, Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, Math. Comput. Simulation 55 (1–3) (2001) 271–280.

[8] I.M. Sobol’, Yu.L. Levitan, Experiments in computing integrals of high multiplicity, Prob. Comput. Appl. Math., Tashkent 51 (1978) 138–145 (in Russian).

[9] I.M. Sobol’, B.V. Shukhman, Integration with quasirandom sequences: numerical experience, Int. J. Modern Phys. C 1995, 6, N 2, 263–275.

[10] I.M. Sobol’, V.I. Turchaninov, Yu.L. Levitan, B.V. Shukhman, Quasirandom Sequence Generators, Keldysh Institute of Applied Mathematics, Moscow, 1992, 22 pp.