Journal scientific and applied research, vol. 12, 2017
Association Scientific and Applied Research
International Journal

Original Contribution ISSN 1314-6289

A COMPARISON OF QUASI MONTE CARLO METHODS BASED ON FAURE AND SOBOL SEQUENCES FOR COMPUTATION OF MULTIDIMENSIONAL INTEGRALS

Venelin Todorov[1], Valeri Dzurov[2], Valentin Dimitrov[2]

1 IICT, BULGARIAN ACADEMY OF SCIENCES
E-mail: venelinltodorov@gmail.bg,

2 ROUSSE UNIVERSITY "ANGEL KANCHEV"
E-mail: vdzhurov@yahoo.com

ABSTRACT: *In this paper we implement and analyze the performance of Faure quasi-random sequence. We compare the results with the Sobol quasi-random sequence which is the most widely used quasi-Monte Carlo method. Also some experiments between Faure sequence and the plain (Crude) Monte Carlo method are given. We consider a case study with a non-smooth integrand function. We show that the Sobol sequence has some advantageous over the Faure sequence.*

KEYWORDS: *Quasi-Monte Carlo algorithms, multidimensional integrals, Faure sequence, Sobol sequence, Crude Monte Carlo algorithm, applications.*

Introduction

High dimensional integrals are usually solved with Monte Carlo algorithms and quasi Monte Carlo algorithms. Monte Carlo method is the only viable method for high-dimensional problems since its convergence is independent of the dimension. Monte Carlo methods give statistical estimates for the functional of the solution by performing random sampling of a certain random variable whose mathematical expectation is the desired functional. Monte Carlo methods are methods of approximation of the solution to problems of computational mathematics, by using random processes for each such problem, with the parameters of the process equal to the solution of the problem. The method can guarantee that the error of Monte Carlo approximation is smaller than a given value with a certain probability [5]. The most important advantage of the Monte Carlo methods is that they are suitable for solving multi-dimensional problems, since the computational complexity increases linearly and not exponentially with the dimensionality [5,14]. The MC method is a widely used tool in many fields of science.

In the last few years new approaches have been developed that outperform standard Monte Carlo in terms of numerical efficiency. It has been found that there can be efficiency gains in using deterministic sequences rather than the random sequences which are a feature of standard Monte Carlo. These deterministic sequences are carefully selected so that they are well dispersed throughout the region of integration. Sequences with this property are known as low discrepancy sequences. These sequences are often more efficient than standard Monte Carlo in evaluating high dimensional integrals if the integrand is sufficiently regular and for many finance applications this is the case. However it is interesting to consider multidimensional integrals of non-smooth function. Such integrals can be used to describe problems in ecology (Genz test functions) or stochastic tomography, where multidimensional integrals are used for statistical measures such as the highest posterior density regions and migration areas.

As we already mention the multidimensional numerical quadratures are of great importance in many practical areas. The Monte Carlo method is known to be only accurate with a tremendous amount of scenarios since its rate of convergence is $O(N^{-1/2})$. Quasi Monte Carlo methods use deterministic sequences that have better uniform properties measured by discrepancy. They are usually superior to the Monte Carlo method as they have a convergence rate of $((\log N)^d/N)$, where N is the number of samples and d is the dimensionality of the problem under consideration.

Quasi Monte Carlo algorithms for numerical integration

Consider the problem of approximate integration of the multiple integral:

$$\int_{[0,1)^d} f(\mathrm{x})d\mathrm{x} = \int_0^1 d\mathrm{x}^{(1)} \int_0^1 d\mathrm{x}^{(2)} \dots \int_0^1 d\mathrm{x}^{(\mathrm{d})} f(x^{(1)}, x^{(2)}, \dots, x^{(d)}),$$

where $x = (x^{(1)}, \dots, x^{(d)}) \in [0,1)^d$.

For small values of d, numerical integration methods such as Simpson's rule or the trapezoidal rule can be used to approximate the integral. These methods, however, suffer from the so-called curse of dimensionality and become impractical as d increases.

The crude Monte Carlo method has rate of convergence $O(N^{-1/2})$ which is independent of the dimension of the integral, and that is why Monte Carlo integration is the only practical method for many high-dimensional problems.

Much of the efforts to improve Monte Carlo are in construction of variance reduction methods which speed up the computation or to use quasi-random sequences [1]. A quasi-random or low discrepancy sequence, such as the Faure, Halton, Hammersley, Niederreiter or Sobol sequences, is ”less random” than a pseudorandom number sequence, but more useful for such tasks as approximation of integrals in higher dimensions, and in global optimization. This is because low discrepancy sequences tend to sample space ”more uniformly” than random numbers. It is a question of interest to know which sequence outperforms the other. In this study we implement the Faure sequence and make a comparison with the Sobol sequence. The Faure sequences [8] are a digital (0, s)-sequence over F_b with b denoting a prime (original case) or a prime power (general case) greater or equal to s. The s infinite generator matrices $\boldsymbol{C}^{(1)},\ldots,\boldsymbol{C}^{(s)}$ over $\mathbf{F}_b$ are defined by $\boldsymbol{C}^{(i)} = (c_{jr}^{(i)})_{j,\, r \geq 0}$ with

$$c_{jr}^{(i)} = \binom{r}{j} \alpha_i^{r-j},$$

where $\alpha_1,\ldots, \alpha_s$ denote s distinct elements from $\mathbf{F}_b$ and the conventions $\alpha^0 = 1$ for all $\alpha \in \mathbf{F}_b$ and $\binom{r}{j} = 0,\ j > r.$

For $\alpha = 1$, the resulting matrix is the infinite Pascal matrix modulo the characteristic of $\mathbf{F}_b$; for $\alpha = 0$, it is the infinite identity matrix. If $s = 1$ and $\alpha_1 = 0$, the resulting (0, 1)-sequence is identical to the van der Corput sequence in the same base [17]. The algorithm for the Faure sequence follows the method of Henri Faure in [8] for computing quasi-random numbers. It is a merging and adaptation of the routines INFAUR and GOFAUR from ACM TOMS 647. We use of persistent variables to improve the MATLAB implementation. The parameters of the Faure algorithm are described below. The input is an integer DIM_NUM, the spatial dimension, which should be at least 2. The other parameter is integer SEED, which is the seed, that indicates the index of the element of the sequence to be calculated. If SEED is negative, it is effectively replaced by a more suitable value. The output is a real QUASI(DIM_NUM), the next quasi-random vector. For the output the appropriate value of SEED have to be used on the next call, if the next element of the sequence is desired. For the Sobol sequence we use an implementation that is an adapation of the INSOBL and GOSOBL routines in ACM TOMS Algorithm 647 [9] and ACM TOMS Algorithm [1,2,3,4]. The original code can only compute the ”next” element of the sequence [11]. The revised code allows the user to specify the index of the desired element. The algorithm has a maximum spatial dimension of 40 since MATLAB doesn’t support 64 bit integers. A remark by Joe and Kuo [10] shows how to extend the algorithm from the original maximum spatial dimension of 40 up to a maximum spatial dimension of 1111. The FORTRAN90 and C++ versions of the code has been updated in this way [12,15], but updating the

MATLAB code has not been simple, since MATLAB doesn't support 64 bit integers. We use algorithm that generates a new quasi-random Sobol vector with each call. The routine adapts the ideas of Antonov and Saleev [1,13] . The parameters of the algorithm are an integer *DIMNUM* , the number of spatial dimensions. The algorithm starts with integer *SEED*, the "seed" for the sequence. This is essentially the index in the sequence of the quasi-random value to be generated. On output, *SEED* has been set to the appropriate next value, usually simply *SEED* + 1. If *SEED* is less than 0 on input, it is treated as though it were 0. An input value of 0 requests the first (0-th) element of the sequence. Output is the real *QUASI*(*DIMNUM*), the next quasi-random vector [5].

Numerical example and results

Table 1: The relative error for 4 dimensional integral of a non-smooth function with Faure and Sobol QMC

N	Faure	Time,s	Sobol	Time,s
100	5.05e-2	0.03	5.67e-3	0.02
1000	8.84e-3	0.43	8.91e-4	0.21
10000	2.10e-3	3.93	5.51e-4	1.78
100000	4.97e-4	42.3	5.79e-5	18.6
10000000	1.09e-4	200	7.70e-6	119

Table 2: The relative error for 4 dimensional integral of a non-smooth function with Faure QMC and Crude MC

N	Faure	Time,s	Crude	Time,s
100	5.05e-2	0.03	2.82e-2	0.003
1000	8.84e-3	0.43	1.47e-2	0.01
10000	2.10e-3	3.93	8.12e-3	0.11
100000	4.97e-4	42.3	1.30e-3	1.36
10000000	1.09e-4	200	7.35e-4	13.08

Table 3: The computational time for 4 dimensional integral of a non-smooth function with the three methods

time in seconds	Faure	Sobol	Crude
0.1	1.82e-2	1.86e-3	8.12e-3
1	5.26e-3	7.15e-4	2.31e-3
5	1.86e-3	2.03e-4	9.81e-4
10	8.22e-4	9.75e-5	7.67e-4
60	3.90e-4	8.91e-6	5.11e-4

We will test Faure and Sobol sequence for evaluating the following multidimensional integral of non-smooth integrand function, taken from the paper of Ivan Dimov and Rayna Georgieva [6]:

$$f_1(x_1, x_2, x_3, x_4) = \sum_{i=1}^{4} |(x_i - 0.8)^{-1/3}|,$$

where the integration is over the unit 4-dimensional hypercube. The referent value of the integral is 7.22261 [6,7].

As can been seen from the table for 4 dimensional integral, the low discrepancy sequence of Sobol produces more rapid convergence, and lower errors, than the Faure sequence and the pseudorandom sequence. This is as anticipated since the pseudo randomly obtained averages converge at the rate O(N−1/2),while the quasi randomly obtained averages converge at a rate closer to O(N−1). As expected the Sobol sequence gives better results than the Faure sequence. It is interesting to see that for a given number of samples the Faure sequence gives better relative error than the Crude Monte Carlo algorithm – see Table 2, but the latter is between 10 and 20 times faster. It is interesting that for a preliminary given time up to 10s the Crude MC gives a little bit better results than the Faure sequence – see Table 3, but for 60s the Faure sequence starts producing better results. This means that Faure sequence has advantage over Crude MC only or large number of samples and for higher computational time. For a preliminary given time of only 1s Sobol sequence gives relative error of 7.15e-4 and Faure sequence gives relative error of 5.26e-3 which is already a sufficient accuracy. Definitely for larger number of points and for a preliminary given time, the advantage of Sobol quasi-random sequence in place of both Crude MC and Faure QMC should become even more pronounced – see Table 1 and Table 3. For this example of non-smooth integrand function even the first derivative does not exist. Such kind of applications appears also in some important problems in financial mathematics and stochastic tomography. It should be mentioned that in the case of a smooth integrand functions the results with the methods under consideration are even more precise. In the future a comparison with other sequences like Halton sequence will be done.

Conclusion

In this paper we analyze the performance of plain Monte Carlo algorithm and two quasi Monte Carlo methods for multidimensional integrals. The Sobol quasi-random sequence is compared with the Faure sequence and the results are very precise for the multidimensional integrals under consideration, which shows the strength of the presented algorithm for relatively low dimensions. We consider an example of non-smooth integrand function which is the more complicated case compared to examples of smooth integrand functions.

For this particular example the stochastic algorithms under consideration produce reliable results. While Sobol sequence is a lot better than the pseudorandom sequence, the Faure sequence gives similar results to the Crude Monte Carlo algorithm. The experiments show that for a preliminary given time the Crude Monte Carlo algorithm outperform the Faure sequence, but for a given number of samples the Faure sequence gives better results than the pseudorandom sequence. The multidimensional integral under consideration is with non-smooth integrand which is more difficult case than the smooth integrand function. This multidimensional integral can be applied to various problems where data is taken in randomized way like stochastic tomography connected with people migration. International migration is a topic that is attracting a significant level of interest in current political debate and is high on the agenda for policy makers in central and local government [16]. One area of debate is the impact of student migration, for example, on net migration. Stochastic methods under consideration are an efficient way to solve problems like the mentioned above that are described with multidimensional integrals of both smooth and non-smooth functions. It will be interesting to compare Faure sequence with Halton and Niederreiter sequences, but this will be an object of a future study.

Acknowledgement

This work was supported by Program for career development of the Young scientists, BAS, Grant No. DFNP-91/04.05.2016 as well as the Bulgarian National Science Fund under Grant DFNI I02-20/2014.

References

[1]. I.A. Antonov, V.M. Saleev, An Economic Method of Computing LP Tau-Sequences, USSR Computational Mathematics and Mathematical Physics, Volume 19, 1980, pages 252-256.

[2]. Paul Bratley, Bennett Fox, Algorithm 659: Implementing Sobol's Quasirandom Sequence Generator, ACM Transactions on Mathematical Software, Volume 14, Number 1, March 1988, pages 88-100.

[3]. Paul Bratley, Bennett Fox, Harald Niederreiter, Implementation and Tests of Low Discrepancy Sequences, ACM Transactions on Modeling and Computer Simulation, Volume 2, Number 3, July 1992, pages 195-213.

[4]. Paul Bratley, Bennett Fox, Linus Schrage, A Guide to Simulation, Second Edition, Springer, 1987, ISBN:0387964673, LC: QA76.9.C65.B73.

[5]. I. Dimov, Monte Carlo Methods for Applied Scientists, New Jersey, London, Singapore, World Scientific, 2008, 291p.

[6]. I. T. Dimov, R. Georgieva, Tz. Ostromsky, Z. Zlatev. Advanced Algorithms for Multidimensional Sensitivity Studies of Large-scale Air Pollution Models based on Sobol Sequences. Special issue of *Computers*

and Mathematics with Applications 65 (3), “Efficient Numerical Methods for Scientific Applications”. Elsevier, 2013, 338 – 351. ISSN: 0898-1221
[7]. I. T. Dimov, R. Georgieva. Multidimensional Sensitivity Analysis of Large-scale Mathematical Models. O.P. Iliev et al. (eds.), Numerical Solution of Partial Differential Equations: Theory, Algorithms, and Their Applications, Springer Proceedings in Mathematics & Statistics 45, Springer Science Business Media, New York, 2013, 137 – 156.
[8]. Henri Faure, Discrepance de suites associees a un systeme de numeration (en dimension s), Acta Arithmetica, Volume 41, 1982, pages 337-351.
[9]. Bennett Fox, Algorithm 647: Implementation and Relative Efficiency of Quasirandom Sequence Generators, ACM Transactions on Mathematical Software,Volume 12, Number 4, pages 362-376, 1986.
[10]. Stephen Joe, Frances Kuo, Remark on Algorithm 659: Implementing Sobol's Quasirandom Sequence Generator, ACM Transactions on Mathematical Software, Volume 29, Number 1, March 2003, pages 49-57.
[11]. Harald Niederreiter, Random Number Generation and quasi-Monte Carlo Methods, SIAM, 1992, ISBN13: 978-0-898712-95-7, LC: QA298.N54.
[12]. William Press, Brian Flannery, Saul Teukolsky, William Vetterling, Numerical Recipes in FORTRAN: The Art of Scientific Computing, Second Edition, Cambridge University Press, 1992, ISBN: 0-521-43064-X, LC: QA297.N866.
[13]. Ilya Sobol, Uniformly Distributed Sequences with an Additional Uniform Property, USSR Computational Mathematics and Mathematical Physics,
[14]. I. Sobol, D. Asotsky, A. Kreinin , S. Kucherenko. Construction and Comparison of High-Dimensional Sobol’ Generators, 2011, Wilmott Journal, Nov, pp. 64-79 Volume 16, 1977, pages 236-242.
[15]. Ilya Sobol, YL Levitan, The Production of Points Uniformly Distributed in a Multidimensional Cube (in Russian), Preprint IPM Akademii Nauk SSSR, Number 40, Moscow 1976.
[16]. http://dera.ioe.ac.uk/25342/1/internationalstudentmigration_tcm77-431150.pdf Population briefing. International student migration. What do the statistics tell us?
[17]. http://mint.sbg.ac.at/desc_SFaure-ori.html#bib-SFaure-3