Journal scientific and applied research, vol. 20, 2021
International Journal

Original Contribution

ISSN 1314-6289

A COMPARISON OF SEVERAL STOCHASTIC TECHNIQUES FOR COMPUTATION OF MULTIDIMENSIONAL INTEGRALS

Venelin Todorov[1,2], Valerij Dzhurov[3], Ilian Tzvetkov[3]

[1]*INSTITUTE OF MATHEMATICS AND INFORMATICS, BULGARIAN ACADEMY OF SCIENCES, DEPARTMENT OF INFORMATION MODELING, ACAD. GEORGIBONCHEV STR., BLOCK 8, 1113 SOFIA, BULGARIA*

[2]*INSTITUTE OF INFORMATION AND COMMUNICATION TECHNOLOGIES, BULGARIAN ACADEMY OF SCIENCES, DEPARTMENT OF PARALLEL ALGORITHMS, ACAD. G. BONCHEV STR., BLOCK 25 A, 1113, SOFIA, BULGARIA*

[3]*ROUSSE UNIVERSITY ”ANGEL KANCHEV”*

E-mail: vtodorov@math.bas.bg, venelin@parallel.bas.bg

ABSTRACT: *A comprehensive experimental study based on Sobol sequence, Importance Sampling and Fibonacci based lattice rule has been done. This is the first time the Sobol sequence has been compared with the Importance sampling method for the problem under consideration. The numerical tests show that the stochastic algorithms under consideration are efficient tool for computing multidimensional integrals. In order to obtain a more accurate and reliable interpretation of the results this is very important.*

KEYWORDS: *Monte Carlo methods, multidimensional integrals, Sobol sequence, Fibonacci lattice rule, Importance sampling algorithm.*

Introduction

High dimensional integrals are usually solved with Monte Carlo algorithms. Monte Carlo method is the only possible 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 [24] 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 [8].

For small values of the dimensionality of the integral d, numerical integration methods such as Simpson's rule or the trapezoidal rule (see Davis and Rabinowitz [19]) can be used to approximate the integral. These methods, however, suffer from the so-called curse of dimensionality and be-come impractical as s increases beyond 3 or 4. The Crude Monte Carlo method has rate of convergence $O(N^{-1/2})$, where N is the number of samples, 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 especially in air pollution modelling, quantum mechanics, Bayesian statistics and international migration forecasting, see [5,6,7,9,10,11,16,17,18].

Importance Sampling

Importance sampling involves multiplying the integrand by 1 to yield an expectation of a quantity that varies less than the original integrand over the region of integration. For example, let h(x) be a density for the random variable X [21]. All we need to do to have a Monte Carlo estimator with zero variance is use and make sure that our density h is proportional to the function g. The ability to simulate independent random variables from h(x), or the ability to compute the density h(x), itself, implies that the normalizing constant of the distribution is computable, which in turn would imply that the original integral involving g(x) is computable. While h(x) might be roughly the same shape as g(x), serious difficulties arise if h(x) gets small much faster than g(x) out in the tails. In such a case, though it is improbable (by definition) that you will realize a value x_i from the far tails of h(x), if you do, then your Monte Carlo estimator will take $g(x_i)=h(x_i)$ for such an improbable x_i may be orders of magnitude larger than the typical values g(x)=h(x) that you see [23].

Quasi Monte Carlo algorithm based on Sobol sequence

For the Sobol sequence we use an implementation that is an adaptation of the INSOBL and GOSOBL routines in ACMTOMS Algorithm 647 and ACMTOMS 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 [25].

The algorithm has a maximum spatial dimension of 40 since MATLAB doesn't support 64-bit integers. Aremark by Joe and Kuoshow show to extend the algorithm from the original maximum spatial dimension of 40 up to a maximum spatial dimension of 1111 [25]. The FORTRAN 90 and C++ versions of the code has been updated in this way [12,14,15], 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 DIM_{NUM}, 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 [1,2].

Fibonacci based lattice rule

The monographs of Sloan and Kachoyan [20] and Wang and Hickernell [22] provide comprehensive expositions of the theory of integration lattices.

Let n be an integer, and a=($a_1,...,a_s$) be an integer vector modulo n. A set of the form [15]

$$P_n = \left\{ \left\{ \frac{ak}{n} \right\} = \left(\left\{ \frac{a_1 k}{n} \right\}, \ldots, \left\{ \frac{a_s k}{n} \right\} \right) \mid k = 1, \ldots, n \right\} \quad (1)$$

is called a lattice point set, where {x} denotes the fractional part of x.The vector a is called a lattice point or generator of the set. As one can see, the formula for the lattice point set is simple to program. The difficulty lies in finding a good value of a, such that the points in the set are evenly spread over the unit cube. The choice of goodgenerating vector, which leads to small errors, is not trivial. Complicated methods from theory of numbers have been used. We consider the following generating vector based on generalized Fibonacii numbers of corresponding dimensionality:

$$a = (1, F^{(s)}_{l+1}, ..., F^{(s)}_{l+s-1})), \quad n_l = F^{(s)}_l, \quad (2)$$

where

$$F^{(s)}_{l+s} = F^{(s)}_l + F^{(s)}_{l+1} + ... + F^{(s)}_{l+s-1}, l = 0, 1, \ldots \quad (3)$$

with initial conditions

$$F^{(s)}_0 = F^{(s)}_1 = \ldots = F^{(s)}_{s-2} = 0, F^{(s)}_{s-1} = 1, \quad (4)$$

For l=0,1,…

The discrepancy of the set obtained by using the vector described above is asymptotically estimated in [20,22].

The number of calculation required to obtain the generating vector is O(ln n_l). The generation of a new point requires constant number of operations, thus to obtain a lattice set of the described kind consisting of n_l points, O (lnn_l) number of operations are necessary.

Numerical example

We want to compute the following 7 dimensional integral:

$$\int_{[0,1]^7} e^{1-\sum_{i=1}^{3} sin(\frac{\pi}{2}.x_i)}.arcsin(sin(1)+\frac{\sum_{j=1}^{7} x_j}{200}) \approx 0.75151101. \quad (5)$$

We will use the IS algorithm with probability density function

$$f(\overline{x}) = \frac{1}{\eta} e^{1-\sum_{i=1}^{3} sin(\frac{\pi}{2}.x_i)}. \quad (6)$$

The value of η must be found separately. It is equal to the value of the integral

$$\int_{[0,1]^3} e^{1-\sum_{i=1}^{3} sin(\frac{\pi}{2}.x_i)}.$$

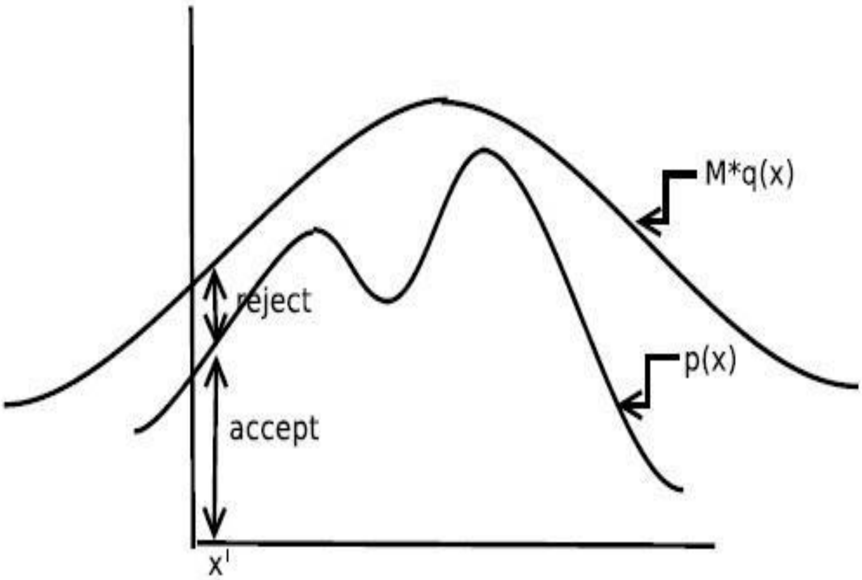

Figure 1. The acceptance-rejection method (the method of selection)

We evaluate the last integral with Crude Monte Carlo method for a number of samples N = 10^5. After that we use the method of selection (the acceptance-rejection method). The idea of the method is given by the Figure 1.

We make a comparison between the quasi-random Sobol sequence (SOBOL), the Importance sampling (IS) and the Fibonacci based lattice rule (FIBO) which is the goal of this paper. As we expected SOBOL and FIBO outperforms IS by far. The results are given in the Tables 1 and 2 below. Each table contains information about the stochastic approach which is applied, the obtained relative error, the needed CPU time and the number of points. Table 1 shows the relative error for a given number of samples, while in Table 2 is presented the relative error for a fixed computational time, which is a measure of the computational complexity of the algorithms.

It can be seen than the CPU time hypercube sampling and FIBO method is the smallest, while Importance sampling method needs much higher time. Also, we see that for a large number of samples SOBOL and FIBO outperforms the importance sampling.

N	FIBO	time	SOBOL	time	IS	time
10^3	1.03e-3	0.15	2.27e-4	0.76	6.61e-3	3.1
$5x10^3$	2.71e-2	0.75	2.13e-4	3.71	3.35e-3	15.8
10^4	3.34e-4	1.52	1.22e-4	7.75	1.09e-3	30
$25x10^3$	2.73e-4	3.51	8.51e-5	27	7.51e-4	74
10^5	1.62e-5	13.8	4.71e-5	72	5.41e-4	315
10^6	1.02e-6	104	9.45e-6	697	2.53e-4	3056

Table1. Relative error for 7-dimensional integral for a given number of realizations of the random variable

time, sec.	FIBO	SOBOL	IS
0.1	1.38e-3	1.85e-3	5.51e-2
1	2.87e-4	1.85e-4	2.31e-2
5	1.16e-4	9.79e-5	8.05e-3
10	5.28e-5	8.36e-5	4.91e-3
20	2.26e-5	5.46e-5	2.58e-3
100	1.61e-6	2.21e-5	7.18e-4

Table2. Relative error for 7-dimensional integral for a preliminary given time in seconds

In Table 1 are presented the relative error for the 7-dimensional integrals with Fibonacci lattice sequence (FIBO), Sobol quasi-random sequence (SOBOL) and Importance sampling (IS) for a fixed number of points and for a fixed computational time in Table 2. Obviously FIBO and SOBOL has the lowest computational complexity and are the fastest algorithm, while the Importance sampling is much slower. As can be seen from the results for 7-dimensional integral, the Importance sampling produces the worst results. It is interesting to see that FIBO and SOBOL gives errors of the same magnitude - see Table 1, but the Fibonacci lattice sequence has the advantage being faster for a preliminary given time in seconds - see Table 2. For large dimensions of the integral the Sobol sequence outperforms the Fibonacci based lattice rule, but this will be an object of a future study.

Conclusion

In this paper we analyze the performance of Sobol quasi random sequence, Fibonacci based lattice rule and Importance sampling algorithm for computation of multidimensional integrals. Stochastic methods under consideration are an efficient way to solve problems in forecasting international migration. The problem of accurate evaluation of multidimensional integrals is discussed. This is the first time a particular 1-rank lattice rule based on Fibonacci generating vector is compared with the importance sampling technique. We make a comparison with the Sobol approach and it gives closer results to the Fibonacci method. It is a crucial element since this may be important for improving the results in important areas for control of migration flows. This paper is funded by “Scientific Research” subsidy by University of Ruse 2020.

Acknowledgement

The author Venelin Todorov is supported by the Bulgarian National Fund of Science under Project KP-06-M32/2–17.12.2019 "Advanced Stochastic and Deterministic Approaches for Large-Scale Problems of Computational Mathematics" and Project DN 12/5-2017, "Efficient Stochastic Methods and Algorithms for Large-Scale Problems".

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]. P. 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]. P. Bratley, Bennett Fox, Linus Schrage, A Guide to Simulation, Second Edition, Springer, 1987, ISBN: 0387964673, LC: QA76.9.C65.B73.
[4]. 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.
[5]. L.A. Brown, (1970). On the use of Markov chains in movement research. Economic Geography, 46 (Suppl.), 393–403.
[6]. H. Brücker and B. Siliverstovs (2005). On the Estimation and Forecasting of International Migration: How Relevant is Heterogeneity Across Countries. IZA Discussion Paper 170. Bonn: InstitutzurZukunft der Arbeit.
[7]. P. Congdon, 2003, Applied Bayesian Modelling. Chichester: John Wiley.
[8]. I. Dimov, Monte Carlo Methods for Applied Scientists, New Jersey, London, Singapore, World Scientific, 2008, 291p.
[9]. Dimov, I. Chernogorova, T., Vulkov, L. Coordinate Transformation Approach for Numerical Solution of Environmental Problems. Mathematical Problems in Meteorological Modelling, series Mathematics in Industry, 24, Springer International Publishing, 2016, ISBN:978-3-319-40155-3, ISSN:1612-3956, DOI:10.1007/978-3-319-40157-7_7, 117-127
[10]. Dimov, I., M. Nedjalkov, J. M. Sellier, S. Selberherr. Neumann Series Analysis of the Wigner Equation Solution. Progress in Industrial Mathematics at ECMI 2014, Mathematics in Industry, 22, Springer International Publishing, G. Russo, V. Capasso, G. Nicosia, V. Romano (ed), 2016, ISBN:978-3-319-23412-0, DOI:10.1007/978-3-319-23413-7_97, 701-707
[11]. Dimov, I., Zlatev, Z., Georgiev, K., Margenov, S.. Numerical algorithms for scientific and engineering applications. Journal of Computational and Applied Mathematics, 310, Elsevier, 2017, ISSN:0377-0427, 1-4.
[12]. Henri Faure, Discrepance de suites associees a un systeme de numeration (en dimension s), ActaArithmetica, Vol. 41, 1982, 337-351.
[13]. John Hammersley, Monte Carlo methods for solving multivariable problems, Proceedings of the New York Academy of Science, Volume 86, 1960, 844-874.
[14]. J.M. Hammersley, D.C. Handscomb, (1964). Monte Carlo Methods.
[15]. L.K. Hua, Y. Wang: Applications of Number Theory to Numerical analysis, 1981.

[16]. J.A. Hoeting, D. Madigan, A. E. Raftery, and C. T. Volinsky (1999). Bayesian Model Averaging: A Tutorial. Statistical Science, 14 (4), 382–417.
[17]. Lin S., “Algebraic Methods for Evaluating Integrals in Bayesian Statistics,” UC Berkeley, May 2011.
[18]. McKay, M.D., Beckman, R.J., Conover, W.J.: A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics 21(2), 23945 (1979)
[19]. Rabinowitz. P. and Davis P.: Methods of Numerical Integration. Academic Press, London, 2nd edition, (1984).
[20]. I.H. Sloanand P.J. Kachoyan, Lattice methods for multiple integration: Theory, error analysis and examples, (1987), J. Numer. Anal., 24, 116-128.
[21]. Van der Corput, J.G.: Verteilungsfunktionen I-VIII. Proc. Akad. Amsterdam, Vol. 38 (1935) 813-821, 1058-1066, Vol.39 (1936) 10-19, 19-26, 149-153, 339-344.
[22]. Y. Wang and F.J. Hickernell, An historical overview of lattice point sets, (2002).
[23]. Lectures: Monte Carlo methods and Importance sampling: http://ib.berkeley.edu/labs/slatkin/eriq/classes/guest_lect/mc_lecture_notes.pdf
[24]. https://www.slideshare.net/guisaga/monte-carlo-simulation-paul-wilmott
[25]. Sobol sequence generator, 2010-09-16. Retrieved 2013-12-20.