

Article

# Monte Carlo Methods and the Koksma-Hlawka Inequality

Sergey Ermakov \*  and Svetlana Leora

The Faculty of Mathematics and Mechanics, St. Petersburg State University, 199034 St. Petersburg, Russia

\* Correspondence: sergej.ermakov@gmail.com

Received: 30 June 2019; Accepted: 7 August 2019; Published: 9 August 2019



**Abstract:** The solution of a wide class of applied problems can be represented as an integral over the trajectories of a random process. The process is usually modeled with the Monte Carlo method and the integral is estimated as the average value of a certain function on the trajectories of this process. Solving this problem with acceptable accuracy usually requires modeling a very large number of trajectories; therefore development of methods to improve the accuracy of such algorithms is extremely important. The paper discusses Monte Carlo method modifications that use some classical results of the theory of cubature formulas (quasi-random methods). A new approach to the derivation of the well known Koksma-Hlawka inequality is pointed out. It is shown that for high ( $s > 5$ ) dimensions of the integral, the asymptotic decrease of the error comparable to the asymptotic behavior of the Monte Carlo method, can be achieved only for a very large number of nodes  $N$ . It is shown that a special criterion can serve as a correct characteristic of the error decrease (average order of the error decrease). Using this criterion, it is possible to analyze the error for reasonable values of  $N$  and to compare various quasi-random sequences. Several numerical examples are given. Obtained results make it possible to formulate recommendations on the correct use of the quasi-random numbers when calculating integrals over the trajectories of random processes.

**Keywords:** Monte Carlo method; quasi-Monte Carlo method; Koksma-Hlawka inequality; quasi-random sequences; stochastic processes

## 1. Introduction

Let  $\zeta(t, \omega)$  be a random process,  $t_1 < t_2 < \dots < t_n$  are given time moments, and  $F_n(t_1, x_1, \dots, t_n, x_n) = P(\zeta(t_1, \omega) \leq x_1; \dots; \zeta(t_n, \omega) \leq x_n)$  are finite-dimensional distributions of the process. The Monte Carlo method of modeling a process  $\zeta(t, \omega)$  usually consists of modeling its finite-dimensional distributions. Extensive literature is devoted to the algorithms of such modeling (for example, [1] and some cited sources). The methods developed in these papers allow us to express realizations of a random vector  $\Xi = P(\zeta(t_1, \omega); \dots; \zeta(t_n, \omega))$  through a certain set of randomizations of a random variable  $(\alpha_1, \dots, \alpha_M)$  uniformly distributed on  $[0, 1]$ , where  $M \geq n$  and, generally speaking, these can be random.

Ultimately, the problem solution  $\Phi$  is represented as the expectation of a certain function  $\Psi$  of a random vector  $\Xi$  for selected finite values  $n$  and  $M$ . That means  $\Phi = E\Psi(\Xi)$ , and the computational process consists in multiple ( $N$  times) calculations of the independent realizations  $\Xi^j$  of the vector  $\Xi$  and in estimation of  $\Phi$  using the arithmetic mean  $E\Psi(\Xi) \approx 1/N \sum_{j=1}^N \Psi(\Xi^j)$ . Let us notice that the value  $M = M_j$  may depend on the number of the realization  $j$ . The maximum value  $\bar{M} = \max_j M_j$  is called a constructive dimension of the algorithm. Finally, recall that  $\Xi^j$  is expressed in terms of realizations of

uniformly distributed numbers, i.e.,  $\Phi$ , the integral over the  $\overline{M}$ -dimensional unit hypercube, has the following form

$$\Phi = E\Psi(\Xi(\alpha_1, \dots, \alpha_{\overline{M}})) = \int_0^1 d\alpha_1 \dots \int_0^1 d\alpha_{\overline{M}}.$$

Further, it is convenient to change the designations and talk about the calculation of the integral  $J = \int_{D_s} f(X)dX$  using the  $s$ -dimensional unit hypercube,  $X = (x_1, \dots, x_s)$ ,  $D_s = \{X : 0 \leq x_i \leq 1; i = 1, \dots, s\}$ . The problem of calculating this integral using the Monte Carlo method is well known. If it is estimated using the sum  $J \approx 1/N \sum_{j=1}^N f(\alpha_1^j, \dots, \alpha_s^j)$ , where  $\alpha_i^j$  are independent realizations of a random variable uniformly distributed on  $[0, 1]$ , and  $f$  is a quadratically integrable function, then for the error it is possible to construct a confidence interval of width  $O(N^{-1/2})$ . At the same time, a number of articles [2–4] based on the theory of numbers considerations, pointed out sequences of  $s$ -dimensional vectors  $Y_1, \dots, Y_N$ , for which the error  $J - 1/N \sum_{j=1}^N f(Y_j)$  decreased as  $\ln^s N/N$ , for functions that have the first partial derivative with respect to each variable. This result is obviously almost  $\sqrt{N}$  times better than the Monte Carlo method. The sequences  $Y_1, \dots, Y_N$ , possessing the property mentioned above, are called quasi-random. Extensive literature is devoted to their properties and applications (see, for example, [5] and the bibliography available there).

As a rule, the authors consider that quasi-random sequences are significantly better than the pseudo-random sequences used in the Monte Carlo method. The legitimacy of such comparisons is studied in detail below.

## 2. Koksma-Hlawka Inequality and Random Quadrature Formulas

One of the well-known approaches to constructing the sum

$$K_N [f] = \sum_{j=1}^N A_j f(X_j), \tag{1}$$

where  $A_j$  are constants,  $X_j = (x_1^{(j)}, \dots, x_s^{(j)}) \in D \subset R^s$ , which is used in the integral  $\int_D f(X)dX$  calculation, is as follows. It is assumed that  $f \in \mathcal{F}$  belongs to a linear normed space of functions, the error of the integration formula

$$R_N [f] = \int_D f(X)dX - K_N [f] \tag{2}$$

is considered as a functional in this space and parameters of the formula (1) are  $A_j$  and  $X_j, j = 1, \dots, N$  are chosen so as to minimize the norm of the functional [6].

This task is usually very difficult. It is enough to note that it is possible to obtain the explicit expression of the above-mentioned functional norm only in a few particular cases. One of these cases is  $A_j = 1/N, j = 1, \dots, N, D$  is a unit hypercube  $D_s = \{X : 0 \leq x_l \leq 1, l = 1, \dots, s\}$  and  $\mathcal{F}$  is a space of functions of bounded variation in the Hardy-Krause sense. In this case one can use the Koksma-Hlawka inequality [7,8]

$$|R_N [f]| \leq V(f) \cdot D^*(X_1, \dots, X_N), \tag{3}$$

where  $V(f)$  is the variation mentioned above, and  $D^*(X_1, \dots, X_N)$  is the error norm, which is called a discrepancy in the non-Russian literature or a deviation in the Russian one. Sometimes it is also called a star discrepancy. In the number-theoretical sense, this quantity characterizes the uniformity of the sequences distribution and equals to

$$\sup_X |S(X) - \frac{1}{N}A(X_1, \dots, X_N)|, \tag{4}$$

where  $S(X)$  is the volume of the multidimensional box  $\Delta(X) = \{Y : y_l \leq x_l, l = 1, \dots, s\}$ , and  $A(X_1, \dots, X_N)$  is a number of points of sequences belonging to  $\Delta(X)$ .

Similar explanations regarding the value  $V(f)$  can be found in Appendix A for this article. For our purposes, it suffices to note that for  $N \rightarrow \infty$ , one knows the upper bound for the asymptotic behavior of  $D^*(X_1, \dots, X_N)$  [9], namely

$$D^*(X_1, \dots, X_N) \leq O\left(\frac{\ln^s N}{N}\right). \tag{5}$$

The authors of [2–4] indicate of the algorithms for constructing sequences  $X_1, \dots, X_N$  for which equality is achieved in (5). These sequences are called quasi-random due to the formal similarity of algorithms of quasi-random numerical integration with the Monte Carlo method of calculating multiple integrals. As we have already said, the authors of numerous articles devoted to the study and use of quasi-random numbers usually note that the Monte-Carlo method has an asymptotic error decrease of the same order as  $O(N^{-1/2})$ , while from (5) we can conclude that quasi-random methods provide an error decrease as  $O(N^{-1})$ , more precisely  $O(N^{-1+\epsilon})$  for any arbitrarily small  $\epsilon$ .

The proof of Inequality (3) given in the literature is quite large and complex. As we have already noted, it can easily be obtained by means of functional analysis. For  $s = 2$  we showed it in the application. The general case simply requires more complex definitions. It can be noted that using the theory of cubature formulas [1], one can obtain analogs of Inequality (3) for Sobolev functions and many other classes of functions and specify sequences for these classes that have faster order of the error decrease. However, at the same time, computational algorithms are usually significantly complicated.

Other important applications of classical computational mathematics arise while estimating the error of quasi-random methods. The construction of the confidence interval in the Monte Carlo method automatically gives an estimate of the error, but for the quasi-Monte Carlo in its pure form there is only the Inequality (3), which is of little use when solving a specific problem. This difficulty can be overcome by randomizing quasi-random points.

Suppose  $Y_1, \dots, Y_N$  are quasi-random vectors of dimension  $s$ , and  $\vec{\beta}_1, \dots, \vec{\beta}_m$  are realizations of vectors of the same dimension, uniformly distributed in  $D_s$ . The sum  $S_{m,N}(f) = 1/m \sum_{l=1}^m 1/N \sum_{k=1}^N f(\{Y_k + \vec{\beta}_l\})$  is an unbiased estimate of the integral

$$ES_{m,N}(f) = \int_{D_s} f(X) dX. \tag{6}$$

The curly brackets denote the operation of taking the fractional part performed on the components of the vector. The error for this randomized sum can be approximated using the central limit theorem. The following statement is trivial

$$J \approx \frac{1}{N} \sum_{k=1}^N f(\{Y_k + \vec{\beta}\}), \tag{7}$$

where  $\vec{\beta}$  is a vector uniformly distributed in  $D_s$ , is a random quadrature formula with one free node. The theory of such formulas is given in detail in [1]. With the help of this theory, one can establish for which class of functions the formula is exact. A number of papers consider other methods for randomizing quasi-random numbers, known as scrambling methods.

### 3. Numerical Error Estimation Experiment: Monte Carlo and Quasi-Monte Carlo

In this paper we show that references to Inequality (3) when evaluating a computational algorithm are not completely correct, at least for  $s > 5$ , and we suggest some correct approaches to determining the asymptotical error decrease of the quasi-random methods. It can be immediately noted that in the case of the Monte Carlo (MC) method, we are talking about the width of the confidence interval and the asymptotic behavior in the central limit theorem is already seen for small  $N$  ( $N > 5$ , for example). The situation is different for the expression (5). The multiplier  $\ln^s N$  plays a significant role already for

$s \geq 4$ . The asymptotics of order  $O(N^{-1+\epsilon})$  can be reached when  $N \rightarrow \infty$ , but as follows from Table 1, for  $s = 5$  the rate of error decrease of the two methods becomes approximately equal with  $N^{12}$ , for  $s = 10$ , with  $N^{40}$ , for  $s = 15$  with  $N = 10^{65}$ .

**Table 1.** Asymptotical error decrease.

$s$	5			10			15		
$N$	$10^6$	$10^{12}$	$10^{14}$	$10^6$	$10^{18}$	$10^{40}$	$10^6$	$10^{30}$	$10^{65}$
$\frac{\ln^s N}{N}$	0.5	$10^{-5}$	$10^{-8}$	$10^5$	$10^{-2}$	$10^{-21}$	$10^{11}$	$10^{-3}$	$10^{-33}$
$\frac{1}{\sqrt{N}}$	$10^{-3}$	$10^{-6}$	$10^{-8}$	$10^{-3}$	$10^{-9}$	$10^{-20}$	$10^{-3}$	$10^{-15}$	$10^{-33}$

The data given in Table 1 clearly confirm the unsuitability of Inequality (5) to show the advantages of the quasi-Monte Carlo (QMC) method for calculating integrals of large multiplicity and, with increasing multiplicity of the integral, this situation worsens. Many authors (for example, [10]) confirm that the real error decrease for different values of  $N$  for  $s > 5$  does not obey inequality (5), but behaves like  $N^{-1+\epsilon}$  with  $0 \leq \epsilon \leq 1/2$ . Thus, one should speak about the quality of a particular quasi-random sequence only for reasonable values of  $N$ , when the asymptotic behavior indicated by equality (3) hasn't been fulfilled yet and one should introduce a reasonable quality criterion only from empirical considerations. First, we discuss the behavior of the integration error in some numerical examples. Let  $f(X)$  be defined and integrable in the  $s$ -dimensional unit cube  $D_s$  and the integral  $\int_{D_s} f(X)dX$  is calculated using the cubature formula (1).

Consider the integrals:

$$I_1 = \int_{D_s} 2^s e^s / (e - 1)^s \prod_{i=1}^s x_i e^{-x_i^2} dX, I_2 = \int_{D_s} (8/\pi)^s \left( \prod_{i=1}^s x_i (1 - x_i) \right)^{1/2} dX. \tag{8}$$

$$I_3 = \int_{D_s} \prod_{i=1}^s \frac{|4x_i - 2| + i}{i + 1} dX, I_4 = \int_{D_s} \sqrt{1/|\sum_{i=1}^s (x_i - 1/2)^2 - 1/4|} dX. \tag{9}$$

The exact value of the integrals  $I_1, I_2, I_3$  are known, they are equal to 1 for all  $s$ . Table 2 shows the absolute error of the QMC method, calculated for  $N = 10^6$  and for the quasi-random Sobol (ErrS) and Halton (ErrH) sequences.

**Table 2.** The absolute error of the quasi-Monte Carlo (QMC) method.

$I_i$	$I_1$			$I_2$			$I_3$		
$s$	10	15	20	10	15	20	10	15	20
ErrS	$10^{-4}$	$10^{-4}$	$10^{-3}$	$10^{-6}$	$10^{-4}$	$10^{-4}$	$10^{-7}$	$10^{-6}$	$10^{-6}$
ErrH	$10^{-4}$	$10^{-3}$	$10^{-3}$	$10^{-6}$	$10^{-4}$	$10^{-4}$	$10^{-7}$	$10^{-7}$	$10^{-7}$
$\frac{\ln^s N}{N}$	$10^5$	$10^{11}$	$10^{16}$	$10^5$	$10^{11}$	$10^{16}$	$10^5$	$10^{11}$	$10^{16}$

The obtained calculations lead to the following conclusions. The error values differ by no more than one order of magnitude for the considered sequences. The well-known estimate  $\ln^s N/N$  cannot serve as a reasonable estimate of the error in this case (see the last line in the Table 2). As already noted, the practical application of quasi-Monte Carlo for large  $s$  is limited. The calculations of integrals by the QMC method, the exact value of which is unknown, do not allow us to compare the accuracy of the various quasi-sequences used among themselves.

#### 4. The Criterion of Decreasing Residual

We show that the use of the randomized quasi-Monte Carlo (RQMC) procedure allows us to effectively estimate the error of the quasi-Monte Carlo method. Consider a new criterion, that will allow us to judge about the error decrease depending on  $N$ , and will provide an opportunity to compare the quality of various quasi-random integration methods.

As a quality criterion, we will use the average order of the error decrease.

Consider the error change interval  $[N_1, N_2]$  for some randomized quasi-random quadrature formula. The average value for a given number of realizations  $N, N \in [N_1, N_2]$  is denoted  $\alpha$  and we will approximate  $R_f(N)$  using the least squares method with the function  $y = a + b \cdot N^{-\alpha}$ . The value of  $\alpha$  is called the average order of the error decrease. In the case when the exact value of the integral is unknown we use the error estimate obtained by randomization.

Let us estimate the average order of the error decrease of the numerical integration for integrals  $I_4$ . The exact value of these integrals is not known. The calculations were carried out for  $n = 10^5$  (the number of nodes) and  $M = 10$  (the number of randomizations) for quasi-random sequences of the Sobol and Halton method. Moreover, the total number of nodes is  $N = 10^6$ . The value of random error  $R_f(N)$  is approximated on the following intervals:  $\Delta r_1 = [99,700,99,800], \Delta r_2 = [99,800,99,900], \Delta r_3 = [99,900,100,000]$ . The results of the calculations are given in Tables 3 and 4, where  $\delta$  is the average order of the error decrease value on a given interval.

**Table 3.** The average decreasing order of the error  $\alpha = \alpha(\Delta)$  for the integral  $I_4$ . Randomized quasi-Monte Carlo (RQMC).

Method	s = 15				s = 20			
	$\Delta r_1$	$\Delta r_2$	$\Delta r_3$	$\delta$	$\Delta r_1$	$\Delta r_2$	$\Delta r_3$	$\delta$
RQMC, Sobol	0.84	0.85	0.87	0.85	0.95	1.0	1.0	0.98
RQMC, Halton	0.70	0.64	0.61	0.65	0.85	0.92	0.98	0.92

**Table 4.** The average decreasing order of the error  $\alpha = \alpha(\Delta)$  for the integral  $I_1$ .

Method	s = 15				s = 20			
	$\Delta r_1$	$\Delta r_2$	$\Delta r_3$	$\delta$	$\Delta r_1$	$\Delta r_2$	$\Delta r_3$	$\delta$
RQMC, Sobol	0.59	0.61	0.61	0.61	0.65	0.63	0.52	0.47
RQMC, Halton	0.61	0.61	0.60	0.61	0.53	0.52	0.52	0.60

Conducted calculations show the relative stability of the estimate of  $\alpha$  for changing  $N_1$  and  $N_2$ .

#### 5. Conclusions

Error analysis of numerical methods plays a major role in the choice of algorithms for solving a problem. The main goals of this article are to propose a new quality criterion for algorithms for calculating multidimensional integrals; point out the incorrect use of the Koksma-Hlawka inequality when comparing the asymptotic error behavior of the Monte Carlo and the quasi-Monte Carlo methods for calculating integrals; and propose a new quality criterion for calculation algorithm integrals with large multiplicity.

This will allow, in particular, to choose the ratio between the number of random and quasi-random components of the nodes used in quadrature formulas, when their number (constructive dimension) is very large. The results obtained are confirmed by numerical examples. It makes possible to judge the comparative quality of various quasi-random sequences in some cases.

The results obtained, can be useful in solving other problems (for example, optimization problems). However, this requires separate studies that are beyond the scope of this work.

**Author Contributions:** Conceptualization, S.E.; methodology, S.E. and S.L.; software, S.L.; validation, S.E. and S.L.; formal analysis, S.E.; investigation, S.E. and S.L.; writing—original draft preparation, S.E.; writing—review and editing, S.E. and S.L.; supervision, S.E.; project administration, S.E.

**Funding:** This research is supported by the Russian Foundation for Basic Research, under Grant number 17-01-00267-a.

**Conflicts of Interest:** The authors declare no conflict of interest.

### Appendix A

For the reader’s convenience a brief proof of the inequality Koksma-Hlawki in the two-dimensional case is given.

Let  $f(x, y)$  have integrable derivatives  $\partial f(x, 1)/\partial x$ ,  $\partial f(1, y)/\partial y$  and  $\partial^2 f(x, y)/\partial x\partial y$  in unit square  $D_2 = \{0 \leq x \leq 1; 0 \leq y \leq 1\}$ . Then we have

$$f(x, y) = f(1, 1) + \int_x^1 f'_x(u, 1)du + \int_y^1 f'_y(1, v)dv - \int_x^1 \int_y^1 f''_{xy}(u, v)dudv.$$

Elementary transformations allow to obtain from

$$R_N[f] = \int_{D_2} f(x, y)dxdy - \frac{1}{N} \sum_{j=1}^N f(x_j, y_j)$$

the expression

$$\begin{aligned} R_N[f] &= \int_0^1 f'_x(u, 1)[-u + \frac{1}{N} \sum \Theta(x_j - u)]du \\ &\quad + \int_0^1 f'_y(1, v)[-v + \frac{1}{N} \sum \Theta(y_j - v)]dv \\ &\quad + \int_0^1 \int_0^1 f''_{xy}(u, v)[uv - \frac{1}{N} \sum \Theta(x_j - u)\Theta(y_j - v)]dudv, \end{aligned} \tag{A1}$$

where

$$\Theta(z) = \begin{cases} 1 & z < 0 \\ 0 & z > 0 \\ 1/2 & z = 0. \end{cases}$$

If we denote

$$V(f) = \int_0^1 |f'_x(u, 1)|du + \int_0^1 |f'_y(1, v)|dv + \int_0^1 \int_0^1 |f''_{xy}(u, v)|dudv,$$

that we have

$$|R_N[f]| \leq V(f) \cdot \max(K_1(x_1, \dots, x_N), K_1(y_1, \dots, y_N), K_2(x_1, y_1; \dots; x_N, y_N)),$$

where

$$K_1(x_j, \dots, x_N) = \sup_u |u - \frac{1}{N} \sum \Theta(x_j - u)|,$$

$$K_1(y_j, \dots, y_N) = \sup_v |v - \frac{1}{N} \sum \Theta(y_j - v)|,$$

$$K_2 = \sup_{uv} |uv - \frac{1}{N} \sum \Theta(x_j - u)\Theta(y_j - v)|.$$

However, as you can see,

$$K_1(x_1, \dots, x_N) \leq K_2(x_1, y_1; \dots; x_N, y_N), K_1(y_1, \dots, y_N) \leq K_2(x_1, y_1; \dots; x_N, y_N)$$

and  $|R_N[f]| \leq V(f) \cdot K_2(x_1, y_1; \dots; x_N, y_N)$ ,  $K_2(x_1, y_1; \dots; x_N, y_N)$  and there is discrepancy of a sequence of points.

In general,  $D^*(X_1, \dots, X_N)$  and  $V(f)$  are defined as follows

$$D^*[f] = \sup_{x_1, \dots, x_s} \left| \prod_{l=1}^s x_l - \frac{1}{N} A(X_1, \dots, X_N) \right|,$$

here  $\prod_{l=1}^s x_l$  is the volume of the hyperparallelepiped  $\Gamma = \{Y : 0 \leq y_l \leq x_l, l = 1, \dots, s\}$ , and  $A(X_1, \dots, X_N)$  is the number of sequence points  $X_1, \dots, X_N$ , belonging to this parallelepiped.

## References

1. Ermakov, S.M. *Monte Carlo Method and Related Questions*; Fizmatlit: Moscow, Russia, 1975. (In Russian)
2. Halton, J.H. On the Efficiency of Certain Quasi-random Sequence of Points Inevaluating Multi-dimensional Integrals. *Numer. Math.* **1960**, *1988*, 84–90. [[CrossRef](#)]
3. Sobol, I.M. *Multidimensional Quadrature Formulas and Haar Functions*; Nauka: Moscow, Russia, 1969. (In Russian)
4. Korobov, N.M. *Number-Theoretic Methods in Approximate Analysis*; MZNMU: Moscow, Russia, 2004. (In Russian)
5. Lemieux, C. *Monte Carlo and Quasi-Monte Carlo Sampling*; Springer: New York, NY, USA, 2009.
6. Sobolev, S.L. *Introduction to the Theory of Cubature Formulas*; Nauka: Moscow, Russia, 1974. (In Russian)
7. Koksma, J.F. Een algemeene stelling inuit de theorie der gelijkmatige verdeeling modulo 1. *Math. (Zutphen B)* **1942**, *11*, 7–11.
8. Hlawka, E. Discrepancy and Riemann Integration. In *Studies in Pure Mathematics*; Academic Press: London, UK, 1971; pp. 21–129.
9. Kuipers, L; Niederreiter, H. *Uniform Distribution of Sequences*; John Wiley and Sons: New York, NY, USA, 1974.
10. Levitan, Y.L.; Markovich, N.I.; Rozin, S.R.; Sobol, I.M. On quasi-random sequences for numerical calculations. *Zh. Vychisl. Math Mat. Phys.* **1988**, *28*, 755–759.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).