"Variance reduction via region mapping: QMC variations for integration formulas with one free node."

Anton A. Antonov

tonytonov@mail.ru

Co-author: S.M. Ermakov

Saint-Petersburg State University

MCQMC Warsaw, August 17, 2010

Antithetic variate and region mapping

Many of the modern studies use „antithetic variate“ method in case of multidimensional integrals calculation. (see Hammersley, [1]) According to it, each point on the step of MC algorithm should be reflected from all coordinate planes (when speaking about integration over the hypercube $[-1,1]^d, d \geqslant 1$). The experiments indicate significant advantage in computation efforts in some cases; with a view to variance, such procedure is always not worse than the original one. However, the problem of symmetrization becomes too complicated in multidimensional case: the number of nodes grows as 2^d. In connection with that, there appears an idea of examining a class of quadrature formulas, which generalizes the notion of nodes transformation. The translation into QMC terms is also of interest.

Random quadrature formulas with one free node are built as follows. Let $(\mathfrak{D}, \mathfrak{A}, \mu)$ be a finite-dimensional Euclidean space with finite Lebesgue measure $\mu(\mathfrak{D})$ and $K_n[f] = \sum_{i=1}^{n} A_i(x_1, \ldots, x_n) f(x_i)$ is a quadrature sum, approximating an integral $\int_{\mathfrak{D}} \varphi_1(x) f(x) dx$ and having all nodes being dependent on one of them, for example, x_1:

$$x_i = T_i(x_1), i = 2, \ldots, n.$$

The first node is taken randomly. Transformations T_i have Jacobians $|\Phi_i|$. The most natural situation is when transformations T_i form a cyclic group of order n, generated by T:

$$T_i = T^i, i = 0, \ldots, n-1, \text{ where } T = T_1 \text{ and } T^n = T^0 \neq T^i \text{ for } i < n.$$

The following result may be obtained within these assumptions(Granovskij, Ermakov, [2], [3]).

Theorem 1. *Let a numeric quadrature with one free node*

$$K_n[f] \approx \int \varphi_1(x) f(x) dx,$$

be exact for orthonormal functions $\varphi_1, \ldots, \varphi_N$. *Sum* $K_n[f]$ *is an unbiased estimate of integral* $J = \int \varphi_1(x) f(x) dx$, *if and only if transformations* T_i *of space* $\mathfrak{D}$ *satisfy the equation system*

$$\sum_{i=1}^{n} \varphi_q(x_i) \varphi_1(x_i) |\Phi_i| = 0,$$

where $q = 2, \ldots, N$. *In that case the formula coefficients are defined with equalities*

$$A_i(x_1, \ldots, x_n) = \frac{\varphi_1(x_1) |\Phi_{i-1}|}{\sum\limits_{i=1}^{n} \varphi_1^2(x_i) |\Phi_{i-1}|}, i = 1, \ldots, n,$$

while the density of x_1 *distribution concerning Lebesgue measure looks like*

$$u(x_1) = \frac{g(x_1)}{n} \sum_{i=1}^{n} \varphi_1^2(x_i) |\Phi_{i-1}|,$$

where $g(x)$ *is an arbitrary measurable nonnegative function, satisfying the condition*

$$n^{-1} \sum_{i=1}^{n} g(x_i) = 1.$$

Interesting to notice that in case of orthogonal transformation all the Jacobians are equal to unity. Moreover, if $g(x)$ is taken to be

$$g(x) = \frac{n\varphi_1^2(x)}{\sum\limits_{i=1}^{n} \varphi_1^2(x_i)|\Phi_{i-1}|},$$

the density function will look like

$$u(x) = \varphi_1^2(x).$$

Finally, put $\varphi_1(x) = 1$ in order to simplify the final formula:

$$\int f(x)\mu(dx) \approx \frac{\mu(\mathfrak{D})}{\sum\limits_{i=0}^{n-1} |\Phi_i|} \sum_{i=0}^{n-1} f(T^i(x))|\Phi_i|,$$

while the density looks like

$$u(x) = \frac{g(x)}{n\mu(\mathfrak{D})} \sum_{i=0}^{n-1} |\Phi_i|.$$

Again only orthogonal transformations are considered, additionally $g(x) = 1$. In that case we obtain $u(x) = \frac{1}{\mu(\mathfrak{D})}$, which means uniform distribution of the free node, and

$$\int f(x)dx \approx \frac{\mu(\mathfrak{D})}{n} \sum_{i=0}^{n-1} f(T^i(x)).$$

Into the terms of QMC...

Let us consider QMC approach. We take the first m points of any uniformly distributed sequence in $\mathfrak{D}$ and apply the transformation group T_i. The final lattice will consist of mn nodes. The reason of such procedure is explained by the next statement.

Theorem 2. *Using previously introduced designations, under the condition* $m \to +\infty$

$$\frac{\mu(\mathfrak{D})}{mn} \sum_{j=1}^{m} \sum_{i=0}^{n-1} f(T^i(x_j) \longrightarrow \int_{\mathfrak{D}} f(x)dx.$$

The proof is based on the properties of orthogonal transormation in finite-dimensional Euclidean spaces. It is possible to show that the discrepancy of the final set is not greater than of the initial one. After it is proven, the reference to Koksma-Hlawka inequality finishes the reasoning. It must be noticed that all assumptions and simplifications in the second theorem are necessary and without them the proof fails: one can show that the discrepancy behaviour may get worse after the trasformations.

Remark 1. Previously mentioned „antithetic variate“ method stays in the network of the discussed costructions: for a hupercube of dimension d the group T_i is a cyclic automorphism group of order 2^d.

Example 1: antithetic variate

Let us apply „antithetic variate“ as an example. Given one node, the others are obtained by successive sign changes in each coordinate. This problem is equivalent to the problem of getting all binary combinations of order d. Take a ten-dimensional space and a mixture of two multivariate normal densities with identity covariance matrix and different means as a test fuction. A comparative study of MC and QMC with Halton points is given below.

The true value of the integral is 0.02387006.

	Number of nodes	Value	Rel.error	Abs.error	Conf.int.
MC	1024	0.0143	0.3987	0.0095	-
	2048	0.0161	0.3238	0.0077	0.0098
	4096	0.0204	0.1439	0.0003	0.0107
	8192	0.0268	0.1212	0.0029	0.0057
	16384	0.0226	0.0516	0.0013	0.0049
	32768	0.0233	0.0255	0.0006	0.0044
	...	...	...	...	...
	262144	0.0243	0.0165	0.0004	0.0014
	Number of nodes	Value	Rel.error	Abs.error	-
QMC	1024	0.0746	2.124	0.0507	-
	2048	0.0688	1.883	0.0449	-
	4096	0.0589	1.4685	0.035	-
	8192	0.0469	0.9664	0.023	-
	16384	0.0351	0.4702	0.0112	-
	32768	0.0292	0.2238	0.0053	-
	...	...	...	...	-
	262144	0.0248	0.038	0.0009	-

Example 1: antithetic variate

On the one hand, the simplicity of MC realization and quickly obtained comparatively exact answer say for this method. However, it must be taken into account that because of the high dimensity for QMC only a few first Halton points are used. That is why asymptotic properties may influence the result only when the function evaluation number is quite large. Nevertheless, the residual is monotonically decreasing.

Generally speaking, this method becomes too complicated in higher dimensions. To save some good properties, it is possible to apply „antithetic variate“ on first coordinates only. This method is gaining more and more popularity in recent studies.

As an additional study let us compare the performance of „antithetic variate“ on all coordinates and only on the origin of coordinates. The computation results are given below.

Full antithetic		Center antithetic	
Number of nodes	Rel.error	Number of nodes	Rel.error
1024	0.3987	2	0.4373
2048	0.3238	4	0.4336
4096	0.1439	8	0.1911
8192	0.1212	16	0.0432
16384	0.0516	32	0.0532
32768	0.0255	64	0.15
...	...	...	...
262144	0.0165	512	0.0319

Formulas with one free node

Another method of obtaining quadratures is to fix some of the nodes and to apply usual MC or QMC algorithm to others. For instance, for a scheme with n nodes it is convenient to fix $n-1$ nodes, leaving only one to be free ([2]). Let x_1 be the free node, and $x_i = \hat{x}_i$ for $i = 2, \ldots, n$ be the fixed nodes for an interpolation quadrature sum $K_n[f] = \frac{\Delta(f,Q)}{\Delta(Q)}$ (where $Q = (x_1, \hat{x}_2, \ldots, \hat{x}_n)$) for integral estimate $\int \varphi_1(x) f(x) dx$. The formula coefficients are $A_i(Q) = \frac{\Delta_i^{(1)}(Q)}{\Delta(Q)}$ (here $\Delta_i^{(1)}(Q)$ denotes an algebraic cofactor of an element $\varphi_1(x_i)$ of a determinant Δ). Applying the same technique as for formulas with one free node, one may get

$$u(x_1) = \frac{\Delta(Q)}{\Delta_i^{(1)}(Q)} \varphi_1(x_1),$$

and if this function has distribution density properties, then the estimate based on this formula will be unbiased. The strictest condition here is non-negativeness in all domain. If $\varphi_1(x) = \text{const}$ and $\varphi_i(x)$ are algebraic polynomials, then it can be shown ([2]) that this condition holds only if all of $\hat{x}_i$ are situated on the domain border.

Example 2: obtaining a new formula

In this example it is explained how such kind of formulas can be built. Let us take $d = 5$ in order to simplify computations performed in hand. We shall be obtaining a formula, which will be exact for a constant and linear polynomials for each coordinate. That is why $\varphi_1(x) = C_1$, and $\varphi_i(x) = C_2 x_i$ for $i = 2, 3, 4, 5$ (constants C_1 and C_2 are chosen to satisfy the property of normalization and they do not affect the work with determinants, so they are omitted). It is natural to put the fixed nodes so that the formula could be easily generalized for higher dimensions. Let those nodes stay in hupercube corners: $\hat{x}_2 = (1,1,1,1,1), \hat{x}_3 = (1,1,1,1,-1), \hat{x}_4 = (1,1,1,-1,-1), \hat{x}_5 = (1,1,-1,-1,-1), \hat{x}_6 = (1,-1,-1,-1,-1)$. The determinant $\Delta(Q)$ will look like (here the free node has coordinates $x_1 = (a, b, c, d, e)$)

$$\begin{Vmatrix} 1 & a & b & c & d & e \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & -1 \\ 1 & 1 & 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 \end{Vmatrix}$$

The determinant equals $16 - 16a$, while the minor values are equal to 16, $8a + 8e$, $8e - 8d$, $8c - 8d$, $8c - 8b$, $8a - 8b$. The final formula with normalizing constants:

Example 2: obtaining a new formula

$$\int_{[-1,1]^5} f(x1,x2,x3,x4,x5)\,dx_1dx_2dx_3dx_4dx_5 \approx \frac{32}{1-a}f(a,b,c,d,e)+$$

$$+\frac{32}{2-2a}\Big(-(a+e)f(1,1,1,1,1)+(e-d)f(1,1,1,1,-1)+$$

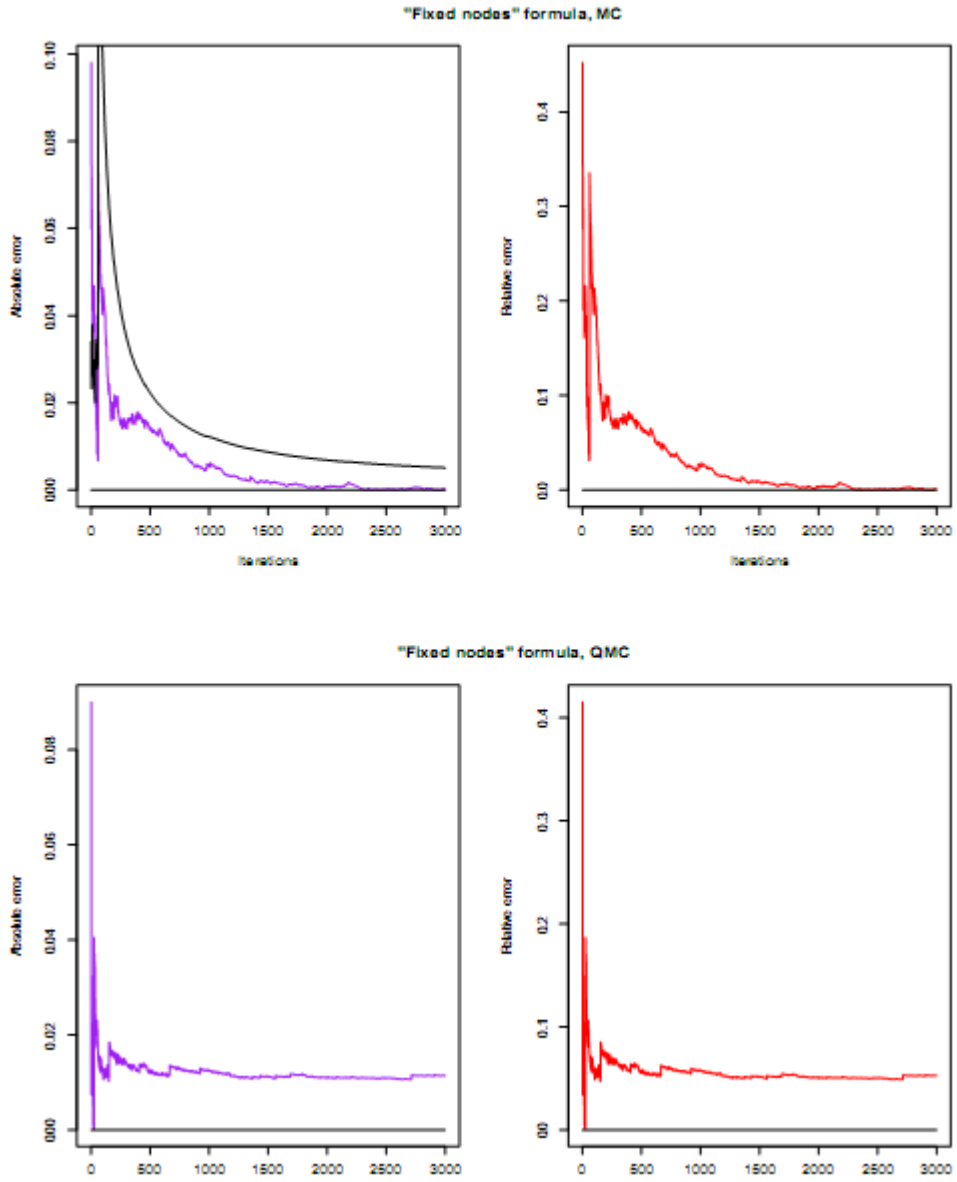
$$+(d-c)f(1,1,1,-1,-1)+(c-b)f(1,1,-1,-1,-1)+(b-a)f(1,-1,-1,-1,-1)\Big).$$

The density function is

$$u(a,b,c,d,e)=\frac{1-a}{32}$$

and can be easily simulated by an inverse function method: the only thing to do is to solve a quadratic equation. Let us apply the formula with MC and QMC methods for the previously described test function.

Example 2: formula performance



Example 2: reasonable ineffectiveness

Here MC procedure demonstrates the correct behaviour in the bounds of confidence intervals. At the same time, QMC method has no convergency at all! Let us try to investigate the reason for it. The density function shows that a point close to $+1$ on the first variable must be a rare event. Such points do bring in large error, because the function value is multiplied by a large number $\frac{32}{1-a}$. The true value is 0.2168633. The quasirandom points with relative error more than 100% are listed below.

```
> nodes.v[1,ans>2]
X257     X392     X595     X662     X797     X932     X1067
0.950617 0.975308 0.983471 0.958847 0.983539 0.942386 0.967078
X1200    X1202    X1238    X1472    X1607    X1805    X1863
0.984973 0.991769 0.9872   0.953360 0.978052 0.986476 0.99040
X1877    X2012    X2147    X2282    X2355    X2410    X2417
0.961591 0.986282 0.945130 0.969821 0.971450 0.987979 0.994513
X2488    X2687    X2822    X2960
0.99360  0.956104 0.980795 0.972952
```

The reason of such Halton sequence behaviour seems to be caused by their tendency to fall too close to the hypercube border. Further studies are necessary to achieve good convergency (for instance, Sobol points with scrambling may solve the problem and include QMC method into the list of effective instruments for this case).

Example 3: six nodes formula

Numeric quadratures with m free nodes with $1 < m < n$ are not studied significantly, so there exist only few such formulas. One of them has two free nodes (it belongs to Ermakov, [3]):

$$\int\limits_{I^d} f(x)dx \approx \frac{1}{3}\left[f(x_1) + f(x_2) + f(x_3)\right],$$

where I^d is a d-dimensional hypercube, and three nodes coordinates are tied with the following conditions: $(x_1^{(i)})^2 + (x_2^{(i)})^2 + (x_3^{(i)})^2 = 1$, $i = 1, 2, \ldots, d$. If x_1 and x_2 are considered to be free nodes, then their joint density is

$$u(x_1, x_2) = \frac{2^s}{\pi^s} \prod_{k=1}^{d} \frac{1}{\sqrt{1 - (x_1^{(k)})^2 - (x_2^{(k)})^2}}.$$

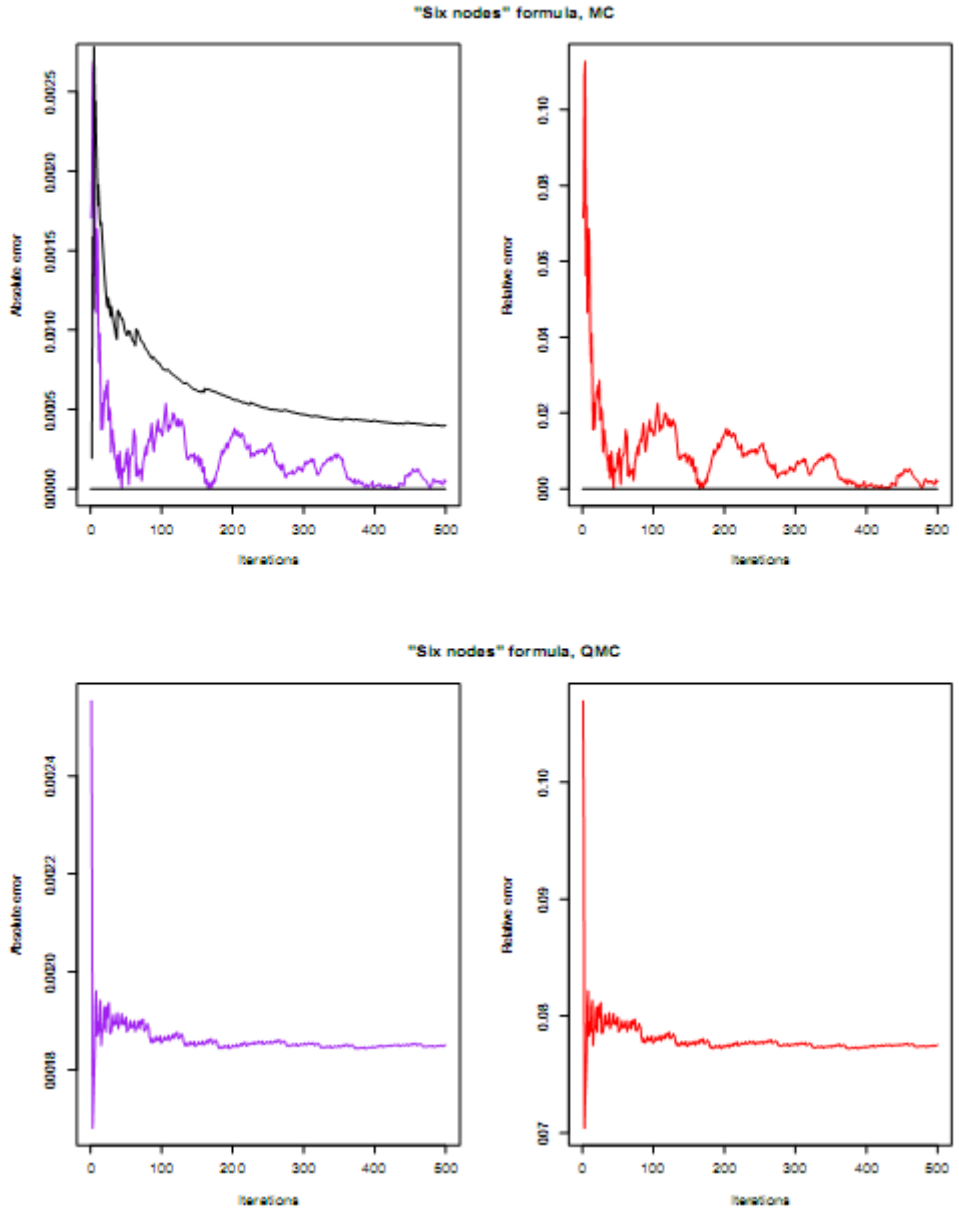
Example 3: modelling

In other words, the point $x_1^{(k)}, x_2^{(k)}, x_3^{(k)}$ is uniformly distributed on a unity sphere, which is convenient for modelling. Additional symmetrization on the origin of coordinates gives us a formula with six nodes in a hypercube and is exact for constants, linear and quadratic functions with respect to $x^{(k)}$, as well as for some of their pairwise products. One of the greatest advantages of this formula is the following fact: the number of nodes does not depend on the dimension.

There exists a special uniform distribution on a three-dimensional sphere modelling method. One of the coordinates may be taken uniformly distributed on the diameter, while the other two have uniform conditional distribution on the corresponding circle. This method is much more advantageous in comparison with rejective sampling, for example.

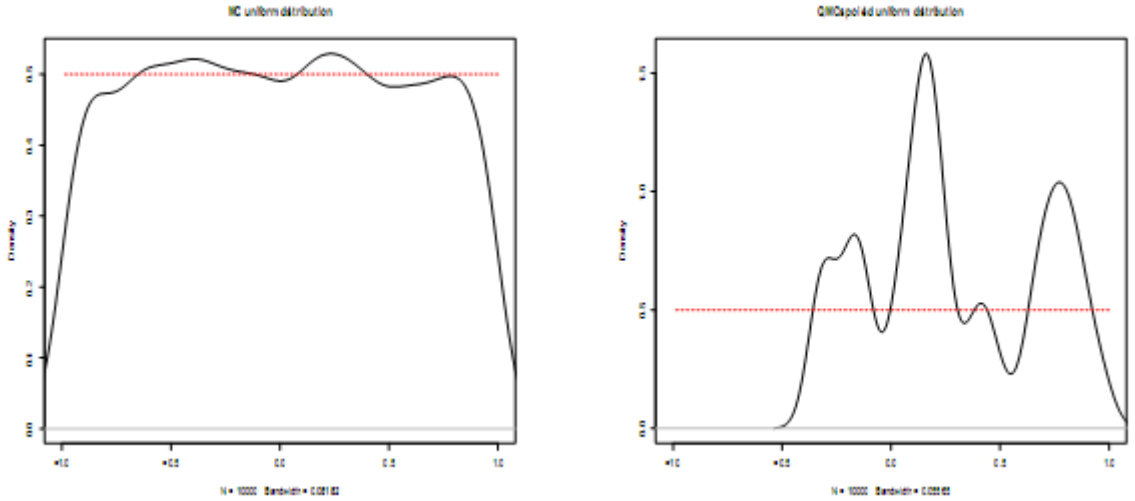
The test function is the same, space dimension is 10. The results are given in pictures.

Example 3: formula performance



Example 3: problem investigation

MC procedure again shows good results, while QMC gives no convergency. This happens because the method described above is partially based on the same rejective sampling, which requires independent realizations of a random quantity. That is why Halton points usage results in some another distribution, not a desired one. Indeed, the graphs below show density estimates. The true distribution must be uniform on $[-1, 1]$, i.e. should have constant density. Both pictures show real distribution.



This explanation indicates that the unconsidered usage of rejection sampling in QMC procedure is extremely dangerous and should be avoided. The problem of QMC convergency in complicated MC schemes is still of particular interest and may be investigated in the nearest future.

References

[1] J.M. Hammersley, K.W. Morton, *A new Monte-Carlo technique – antithetic variates* // Proc. Cambr. Phil. Soc. 52, 1956, pp. 449–474.

[2] B. L. Granovskij and S. M. Ermakov. *The Monte Carlo method* // Translated from Itogi Nauki i Tekhniki (Teoriya Veroyatnostei, Matematicheskaya Statistika, Teoreticheskaya Kibernetika), Vol. 13, pp. 59–108, 1976.

[3] S. M. Ermakov. *Die Monte Carlo Methode und verwandte Fragen* // VEB Deutscher Verlag der Wissenschaften, 1975, Berlin.

[4] H. Niederreiter, *Random Number Generation and Quasi-Monte Carlo Methods* // Dover Publications, Dover, 2006. pp. 22–46.

[5] L. Kuipers, H. Niederreiter, *Uniform Distribution Of Sequences* // Wiley, New York, 1974. pp. 1–31, 88–118.

[6] J. Halton, *On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals* // Numerische Mathematik 2, 1960, pp. 84—90.

[7] I.M. Sobol', *Distribution of points in a cube and approximate evaluation of integrals* // Zh. Vych. Mat. Mat. Fiz. 7, 1967, pp. 784—802.

Thanks for your attention!

Questions?