Multilevel Monte Carlo Methods

Stefan Heinrich

Fachbereich Informatik
Universität Kaiserslautern
D-67653 Kaiserslautern, Germany
`heinrich@informatik.uni-kl.de`,
WWW home page: `http://www.uni-kl.de/AG-Heinrich/Stefan.html`

Abstract. We study Monte Carlo approximations to high dimensional parameter dependent integrals. We survey the multilevel variance reduction technique introduced by the author in [4] and present extensions and new developments of it. The tools needed for the convergence analysis of vector-valued Monte Carlo methods are discussed, as well. Applications to stochastic solution of integral equations are given for the case where an approximation of the full solution function or a family of functionals of the solution depending on a parameter of a certain dimension is sought.

1   Introduction

Monte Carlo is often the method of choice when high dimensional problems have to be solved. Classical Monte Carlo methods approximate a scalar, like a high dimensional integral or a functional of the solution of an integral equation (the value in a point, or a weighted mean). This field is well studied.

In this paper we are concerned with a different question: What if we want to approximate whole functions, like, for example, integrals depending on a parameter, or the solution of an integral equation on a submanifold or even the full solution. Much less is known for these problems. Is Monte Carlo still advisable in these cases? And if yes, what are efficient MC-methods for these situations?

Let us first introduce the basic numerical problems to be studied:

Parametric integration: Compute

$$u(\lambda) = \int_G f(\lambda, t)dt$$

as a function of $\lambda \in \Lambda$, where $\Lambda \subseteq \mathbf{R}^{d_1}$ is the parameter domain, $G \subseteq \mathbf{R}^{d_2}$ the integration domain, and f is a given function on $\Lambda \times G$. Applications include high dimensional integrals in finance.

Integral equations: Let u be the solution of an integral equation

$$u(s) = \int_G k(s, t)u(t)dt + f(s),$$

where $G \subseteq \mathbf{R}^{d_2}$ and k and f are given functions. Consider the following task: Given a family of functionals g_λ ($\lambda \in \Lambda$), where $\Lambda \subseteq \mathbf{R}^{d_1}$ is the parameter domain, compute

$$v(\lambda) := \langle u, g_\lambda \rangle = \int_G u(s) g_\lambda(s) ds$$

as a function of $\lambda \in \Lambda$. For example, if $\Lambda \subseteq G$ and $g_\lambda = \delta_\lambda$ - the delta-function, this means that we want to compute $u(\lambda)$ on a subset Λ of G, e.g. a submanifold, or, if $\Lambda = G$, we seek to compute the full solution function. Applications are transport problems, where various parameter dependent families of functionals are computed (for example, the particle density in space is an average over the velocity).

Clearly, the question of parameter dependence has been touched in many Monte Carlo papers, but a systematic study was conducted by Frolov, Chentsov [2], Sobol [14], [15], who developed the method of dependent tests, and later on by Mikhailov [10], Prigarin [12], Voytishek [17],[18], who called their class of methods discrete-stochastic procedures. The multilevel approach to these problems originates in Heinrich [4] and was further developed in [5],[6], and by Heinrich and Sindambiwe in [7]. Sindambiwe [13] contains extensions to unbounded domains and first numerical experiments, while Keller [8] presents applications to light transport in computer graphics. Further numerical testing is reported by Voytishek and Mezentseva [19]. It is the aim of this paper to give a short introduction into the basic ideas of this method, its background and its applications. We start with an example which explains the crucial ideas in a very simple situation.

2   A simple example

We are given a function $f(\lambda, t) \in C[0,1]^2$ and want to compute

$$u(\lambda) = \int_0^1 f(\lambda, t) dt \qquad (1)$$

for all $\lambda \in \Lambda = [0,1]$. First we consider the

2.1   Standard, one-level approach:

What is the usual, direct way of applying Monte Carlo to this problem? We fix a grid, $\{\lambda_i = \frac{i}{n}, \quad i = 0, 1, \ldots, n\}$, where $n \in \mathbf{N}$, estimate the respective integrals (1) in the node points and interpolate in some way. We estimate

$$u(\lambda_i) \approx \frac{1}{N} \sum_{j=1}^{N} f(\lambda_i, \xi_j),$$

where ξ_j ($j = 1, \ldots, N$) are independent, uniformly distributed on $[0,1]$ random variables. Note that this is (in a simple form) the basic approach of discrete-stochastic procedures ([17], [18]). Our choice here incorporates the method of

dependent tests [2]: although the ξ_j are independent, the same ξ_j is used for all parameters λ_i, $i = 0, \ldots, n$. Due to these dependencies one obtains smooth approximating curves and avoids the statistical fluctuations from node to node. (It is exactly this approach which we need as a starting point for our multilevel method.) Next one approximates the full function by interpolation: For all $\lambda \in \Lambda$

$$u(\lambda) \approx (Pu)(\lambda) = \sum_{i=0}^{n} u(\lambda_i)\varphi_i(\lambda)$$

$$\approx \eta(\lambda) := \sum_{i=0}^{n} \left(\frac{1}{N} \sum_{j=0}^{N} f(\lambda_i, \xi_j) \right) \varphi_i(\lambda) = \frac{1}{N} \sum_{j=0}^{N} (Pf(\cdot, \xi_j))(\lambda).$$

For this introductory example, piecewise linear interpolation suffices, that is, the φ_i are the respective hat functions. The error of such a method is defined to be the root mean square norm deviation (we have chosen the L_2-norm here, which is the simplest in this respect, other norms are considered below):

$$e(\eta) = \left(\mathbf{E} \|u - \eta\|_{L_2}^2 \right)^{1/2} = \left(\mathbf{E} \int_0^1 |u(\lambda) - \eta(\lambda)|^2 d\lambda \right)^{1/2}.$$

Under the simple smoothness assumption $f \in C^{1,0}([0,1]^2)$ (that means, $f \in C([0,1]^2)$ and f is continuously differentiable with respect to λ), it can be shown that

$$e(\eta) = \mathcal{O}(N^{-1/2} + n^{-1}).$$

Obviously, the computational cost (number of arithmetic operations, random number and function value calls) is $\mathcal{O}(nN)$. Consequently, the optimal error-cost relation is reached for $n = \Theta(N^{1/2})$, which gives an error $\mathcal{O}(N^{-1/2})$ at cost $\mathcal{O}(N^{3/2})$.

2.2 Multilevel-Splitting

Now assume that we have a sequence of grids

$$\{\lambda_{\ell i} = \frac{i}{2^\ell}, \quad i = 0, 1, \ldots, 2^\ell\} \quad (\ell = 0, \ldots, m)$$

with the associated interpolation operators

$$P_\ell u = \sum_{i=0}^{n_\ell} u(\lambda_{\ell i})\varphi_{\ell i},$$

so that $P = P_m$. Trivially, we can represent

$$P = P_m = \sum_{\ell=0}^{m} (P_\ell - P_{\ell-1}) \qquad (P_{-1} := 0).$$

Let us still stay with the standard (dependent test) estimator from above, which can now be represented as

$$\eta = \sum_{\ell=0}^{m} \frac{1}{N} \sum_{j=0}^{N} (P_\ell - P_{\ell-1}) f(\cdot, \xi_j). \tag{2}$$

This allows us to have a closer look at the behaviour of this estimator in the various levels: In the following table we give the order of variance and cost related to the respective level (to emphasize the ideas, we act as if we would compute the standard estimator by formula (2) – which actually would not change the order of cost).

level	0	ℓ	m
square root of variance	$N^{-1/2}$	$2^{-\ell} N^{-1/2}$	$2^{-m} N^{-1/2}$
cost	N	$2^{\ell} N$	$2^{m} N$

We see that the variance reaches its maximum essentially at the first level, while the maximal cost is concentrated at the last level. This leads us to the idea of the multilevel approach: we try to balance error and cost in the most efficient way.

2.3 Multilevel approach

Now we adopt representation (2), but moreover, allow the number of samples in each level to vary. In other words, we choose $N_\ell \in \mathbf{N}$ $(l = 0, \ldots, m)$, let $\{\xi_{\ell j}, \ \ell = 0, \ldots, m, \ j = 0, \ldots, N_\ell\}$ be independent, uniformly distributed on $[0, 1]$ random variables, and define the multilevel estimator as

$$\eta^{\text{mult}} = \sum_{\ell=0}^{m} \frac{1}{N_\ell} \sum_{j=1}^{N_\ell} (P_\ell - P_{\ell-1}) f(\cdot, \xi_{\ell j}).$$

A suitable choice of balancing for our simple situation is, for example, $N_\ell \asymp 2^{-3\ell/2} N$ (the notation $\asymp$ is equivalent to the Θ notation). We keep the relation $n = 2^m \asymp N^{1/2}$. Then it can be checked that the error-cost table looks as follows:

level	0	ℓ	m
square root of variance	$N^{-1/2}$	$2^{-\ell/4} N^{-1/2}$	$2^{-m/4} N^{-1/2}$
cost	N	$2^{-\ell/2} N$	$2^{-m/2} N$

It follows that the total stochastic error is $\mathcal{O}(N^{-1/2})$ (the deterministic, systematic error due to the interpolation approximation is $\|u - P_m u\| = \mathcal{O}(2^{-m}) = \mathcal{O}(N^{-1/2})$), and the total cost is $\mathcal{O}(N)$. What did we gain? As compared to the standard, one level method, we saved the grid factor $n (= 2^m)$. That is, we computed an approximation to the whole *family of integrals* with the error $\mathcal{O}(N^{-1/2})$ and cost $\mathcal{O}(N)$ of a standard computation of *one single integral!*

3 A general result

After having explained the ideas at a simple example, we now state general conditions in order to include large classes of domains, smoothness and types of summability into this method and its analysis. At the same time we try to keep the smoothness assumptions minimal in the sense that they are needed only with respect to the parameter variables. Let $\Lambda \subset \mathbf{R}^{d_1}$ and $G \subset \mathbf{R}^{d_2}$ be bounded open sets with Lipschitz boundary. Let $1 \leq q \leq \infty$, $r \in \mathbf{N}$ and assume the following (Sobolev embedding) condition: $r/d_1 > 1/q$. Define

$$W_q^{r,0}(\Lambda \times G) = \left\{ f \in L_q(\Lambda \times G) : \frac{\partial^\alpha f}{\partial \lambda^\alpha} \in L_q, \quad |\alpha| \leq r \right\}$$

$$\|f\|_{W_q^{r,0}} = \left(\sum_{|\alpha| \leq r} \left\| \frac{\partial^\alpha f}{\partial \lambda^\alpha} \right\|_{L_q}^q \right)^{1/q},$$

where $\frac{\partial^\alpha f}{\partial \lambda^\alpha}$ denotes the generalized partial derivative. If $p = \infty$, the integrals are replaced by the essential supremum in the usual way. So $f \in W_q^{r,0}(\Lambda \times G)$ means roughly that $f(\lambda, t)$ is in the standard Sobolev space W_q^r with respect to λ and just in L_q with respect to t.

We proceed in a general way also with the approximating operators. Let

$$P_\ell : W_q^r(\Lambda) \to L_q(\Lambda) \quad (\ell = 0, 1, \ldots)$$

be linear operators of the form

$$P_\ell f = \sum_{i=0}^{n_\ell} f(\lambda_{\ell i}) \varphi_{\ell i}$$

with $\lambda_{\ell i} \in \bar{\Lambda}$ (the closure of Λ), and $\varphi_{\ell i} \in L_q(\Lambda)$. Note that, due to the embedding condition, the $f(\lambda_{\ell i})$ are well-defined. We assume that there exist constants $c_1, c_2, c_3 > 0$ such that

$$c_1 2^{d_1 \ell} \leq n_\ell \leq c_2 2^{d_1 \ell}$$

$$\|I - P_\ell : W_q^r(\Lambda) \to L_q(\Lambda)\| \leq c_3 2^{-r\ell}. \tag{3}$$

Here $I : W_q^r(\Lambda) \to L_q(\Lambda)$ stands for the identical embedding. We do not specify the approximating operators in more detail. We just mention that for standard domains there are plenty of families satisfying these requirements, as e.g. triangular, rectangular or isoparametric finite elements (see Ciarlet [1]). For example, if $r = 2$, and the domain is polyhedral, piecewise linear interpolation suffices.

Based on these tools we now introduce the multilevel estimator:

$$\eta^{\mathrm{mult}} = \sum_{\ell=0}^m \frac{|G|}{N_\ell} \sum_{j=1}^N (P_\ell - P_{\ell-1}) f(\cdot, \xi_{\ell j}) \qquad (P_{-1} := 0).$$

Here $\xi_{\ell j}$ are independent, uniformly distributed on G random variables. In detail, the multilevel estimator looks as

$$\eta^{\mathrm{mult}} = \sum_{\ell=0}^{m} |G| \left[\sum_{i=0}^{n_\ell} \left(\frac{1}{N_\ell} \sum_{j=1}^{N_\ell} f(\lambda_{\ell i}, \xi_{\ell j}) \right) \varphi_{\ell i} \right.$$
$$\left. - \sum_{i=0}^{n_{\ell-1}} \left(\frac{1}{N_\ell} \sum_{j=1}^{N_\ell} f(\lambda_{\ell-1,i}, \xi_{\ell j}) \right) \varphi_{\ell-1,i} \right].$$

Let us mention that we are not restricted to the uniform distribution for the $(\xi_{\ell j})$. We might as well choose another appropriate (e.g. "importance") distribution, say with density π, but then $|G|f(\cdot, \xi_{\ell j})$ in the above estimator should be replaced by $f(\cdot, \xi_{\ell j})/\pi(\xi_{\ell j})$. Finally, we introduce the error criterion: Let $p = \min(2, q)$. We define the error of method η^{mult} as

$$e(\eta^{\mathrm{mult}}) = (\mathbf{E}\|u - \eta^{\mathrm{mult}}\|_{L_q}^p)^{1/p}.$$

Let us explain this choice. Since we deal with Sobolev spaces with summability index q, the natural norm to measure approximation error is the $L_q(\Lambda)$ norm, which we chose here. This, however, does not yet say how we handle the fact that η^{mult} is a vector-valued ($L_q(\Lambda)$-valued) random variable. In the case $2 \leq q \leq \infty$, we take the root mean square norm error (that is, the second moment) as in our introductory example. However if $q < 2$, due to the weak summability assumption, the second moment may not exist, and we therefore then take the q-th moment. The following result which is proved in [6] gives the speed of convergence of the described multilevel Monte Carlo method.

Theorem 1. *Let $1 < q < \infty$, $p = \min(2, q)$. Then there exist constants $c_1, c_2 > 0$ such that for each integer $M > 1$ there is a choice of parameters $m, (N_\ell)_{\ell=1}^{m}$ such that the cost of computing η^{mult} is bounded by $c_1 M$ and for each $f \in W_q^{r,0}(\Lambda \times G)$ with $\|f\|_{W_q^{r,0}} \leq 1$*

$$e(\eta^{\mathrm{mult}}) \leq c_2 \begin{cases} M^{-r/d_1} & \text{if } r/d_1 < 1 - 1/p \\ M^{1/p-1} \log M & \text{if } r/d_1 = 1 - 1/p \\ M^{1/p-1} & \text{if } r/d_1 > 1 - 1/p. \end{cases}$$

Using methods from information-based complexity theory ([16], [11], [3]), it can be shown in a similar way as in [7] that this algorithm is optimal in a very broad sense: No randomized algorithm with cost M can reach a better rate than the above on the given class of functions (up to a possible $\log M$ factor in the case $r/d_1 = 1 - 1/p$).

4   Error estimates via probability in Banach spaces

In this section we explain the tools needed to prove error estimates like the above, that is, for vector-valued random variables. As the analysis of the classical, scalar

valued Monte Carlo method needs the theory of scalar-valued random variables, that is, standard probability theory, we require for our analysis the respective vector-valued tools, which are provided by Banach space probability theory (see [9]). Let us have another, more general look at our problem from this point of view. Suppose we seek to approximate $u \in X$, where X is a Banach space (e.g. of functions, like in our concrete case of section 3, $X = L_q(\Lambda)$). Let $1 \leq p \leq 2$ and let θ be an X-valued random variable with

$$\mathbf{E}\theta = u \quad \text{and} \quad \mathbf{E}\|\theta\|_X^p < \infty.$$

Let, furthermore, $P_0, P_1, \ldots, P_m$ be a sequence of finite rank operators on X. We define an X-valued random variable approximating u by

$$\eta^{\mathrm{mult}} = \sum_{\ell=0}^{m} \frac{1}{N_\ell} \sum_{j=1}^{N_\ell} (P_\ell - P_{\ell-1})\theta_{\ell j} \quad (P_{-1} := 0).$$

Here $(\theta_{\ell j}, \ell = 0, \ldots, m, j = 1, \ldots, N_\ell)$ are independent copies of θ (in the case of section 3 we have $\theta = |G|f(\cdot, \xi) \in L_q(\Lambda) = X$). The error of η^{mult} is defined as

$$e(\eta^{\mathrm{mult}}) = (\mathbf{E}\|u - \eta^{\mathrm{mult}}\|_X^p)^{1/p},$$

and satisfies

$$e(\eta^{\mathrm{mult}}) \leq \|u - P_m u\|_X + (\mathbf{E}\|P_m u - \eta^{\mathrm{mult}}\|_X^p)^{1/p}.$$

So the estimate of the total error reduces to the estimate of the deterministic (systematic) part, $\|u - P_m u\|_X$, which is provided by approximation theory like in (3) and which we will not discuss further, and the stochastic part

$$(\mathbf{E}\|P_m u - \eta^{\mathrm{mult}}\|_X^p)^{1/p},$$

which we will concentrate on. To study it, we need a notion from probability theory of Banach spaces (see [9]). A Banach space X is said to be of type p, $(1 \leq p \leq 2)$ if there is a constant $c > 0$ such that for all N, all $x_1, \ldots, x_N \in X$

$$\mathbf{E}\|\sum_{i=1}^{N} \varepsilon_i x_i\|_X^p \leq c^p \sum_{i=1}^{N} \|x_i\|_X^p, \tag{4}$$

where $(\varepsilon_i)_{i=1}^{N}$ are independent, centered, $\{-1, 1\}$-valued Bernoulli variables. Let us just briefly recall the following facts: For $1 \leq q < \infty$, L_q is of type $\min(2, q)$. Type p implies type p_1 for $p_1 \leq p$. Each Banach space is of type 1, by the triangle inequality (and no space is of type > 2). For $1 \leq q < 2$, L_q is not of type p for $p > q$. L_∞ is not of type p for any $p > 1$. Each finite dimensional space is of type 2. If X is of type p, the type p constant of X is defined to be

$$T_p(X) = \inf\{c > 0 : c \text{ satisfies } (4)\}.$$

The crucial result for us is the following, which can be found in [9], Prop. 9.11.

Proposition 1. *Let X be of type p $(1 \le p \le 2)$ and let ϱ_i $(i = 1, \ldots, N)$ be independent, X-valued random variables with $\mathbf{E}\varrho_i = 0$ and $\mathbf{E}\|\varrho_i\|_X^p < \infty$. Then*

$$\mathbf{E}\|\sum_{i=1}^{N} \varrho_i\|_X^p \le (2T_p(X))^p \sum_{i=1}^{N} \|\varrho_i\|_X^p.$$

From this result we immediately derive the following inequality, which is the basis of the convergence analysis in Theorem 1:

$$\left(\mathbf{E}\|P_m u - \eta^{\mathrm{mult}}\|_X^p\right)^{1/p} \le c \left(\sum_{\ell=0}^{m} N_\ell^{1-p} \mathbf{E}\|(P_\ell - P_{\ell-1})(u - \theta)\|_X^p\right)^{1/p}. \quad (5)$$

Let us make a few remarks about the case $q = \infty$, which was left out in Theorem 1. Since L_∞ is, as mentioned above, not of any nontrivial type, one might wonder if any result like Theorem 1 can hold at all in this situation. It turns out that one can derive a result very close to Theorem 1 also in the case $q = \infty$. Define

$$X_m = \mathrm{span}\left(\bigcup_{\ell=0}^{m} P_\ell(X)\right) = \mathrm{span}\{\varphi_{\ell,i}, \ell = 0, \ldots, m, i = 0, \ldots, n_\ell\}.$$

Then we can make relation (5) more precise:

$$(\mathbf{E}\|P_m u - \eta^{\mathrm{mult}}\|_X^2)^{1/2} \le 2T_2(X_m) \left(\sum_{\ell=0}^{m} N_\ell^{-1} \mathbf{E}\|(P_\ell - P_{\ell-1})(u - \theta)\|_X^2\right)^{1/2}.$$

Now, if X_m is spanned by functions with almost disjoint supports, meaning that for each point $\lambda \in \Lambda$, the number of supports containing λ is bounded by a constant not depending on m (as it usually is the case in the situations mentioned after relation (3)), the spaces X_m are uniformly (in m) isomorphic to $\ell_\infty^{\dim X_m}$. Moreover, it is known that

$$T_2(\ell_\infty^M) \asymp (\log M)^{1/2}.$$

This introduces just a logarithmic factor into the estimates. The following is essentially proved in [7].

Theorem 2. *There exist $c_1, c_2 > 0$ such that for each $M > 1$ there is a choice of parameters such that the cost of η^{mult} is $\le c_1 M$ and for all $f \in W_\infty^{r,0}(\Lambda \times G)$ with $\|f\|_{W_\infty^{r,0}} \le 1$*

$$\left(\mathbf{E}\|u - \eta^{\mathrm{mult}}\|_{L_\infty}^2\right)^{1/2} \le c_2 \begin{cases} M^{-r/d_1}(\log M)^{r/d_1} & \text{if } r/d_1 < 1/2 \\ M^{-1/2}(\log M)^{3/2} & \text{if } r/d_1 = 1/2 \\ M^{-1/2}(\log M)^{1/2} & \text{if } r/d_1 > 1/2. \end{cases}$$

The results are optimal, including the logarithmic factor (except, possibly, for a factor $\log M$ in the case $r/d_1 = 1/2$).

5 Integral equations

Finally, we want to explain how these methods can be used also for integral equations

$$u(s) = f(s) + \int_G k(s,t)u(t)dt.$$

Recall that we want to compute the function

$$v(\lambda) := \langle u, g_\lambda \rangle.$$

We have for $\lambda \in \Lambda$

$$v(\lambda) = \langle f, g_\lambda \rangle + \mathbf{E}_\omega \theta(\lambda, \omega),$$

where the random variable θ is constructed from the trajectory

$$\omega = (t_0, t_1, \ldots, t_\nu)$$

of a Markov chain on G with initial density $p_0(t)$ and transition density $p(s,t)$ as follows

$$\theta(\lambda, \omega) = \sum_{i=0}^{\nu} \langle k(\cdot, t_i), g_\lambda \rangle \frac{k(t_i, t_{i-1}) \ldots k(t_1, t_0) f(t_0)}{p(t_i, t_{i-1}) \ldots p(t_1, t_0) p_0(t_0)}.$$

Now observe that

$$\mathbf{E}_\omega \theta(\lambda, \omega) = \int_\Omega \theta(\lambda, \omega) d\mathbf{P}(\omega)$$

is an integral depending on a parameter. Hence we can transform our previously developed method into an analogous one for integral equations.

References

1. Ciarlet, P. G.: *The Finite Element Method for Elliptic Problems*, North-Holland, Amsterdam 1978.
2. Frolov, A. S. and Chentsov, N. N.: On the calculation of certain integrals dependent on a parameter by the Monte Carlo method, Zh. Vychisl. Mat. Mat. Fiz. **2** (1962), 714 – 717 (in Russian).
3. Heinrich, S.: Random approximation in numerical analysis, *Functional Analysis* (Bierstedt, K. D., Pietsch, A., Ruess, W. M. and Vogt, D. editors), pp. 123 – 171, Marcel Dekker, New York, Basel, Hong Kong 1993.
4. Heinrich, S.: Monte Carlo complexity of global solution of integral equations, J. Complexity **14** (1998), 151 – 175.
5. Heinrich, S.: Wavelet Monte Carlo methods for the global solution of integral equations, *Proceedings of the Third International Conference on Monte Carlo and Quasi-Monte Carlo Methods, Claremont 1998* (H. Niederreiter and J. Spanier, editors), pp. 227 – 237, Springer Verlag, Berlin 1999.

6. Heinrich, S.: The multilevel method of dependent tests, Advances in Stochastic Simulation Methods (N. Balakrishnan, V. B. Melas, S. M. Ermakov, editors), pp. 47 – 62, Birkhäuser, Boston, Basel, Berlin 2000.

7. Heinrich, S. and Sindambiwe, E.: Monte Carlo complexity of parametric integration, J. Complexity **15** (1999), 317 – 341.

8. Keller, A.: Hierarchical Monte Carlo Image Synthesis, Mathematics and Computers in Simulation **55** (2001), no. 1 – 3, pp. 79 – 92.

9. Ledoux, M. and Talagrand, M.: *Probability in Banach Spaces*, Springer, Berlin, Heidelberg, New York 1991.

10. Mikhailov, G. A.: *Minimization of Computational Costs of Non-Analogue Monte Carlo Methods*, World Scientific, Singapore 1991.

11. Novak, E.: *Deterministic and Stochastic Error Bounds in Numerical Analysis*, Lecture Notes in Mathematics 1349, Springer, Berlin, Heidelberg, New York 1988.

12. Prigarin, S. M.: Convergence and optimization of functional estimates in statistical modelling in Sobolev's Hilbert spaces, Russian J. Numer. Anal. Math. Modelling **10** (1995), 325 – 346.

13. Sindambiwe, E.: *Optimal algorithms for parametric integration*, Shaker Verlag, Aachen 1999.

14. Sobol, I. M.: The use of ω^2-distribution for error estimation in the calculation of integrals by the Monte Carlo method, U.S.S.R. Comput. Math. and Math. Phys. **2** (1962), 717 – 723 (in Russian).

15. Sobol, I. M.: *Computational Monte Carlo Methods*, Nauka, Moscow 1973 (in Russian).

16. Traub, J. F., Wasilkowski, G. W., Woźniakowski, H.: *Information-Based Complexity*, Academic Press, New York 1988.

17. Voytishek, A. V.: Discrete-stochastic procedures for the global estimation of an integral which depends on a parameter, Comp. Math. Math. Phys. **36** (1996), 997 – 1009.

18. Voytishek, A. V.: Using the Strang-Fix approximation in discrete-stochastic numerical procedures, Monte Carlo Methods and Appl. **3** (1997), 89 – 112.

19. Voytishek, A. V., Mezentseva T. E.: Practical use of multilevel algorithms, Proceedings of the 4-th St. Petersburg Workshop on Simulation, St. Petersburg 2001 (to appear).