Bull. Nov. Comp. Center, Num. Anal., 4 (1993), 71–81

Comparison of two procedures for global stochastic estimation of functions

E.V. Shkarupa and A.V. Voytishek

Numerical stochastic procedures for estimating integrals depending on parameter are considered. The discrete mesh on the domain of definition of parameter is introduced, and the Monte Carlo algorithms for estimating integral in mesh points are used. The independent Monte Carlo estimates and the "depended tests" method are compared. It is proved that the "depended tests" method is "better" in the sense of error asymptotics, but in some special cases (corresponding numerical examples are given) the computational cost for independent estimates may be less.

Introduction

For estimating the function

$$g(x) = \mathrm{M}\xi(x, \omega)$$

(here ω – stochastic parameter; $x \in X$, X – compact in $\boldsymbol{R}^l$, i.e., ξ – stochastic field with l parameters) by the Monte Carlo method one can use the following procedures:

Procedure A. Introduce a mesh $x_0, x_1, \ldots, x_M$ in X, and for every x_i realize the unbiased stochastic estimate

$$\hat{g}_{n_i}(x_i) = \frac{1}{n_i} \sum_{j=1}^{n_i} \xi(x_i, \omega_j^{(i)})$$

and thereafter approximate the function $g(x)$ with respect to these values. Here $(\omega_1^{(i)}, \ldots, \omega_{n_i}^{(i)})$ is a sample of independent identically distributed (i.i.d.) stochastic elements. In general case for different i samples are distributed differently.

"Depended tests" method. Realize the unbiased stochastic estimate

$$\bar{g}_n(x) = \frac{1}{n}\sum_{j=1}^{n}\xi(x,\omega_j).$$

Here the sample of i.i.d. stochastic elements $(\omega_1,\ldots,\omega_n)$ is the same for all x.

Procedure B. Realize mixed strategy: use the same sample $(\omega_1,\ldots,\omega_n)$ for the values $\hat{g}_{n_i}(x_i)$ in the procedure A (here $n_0 = n_1 = \ldots = n_M = n$).

The "depended tests" method was suggested in [1] and investigated in [2–4], see also surveys in [2–4]. Procedures of the type A and B were investigated in [2, 5–7].

Let us note advantages and disadvantages of considered procedures. As a rule, the "depended tests" method is more simple for realization by computer, and it gives smooth approximation for graph of $g(x)$ in the case, when $g(x)$ is a smooth function. Convergence of the method is conditioned by the smoothness of trajectories of the field $\xi(x,\omega)$; for example, in [2, 4] it is shown the following:

Lemma 1. *Let the stochastic field*

$$\tilde{\xi}(x,\omega) = \xi(x,\omega) - g(x)$$

be continuous on X in mean of p-th degree ($p > 1$), and for every natural number k ($1 \le k \le l$) there exist derivatives

$$\frac{\partial^k\tilde{\xi}(\bar{x}_1,\ldots,\bar{x}_l)}{\partial\bar{x}_1^{m_1}\ldots\partial\bar{x}_l^{m_l}}$$

in mean of p-th degree which are continuous on X in mean of p-th degree. Here m_i are equal to 0 or 1, $m_1 + \cdots + m_l = k$, $x = (\bar{x}_1,\ldots,\bar{x}_l)$, i.e., the mixed derivatives of k-th order of no more than the first order with respect to each coordinate are considered. Moreover, let $\mathbf{D}\tilde{\xi}(x) < A < \infty$, $A = \mathrm{const}$.

Then the error of the estimate $\bar{g}_n(x)$ in "depended tests" method has the order $n^{-1/2}$ with respect to probability, i.e., the relation

$$\mathbf{P}\Big\{\sup_{x\in X}|\bar{g}_n(x) - g(x)| \le C\cdot n^{-1/2}\Big\} \to \mathbf{P}\Big\{\sup_{x\in X}|\xi_0(x,\omega)| \le C\Big\},$$

is true, where ξ_0 is continuous (with respect to probability) Gaussian stochastic function.

The procedure A is on the one hand more cumbersome than "depended tests" method, but on the other hand it allows to take into account the peculiarities of the function $g(x)$ using the special choice of parameters n_0, $n_1, \dots, n_M$, M, densities of distributions $q_i(y)$ of stochastic elements $(\omega_1^{(i)}, \dots, \omega_{n_i}^{(i)})$ and the way of approximation of the function $g(x)$, and the convergence of the procedure A can be obtained for more wide class of stochastic functions $\xi(x, \omega)$.

The procedure B, unlike the "depended tests" method, allows to choose the way of approximation of $g(x)$ using values in mesh points.

The procedures A and B differ in the order of convergence speed. This paper is devoted to the more exact investigation of this difference. Here we shall consider uniform (in probability) stochastic metrics C for estimating the error of the procedures A and B.

1. Convergence of procedure A

Let for simplicity $x = [a, b] \subset \boldsymbol{R}$ and the mesh

$$a = x_0 < x_1 < \dots < x_M = b$$

be such that $x_{i+1} - x_i \equiv h = (b - a)/M$. Consider the stochastic function

$$\tilde{g}(x) = \sum_{i=0}^{M} \hat{g}_{n_i}(x_i) \cdot \varphi_i(x). \tag{1.1}$$

It is the approximation of the function $g(x)$ on set of basic functions $\{\varphi_i\}$. In this paper we choose the linear approximation, i.e., φ_i are linear finite elements or "functions-covers" [8]

$$\varphi_0(x) = \begin{cases} (x_1 - x)/(x_1 - x_0), & \text{when } x \in [x_0, x_1], \\ 0, & \text{otherwise,} \end{cases}$$

$$\varphi_i(x) = \begin{cases} (x - x_{i-1})/(x_i - x_{i-1}), & \text{when } x \in [x_{i-1}, x_i], \\ (x_{i+1} - x)/(x_{i+1} - x_i), & \text{when } x \in [x_i, x_{i+1}], \\ 0, & \text{otherwise,} \end{cases}$$

$$i = 1, 2, \dots, M - 1,$$

$$\varphi_M(x) = \begin{cases} (x - x_{M-1})/(x_M - x_{M-1}), & \text{when } x \in [x_{M-1}, x_M], \\ 0, & \text{otherwise.} \end{cases}$$

We shall be interested in the error of approximation (1.1), i.e., the stochastic value $\delta = \sup\limits_{x \in [a,b]} |g(x) - \tilde{g}(x)|$. Let

$$G(x) = \sum_{i=0}^{M} g(x_i)\varphi_i(x)$$

be the linear approximation with exact values of the function g in mesh points. Then

$$\delta \le \sup_{x\in[a,b]} |g(x) - G(x)| + \sup_{x\in[a,b]} |G(x) - \tilde{g}(x)|. \qquad (1.2)$$

The value

$$\delta_1 = \sup_{x\in[a,b]} |g(x) - G(x)|$$

is not stochastic and it's estimation depends on smoothness of the function $g(x)$. For example, if $g(x) \in C^{(2)}[a,b]$ then

$$\delta_1 \le Ch^2 \cdot \|g(x)\|_{C^{(2)}[a,b]}, \qquad (1.3)$$

where the constant C is independent from h and $g(x)$ and

$$\|u\|_{C^{(2)}[a,b]} = \sup_{x\in[a,b]} |u(x)| + \sup_{x\in[a,b]} \left|\frac{\partial u}{\partial x}\right| + \sup_{x\in[a,b]} \left|\frac{\partial^2 u}{\partial x^2}\right|$$

(see [8, Theorem 2.2.2]).

Now consider the stochastic value

$$\begin{aligned} \delta_2 &= \sup_{x\in[a,b]} |G(x) - \tilde{g}(x)| \\ &= \sup_{x\in[a,b]} \left| \sum_{i=0}^{M} \left(g(x_i) - \sum_{j=1}^{n_j} \frac{\xi(x_i, \omega_j^{(i)})}{n_j} \right) \varphi_i(x) \right|. \end{aligned} \qquad (1.4)$$

Note that

$$\delta_2 = \max_{i=0,1,\dots,M} \left| g(x_i) - \hat{g}_{n_i}(x_i) \right|. \qquad (1.5)$$

It is the consequence of the following elementary result:

Lemma 2. *Let $y_1(x)$ be the straight line, which passes the points $(\hat{x}_0, \hat{y}_0)$ and $(\hat{x}_1, \hat{y}_1)$, and $y_2(x)$ be the straight line, which passes the points $(\hat{x}_0, \hat{y}_0 + \hat{\delta}_0)$ and $(\hat{x}_1, \hat{y}_1 + \hat{\delta}_1)$. Then*

$$\sup_{x\in[a,b]} \left| y_2(x) - y_1(x) \right| = \max\left\{ |\hat{\delta}_0|, |\hat{\delta}_1| \right\}.$$

Proof. Note that

$$y_1(x) = \hat{y}_0 + (x - \hat{x}_0)(\hat{y}_1 - \hat{y}_0)/(\hat{x}_1 - \hat{x}_0),$$
$$y_2(x) = \hat{y}_0 + \hat{\delta}_0 + (x - \hat{x}_0)(\hat{y}_1 + \hat{\delta}_1 - \hat{y}_0 - \hat{\delta}_0)/(\hat{x}_1 - \hat{x}_0), \quad x \in [\hat{x}_0, \hat{x}_1].$$

Then

$$\gamma(x) = y_2(x) - y_1(x) = \hat{\delta}_0 + (x - \hat{x}_0)(\hat{\delta}_1 - \hat{\delta}_0), \quad x \in [\hat{x}_0, \hat{x}_1],$$

i.e., $\gamma(x)$ is the linear function and it takes the maximum and minimum values at the points $\hat{x}_0$ and $\hat{x}_1$ and then

$$\begin{aligned}\max_{x\in[\hat{x}_0,\hat{x}_1]} \left|y_2(x) - y_1(x)\right| &= \max\left\{\left|\max_{x\in[\hat{x}_0,\hat{x}_1]} \gamma(x)\right|, \left|\min_{x\in[\hat{x}_0,\hat{x}_1]} \gamma(x)\right|\right\} \\ &= \max\left\{|\gamma(\hat{x}_0)|, |\gamma(\hat{x}_1)|\right\} = \max\left\{|\hat{\delta}_0|, |\hat{\delta}_1|\right\}, \quad \square\end{aligned}$$

Further let us note that

$$\max_{i=0,1,\dots,M} \left|g(x_i) - \hat{g}_{n_i}(x_i)\right| = \max_{i=0,1,\dots,M} \left|\sum_{j=1}^{n_i} \frac{\xi(x_i, \omega_j^{(i)}) - g(x_i)}{n_i}\right|. \tag{1.6}$$

Let $\bar{n} = \min(n_0, n_1, \dots, n_M)$, then

$$\begin{aligned}&\max_{i=0,1,\dots,M} \left|\sum_{j=1}^{n_i} \frac{\xi(x_i, \omega_j^{(i)}) - g(x_i)}{n_i}\right| \le \\ &\frac{1}{\sqrt{\bar{n}}} \max_{i=0,1,\dots,M} \left|\sum_{j=1}^{n_i} \frac{\xi(x_i, \omega_j^{(i)}) - g(x_i)}{\sqrt{n_i}}\right|.\end{aligned} \tag{1.7}$$

Suppose that variance function

$$\sigma^2(x) = \mathrm{M}\big(\xi(x,\omega) - g(x)\big)^2$$

is bounded on X by the value $D = d^2$, $d > 0$. Then

$$\begin{aligned}&\frac{1}{\sqrt{\bar{n}}} \max_{i=0,1,\dots,M} \left|\sum_{j=1}^{n_i} \frac{\xi(x_i, \omega_j^{(i)}) - g(x_i)}{\sqrt{n_i}}\right| \le \\ &\frac{d}{\sqrt{\bar{n}}} \max_{i=0,1,\dots,M} \left|\sum_{j=1}^{n_i} \frac{\xi(x_i, \omega_j^{(i)}) - g(x_i)}{\sigma(x_i)\sqrt{n_i}}\right|.\end{aligned} \tag{1.8}$$

By virtue of the central limit theorem [9] sums

$$\sum_{j=1}^{n_i} \frac{\xi(x_i, \omega_j^{(i)}) - g(x_i)}{\sigma(x_i)\sqrt{n_i}}$$

converge to standard normal stochastic values $\eta_i \simeq N(0,1)$ for $\bar{n} \to \infty$. So for the fixed confidence level $\bar{p}$ and for sufficiently big $\bar{n}$ we have

$$\delta_2 < \frac{d}{\sqrt{\bar{n}}} \cdot \tilde{T}_M, \tag{1.9}$$

where $\tilde{T}_M = \max_{i=0,1,\ldots,M} |\eta_i|$.

The following result is true:

Lemma 3. *If $\{\eta_i\}$ is the sequence of independent standard normal stochastic values, then the asymptotic distribution for $M \to \infty$ of the value*

$$T_M = \max\{\eta_0, \eta_1, \ldots, \eta_M\}$$

is the following:

$$\mathbf{P}\{a_M(T_M - b_M) \le y\} \to \exp(-e^{-y}) = \Lambda(y), \tag{1.10}$$

where

$$a_M = (2\ln M)^{1/2}; \quad b_M = (2\ln M)^{1/2} - \frac{1}{2}(2\ln M)^{-1/2}(\ln\ln M + \ln 4\pi).$$

Note that

$$t_M \stackrel{df}{=} \min\{\eta_0, \eta_1, \ldots, \eta_M\} = -\max\{-\eta_1, \ldots, -\eta_M\},$$

and for sufficiently big M we have

$$\tilde{T}_M = \max\{T_M, -t_M\}. \tag{1.11}$$

Then, taking into account the symmetry of distribution of standard normal stochastic values $\{\eta_i\}$ concerning nought and Theorem 1.8.3 from [10], it is possible to assert that

$$\mathbf{P}\{a_M(\tilde{T}_M - b_M) \le y\} \to \exp(-2e^{-y}) = \Lambda^2(y). \tag{1.12}$$

Thus from (1.2)–(1.12) we have the following result:

Theorem 1. *If*

(a) $g(x) \in C^{(2)}[a,b]$;

(b) $\sigma^2(x) = \mathbf{M}\big(\xi(x,\omega) - g(x)\big)^2 \le d^2$, $d > 0$;

then for the fixed confidence level $\bar{p}$ there exist real constants C_1 and $C_2(\bar{p})$, natural numbers $\hat{M}(\bar{p})$ and $\hat{N}(\bar{p})$, such that for every $\bar{n} > \hat{N}(\bar{p})$ and $M > \hat{M}(\bar{p})$ fulfil

$$\mathbf{P}\Big\{ \sup_{x\in[a,b]} |g(x) - \tilde{g}(x)| \le C_1 h^2 \|g(x)\|_{C^2[a,b]} + d/\sqrt{\bar{n}}\Big[(2\ln M)^{1/2} - (2\ln M)^{-1/2}\big(C_2 - \frac{1}{2}(\ln\ln M + \ln 4\pi)\big)\Big]\Big\} > \bar{p}. \tag{1.14}$$

Remark 1. If $g(x)$ belongs to another functional space (condition (a)), then the first addendum in the right-hand side of inequality (1.13) is also another (see [8, Chapter 2, Section 2]).

Remark 2. In papers [2, 5–7] the dependence from M for the error of the procedures of the type A and B was not taken into account. For the procedure B this ignorance of the dependence from M is justified (see, further Section 2). But in [2, 5–7] there is no exact definition what procedure is used: A or B. So the result of this paper can be regarded as a refinement of the results from [2, 5–7].

2. Convergence of procedure B

Here we can conduct the same reasoning as in Section 1 for the procedure A as far as inequality (1.5)

$$\delta_2 \le \max_{i=0,1,\dots,M} |g(x_i) - \bar{g}_n(x_i)|. \tag{2.1}$$

Taking into account the fact that in the procedure B we choose the same sample $\omega_1, \dots, \omega_n$ for getting all values $\bar{g}_n(x_i)$ we can use Lemma 1. If conditions of this lemma are fulfilled, then, with regard to the inequality

$$\mathbf{P}\{ \max_{i=0,1,\dots,M} |g(x_i) - \bar{g}_n(x_i)| \le C\cdot n^{-1/2}\} \ge \mathbf{P}\{ \sup_{x\in[a,b]} |g(x) - \bar{g}_n(x)| \le C\cdot n^{-1/2}\},$$

for sufficiently big n we have

$$\mathbf{P}\{\max_{i=0,1,\dots,M}|g(x_i)-\bar{g}_n(x_i)|\le C\cdot n^{-1/2}\}\gtrsim\mathbf{P}\{\sup_{x\in[a,b]}|\xi_0(x)|\le C\}. \quad (2.2)$$

Here the sign "$\gtrsim$" means "converges or more than".

As the process $\xi_0(x)$ is continuous with probability 1 on $[a,b]$, it is bounded with probability 1 on $[a,b]$ [11]. So from (1.2)–(1.4), (2.1), (2.2) we obtain the following result:

Theorem 2. *If*

(a) $g(x)\in C^{(2)}[a,b]$;

(b) $\sigma^2(x)=\mathbf{M}\big(\tilde{\xi}(x,\omega)\big)^2=\mathbf{M}\big(\xi(x,\omega)-g(x)\big)^2\le d^2$, $d>0$;

(c) *process $\tilde{\xi}$ is continuous on X together with it's derivative on x in mean of p-th degree, $p>1$;*

then for the fixed confidence level $\bar{p}$ there exist real constants A_1 and $A_2(\bar{p})$ and natural $\hat{N}(\bar{p})$, such that for every $n>\hat{N}(\bar{p})$ fulfil

$$\mathbf{P}\Big\{\sup_{x\in[a,b]}|g(x)-\overset{\approx}{g}(x)|\le A_1\cdot h^2\|g(x)\|_{C^2[a,b]}+A_2\cdot n^{-1/2}\Big\}>\bar{p}, \quad (2.3)$$

where

$$\overset{\approx}{g}(x)=\sum_{i=0}^{M}\bar{g}_n(x_i)\varphi_i(x).$$

For condition (a) of Theorem 2 and for the first addendum in the right-hand side of (2.3) Remark 1 from Section 1 is true.

Comparing Theorem 1 with Theorem 2 we note the following:

1. Estimate (2.3) is asymptotically "better" than corresponding estimate (1.13), but constant A_2 may be rather big for some particular cases and for moderately big M estimates (2.3) and (1.13) are comparable (see numerical results in Section 3).
2. For convergence of the procedure B the smoothness of the stochastic process (in general case – field) $\tilde{\xi}$ in mean of p-th degree is required, but for the procedure A it is not required. However, condition (c) of Theorem 2 can be done weaker (see, for example [12]), but in every case it is necessary for trajectories of the filed $\tilde{\xi}$ to belong to functional space, where the functional $\sup_{x\in X}u(x)$ is continuous, and the choice of such spaces in not very wide [2].

3. Numerical results

Consider the example of the function $g(x)$, where more cumbersome procedure A turns out to be comparable in cost [13] with the "depended tests" method. From Theorem 2 it is obvious that the speed of convergence in the procedure B depends on the values of derivative of integrand on x. So we choose the function $g(x)$ with the sufficiently big absolute value of derivative on any part of it's domain of definition. Let

$$g(x) = \int_0^1 f(x,y)dy = \int_0^1 e^{y/(x+a)}dy = (x+a)(e^{1/(x+a)} - 1),$$

where $x \in [0,1]$ and a is small parameter (in calculations $a = 0.2$). Function $f(x,y)$ has sufficiently big absolute value of derivative on x in the neighbourhood of the point $x = 0$. Construct the mesh

$$x_0 = 0, \quad x_i = i + 0.01, \quad i = \overline{1,20}, \quad x_i = 0.2 + 0.1 + (i-20), \quad i = \overline{21,28},$$

which "exaggerate" near zero.

We estimate the values of the function $g(x)$ in mesh points x_i for the procedure B according to the formula

$$\bar{g}_n(x_i) = \frac{1}{n}\sum_{j=1}^{n} f(x_i, \alpha_j),$$

where α_j is realization of stochastic quantity uniformly distributed on the segment [0,1].

Previously it was noted that the procedure A allows to vary the density of distribution of stochastic parameter ω in different mesh points x_i. On the other hand we must take into account that the separate processing of every mesh point and realization of stochastic quantities with complex densities of distribution lead to the drastic increase of computational expenditures. So we use the following modification of the procedure A. Divide the segment $[0,1]$ to parts

$$T_1 = [0, 0.05], \quad T_2 = [0.05, 0.1], \quad T_3 = [0.1, 0.15],$$
$$T_4 = [0.15, 0.2], \quad T_5 = [0.2, 1].$$

On every part T_k we use the "depended tests" method with the optimal [13] density

$$q_k(y) = \Big(\sum_i f^2(x_i, y)\Big)^{1/2} / I_k, \quad x_i \in T_k, \quad k = 1, 2, 3, 4,$$

here

$$I_k = \int_0^1 \Big(\sum_i f^2(x_i, y)\Big)^{1/2} dy.$$

On the fifth part T_5 we use uniform density. Integrals I_k, $k = 1, 2, 3, 4$ are calculated by the Monte Carlo method with uniform density.

In this case values of the function $g(x)$ in mesh point x_i from the part T_k are estimated by the formula

$$\hat{g}_{n_k}(x_i) = \frac{1}{n_k} \sum_{j=1}^{n_k} \frac{f(x_i, \theta_j)}{q_k(\theta_j)},$$

where θ_j is the realization of stochastic quantity distributed according to the density $q_k(y)$ (here we use "exclusion" method [13]).

Results of calculations are shown in Table.

Table. Comparison of the procedures A and B

Characteristics		Number of part			
		1	2	3	4
Number of tests	(A)	50	50	50	50
	(B)	1000	1000	1000	1000
Absolute error	(A)	0.775	0.153	0.046	0.054
	(B)	0.818	0.266	0.147	0.096
Relative error (%)	(A)	2.63	1.28	0.61	1.20
	(B)	2.77	2.23	1.96	1.76
Selective variance (σ_s^2)	(A)	3.90	0.59	0.14	0.06
	(B)	1420	149	43	17
Cost ($t \cdot \sigma_s^2$)	(A)	4.00	0.52	0.09	0.40
	(B)	38.34	4.05	1.16	0.46

t – computer time for realization of one stochastic value.

Thus, calculations prove the truth of the note in Section 2 that using the procedure A for processing with peculiarities of the function $g(x)$ one can obtain the gain in cost compared with the "depended tests" method.

References

[1] A.S. Frolov, N.N. Chentsov, Using of depended tests in the Monte Carlo method for obtaining of smooth curves, Abstracts of 6-th All Union Conference of Probability Theory and Mathematical Statistics, Vilnius, 1962, 425–437 (in Russian).

[2] A.V. Voytishek, Functional convergence of stochastic estimates and models, Ph.D. Thesis., Novosibirsk, 1989 (in Russian).

[3] S.M. Prigarin, Modelling of stochastic fields and solution of some stochastic problems of atmospheric optics by the Monte Carlo method, Ph.D. Thesis., Novosibirsk, 1990 (in Russian).

[4] A.V. Voytishek, S.M. Prigarin, On functional convergence of estimates and models in the Monte Carlo method, Journal Vychislitelnoy Matematiki i Matematicheskoy Fiziki, Vol. 32, No. 10, 1992, 1641–1651 (in Russian).

[5] G.A. Mikhailov, Cost of Monte Carlo methods for global estimation of solutions to multidimensional problems, Preprint No. 922., Comp. Cent., Sib. Branch, USSR Acad. Sci., Novosibirsk, 1990 (in Russian).

[6] G.A. Mikhailov, Cost of Monte Carlo methods for global estimation of solutions to multidimensional problems, Sov. J. Numer. Anal. Model., Vol. 6, No. 2, 1991, 131–150 (in Russian).

[7] G.A. Mikhailov, Minimization of computational costs of nonanalogue Monte Carlo methods, Series of Soviet and East European Mathematics, Vol. 5, Singapore World Scientific, Singapore, 1991 (in Russian).

[8] G.I. Marchuk, V.I. Agoshkov, Introduction to Projection-Mesh Methods, Nauka, Moscow, 1981 (in Russian).

[9] A.A. Borovkov, Probability Theory, Nauka, Moscow, 1986 (in Russian).

[10] M.R. Leadbetter, G.Lingren and H.Rootzen, Extremes and Related Properties of Random Sequences and Processes, Springer-Verlag, 1986.

[11] A.D. Ventsel, Course of the Theory of Stochastic Processes, Nauka, Moscow, 1975 (in Russian).

[12] S.M. Prigarin, Some applications of the Jain-Marcus theorem in statistical modelling, Methods of Statistical Modelling, Novosibirsk, Comp. Cent., Sib. Branch, USSR Acad. Sci., 1990, 22–28 (in Russian).

[13] S.M. Ermakov, G.A. Mikhailov, Statistical Modelling, Nauka, Moscow, 1982 (in Russian).