ISSN 1995-4239, Numerical Analysis and Applications, 2019, Vol. 12, No. 2, pp. 155–165. © Pleiades Publishing, Ltd., 2019.
Russian Text © The Author(s), 2019, published in Sibirskii Zhurnal Vychislitel'noi Matematiki, 2019, Vol. 22, No. 2, pp. 187–200.

Randomized Monte Carlo Algorithms for Problems with Random Parameters ("Double Randomization" Method)

G. A. Mikhailov[1,2*]

[1] *Institute of Computational Mathematics and Mathematical Geophysics, Siberian Branch, Russian Academy of Sciences, pr. Akad. Lavrent'eva 6, Novosibirsk, 630090 Russia*

[2] *Novosibirsk State University, ul. Pirogova 2, Novosibirsk, 630090 Russia*

Received June 21, 2018; accepted January 21, 2019

Abstract—Randomized Monte Carlo algorithms are constructed by a combination of a basic probabilistic model and its random parameters to investigate parametric distributions of linear functionals. An optimization of the algorithms with a statistical kernel estimator for the probability density is presented. A randomized projection algorithm for estimating a nonlinear functional distribution is formulated and applied to the investigation of the criticality fluctuations of a particle multiplication process in a random medium.

DOI: 10.1134/S1995423919020058

Keywords: *probabilistic model, statistical modeling, random parameter, randomized algorithm, double randomization method, random medium, splitting method, statistical kernel estimator.*

INTRODUCTION

Numerical Monte Carlo methods are usually constructed on the basis of natural or artificially formulated basic probabilistic models. They are especially efficient when the basic models have random parameters and a combined statistical simulation of basic and parametric distributions is used to estimate the sought-for quantities. In this, in fact, a product of the corresponding probabilistic spaces is realized (possibly, many times). In the present paper, such Monte Carlo algorithms are called randomized. The thus formulated "double randomization" method may be explained by considering the integral

$$J(\sigma)\int\limits_W g(w;\sigma)\mathrm{P}\,(dw;\sigma)$$

with a random (possibly functional) parameter σ. Here $\mathrm{P}(dw;\sigma)$ is a probabilistic measure in W with a parameter σ. Let us estimate the mean $\mathrm{E}J(\sigma)$ and the variance $\mathrm{D}J(\sigma)$. If a sufficiently exact estimator $J(\sigma)\approx\hat{J}(\sigma)$ is available, a numerically constructed sample $\{\sigma_i\}$ yields statistical estimates of the required quantities. However, this algorithm can be too costly for real problems. In this case it is reasonable to use double randomization: simulate for a given σ only a small number of points ω according to the distribution $\mathrm{P}(dw;\sigma)$, and then calculate and average the $g(\omega;\sigma)$-values thus obtained.

Optimization of this algorithm on the basis of a criterion of computational cost is considered in Section 1. A more complicated problem of optimizing a randomized algorithm is solved in Section 2 for the case of $J(x):=J(\sigma;x)$, and it is necessary to estimate the function $f(x)=\mathrm{E}J(\sigma;x)$.

The principle of constructing a double randomization algorithm to estimate $\mathrm{D}J(\sigma)$ is not so evident. Since $\mathrm{EE}(g^2(\omega;\sigma)|\sigma)\neq\mathrm{E}J^2(\sigma)$, at least two conditionally independent points ω (at fixed σ) must be simulated here (see Section 2). Note that such an algorithm cannot be constructed and justified only

* E-mail: gam@osmf.sscc.ru

on the basis of "physical considerations." The Monte Carlo algorithm presented in Sections 3 and 4 to estimate the distribution of a nonlinear function $\Phi(J(\sigma))$ using a power series approximation is even more complicated in this regard.

1. OPTIMIZATION OF ESTIMATION OF LINEAR FUNCTIONAL PROBABILISTIC MOMENTS

Consider linear functionals of the form

$$J_k(\sigma) = \int\limits_{R^m} \varphi(x;\sigma) h_k(x;\sigma)\, dx.$$

Here $x \in R^m$, σ is a random (and, possibly, functional) parameter of the problem ("random medium"), $\varphi(x;\sigma) \in L_1(R^m)$ is the solution to the problem with a parameter σ obtained by using a computer probabilistic model, that is, by a basic ensemble of trajectories $\{\Omega\}$ in the phase space R^m, $h_k(x;\sigma) \in L_\infty(R^m)$.

The Monte Carlo method is used to construct unbiased estimators $\xi_k(\Omega;\sigma)$ of the functionals $J_k(\sigma)$, that is, at a fixed σ we have $\mathrm{E}_\Omega \xi_k(\Omega;\sigma) = J_k(\sigma)$. To illustrate this scheme, one may consider the problem of transfer of particles—radiation quanta—with scattering and absorption through a medium with a random density $\sigma(r)$, $r \in R^3$ (see, for instance, [1, 2]). Here $\{\Omega\}$ is an ensemble of trajectories of quanta, which can be defined by a homogeneous Markov chain of collisions of quanta with elements of the substance. In this case the Monte Carlo method is used to average some functionals of the solution of the integro-differential equation of radiation transfer through a random medium. In what follows, randomized algorithms will be formulated for this problem of transfer of particles, although they can be used for any ensembles $\{\Omega\}$ and parameters σ that are implemented numerically.

The double randomization method for the estimation of probabilistic moments of the linear functionals $\{J_k(\sigma)\}$ is defined by using the following relation, which is easy to verify [3]:

$$\mathrm{E}\left[\prod_{k=1}^{s} J_k(\sigma)\right] = \mathrm{E}_{(\{\Omega_k\},\sigma)}\left[\prod_{k=1}^{s} \xi_k(\Omega_k;\sigma)\right]. \tag{1}$$

Here Ω_k $(k = 1, \ldots, s)$ are conditionally independent trajectories of radiation quanta constructed to realize a medium with density σ, and $\xi_k(\Omega;\sigma)$ are unbiased estimators of the functionals $J_k(\sigma)$, that is, $\mathrm{E}_\Omega \xi_k(\Omega;\sigma) = J_k(\sigma)$.

According to the rule of repeated averaging (that is, Fubini's theorem), relation (1) is realized as follows: first construct a realization of the random medium (that is, generally speaking, of the field σ); then construct, in this fixed medium, a series of independent (to be more exact, conditionally independent) trajectories $\{\Omega_k\}$, $k = 1, \ldots, s$, which contributes to statistical estimation of the quantity (1).

From a practical point of view, it is very important that in constructing an unbiased estimator of the moment (1) for a given realization of σ, one only has to construct s elementary estimates of the functional. Specifically, at $s = 1$ only one such estimate can be constructed, since $\mathrm{E}_{(\Omega_1,\sigma)} \xi(\Omega_1,\sigma) = \mathrm{E}\, J(\sigma)$. In this case the use of a series of conditionally independent trajectories can decrease the cost of the estimate by using the formulas of a "splitting method" (see, for instance, [1]). Note that when a trajectory Ω gets into a subdomain of the medium with already chosen values of σ, they cannot be chosen again, otherwise there emerges an "error of overchoice" [2]. Note also that, according to Fubini's theorem, the right-hand side of relation (1) remains finite with $\xi(\Omega_k;\sigma)$ replaced by $|\xi(\Omega_k;\sigma)|$. At $\xi(\Omega_k;\sigma) \geq 0$ relation (1) holds in any case. It is clear that with the above algorithm of estimating the moment (1) one can also construct estimators of moments of order $l < n$. In this case it is reasonable to use all subsequences of order l obtained from the sequence $\Omega_1, \ldots, \Omega_s$.

As noted above, the cost of the double randomization algorithm for estimating $\mathrm{E}\, J(\sigma) = \mathrm{E}_{(\Omega;\sigma)} \xi(\Omega;\sigma)$ can be decreased by using a series of conditionally independent trajectories $\{\Omega_k\}_{k=1,\ldots,n}$ constructed for a fixed σ, that is, by using the "splitting method" estimator [4]

$$\zeta_n = \frac{1}{n}\sum_{k=1}^{n}\xi(\Omega_k;\sigma).$$

In this case (see, for instance, [4])

$$\mathrm{D}\zeta_n = d_0 + d_1/n, \quad \text{where} \quad d_0 = \mathrm{D}_\sigma \mathrm{E}_\Omega \xi(\Omega;\sigma), \quad d_1 = \mathrm{E}_\sigma \mathrm{D}_\Omega \xi(\Omega;\sigma).$$

Here the mean number of computational operations is given by the formula $T_n = t_0 + nt_1$, where t_0 corresponds to a realization of σ, and t_1, to a realization of Ω. The value of n minimizing (up to the transition to the integer part) the cost $S_n = \mathrm{D}\zeta_n T_n$ (see [4]) is

$$n_{\mathrm{opt}} = \sqrt{\frac{t_0}{t_1}\frac{d_1}{d_0}}, \tag{2}$$

and (see, for instance, [1])

$$S_{n_{\mathrm{opt}}} = \left(\sqrt{t_0 d_0} + \sqrt{t_1 d_1}\right)^2 \le T_1. \tag{3}$$

To estimate the efficiency of this randomized algorithm with splitting, consider a model problem of transfer of particles with delta-scattering and absorption through a medium with a random optical thickness of absorption σ uniformly (that is, with a density of $1/\Delta$) distributed over the interval $(\Sigma, \Sigma + \Delta)$ at $\Sigma, \Delta \gg 1$. The randomized algorithm is constructed to calculate the probability of passage $p = \mathrm{E}p(\sigma)$ averaged over the realizations of the medium. In this case $p(\sigma) = \mathrm{e}^{-\sigma}$, and an asymptotic (at $\Sigma, \Delta \to \infty$) estimator $p = \mathrm{E}p(\sigma) \asymp \mathrm{e}^{-\Sigma}/\Delta$ is possible. Here the elementary estimator $\xi(\sigma;\Omega)$ is a "Bernoulli" one, that is, $\xi(\sigma;\Omega) = 1$ with probability $p(\sigma)$ and $\xi(\sigma;\Omega) = 0$ with probability $1 - p(\sigma)$. We have

$$d_0 \asymp p^2\frac{\Delta}{2}, \qquad d_1 \asymp p, \qquad n_{\mathrm{opt}} \asymp \sqrt{\frac{t_0}{t_1}\frac{2}{p\Delta}} \asymp \sqrt{2\frac{t_0}{t_1}\mathrm{e}^{\Sigma}},$$

$$S_{n_{\mathrm{opt}}} \asymp \left(\sqrt{\frac{\Delta}{2}p^2 t_0} + \sqrt{pt_1}\right)^2 \asymp pt_1 \asymp S_1\left(1 + \frac{t_0}{t_1}\right)^{-1},$$

since for the estimator without splitting $S_1 \asymp d_1(t_1 + t_0) \asymp p(t_1 + t_0)$. Thus, here the efficiency of splitting is determined by the ratio t_0/t_1. For real problems this ratio may take large values, since t_1 is bounded by the quantity C/p_{c} in direct modeling of absorption of a particle under collision with a given probability p_{c} [5]. In this case the ratio between d_0 and d_1 and, hence, between $S_{n_{\mathrm{opt}}}$ and S_1, may be close to the above-obtained model ratio. On the other hand, this estimate shows that it is not always reasonable to use $n > 1$.

In real problems it is difficult to obtain analytical estimates of the coefficients in formula (2). Therefore, as noted in [1], they should be estimated from preliminary statistical estimation of the quantities $\mathrm{D}\zeta_n, T_n$ for two values, $n = n_1, n_2$ (which should be as close as possible to n_{opt}), that is, by solving the following equations:

$$d_0 + \frac{d_1}{n_i} = \hat{\mathrm{D}}\zeta_{n_i}, \qquad t_0 + n_i t_i = \hat{T}_{n_i}, \quad i = 1, 2. \tag{4}$$

Any method based on (2) should be corrected in those cases where the realizations of σ are "completed" at successive modeling of the trajectories Ω_k [6]. Here the cost S_n depends on n nonlinearly (and in real problems, in a complicated way), so that the efficient ratio t_0/t_1 and, hence, n_{opt} decreases. However,

it may be assumed that in the representation $T_n = t_0 + n\varphi(n)t_1$ in a vicinity of $n_{\rm opt}$ the function $\varphi(n)$ changes to a lesser extent than n. Hence, Eq. (4) with the expression (2) can be efficiently used to correct the estimate of $n_{\rm opt}$. For instance, if after modeling of a trajectory Ω_k the number of operations for constructing Ω_{k+1} decreases by a comparatively small quantity t_2, we may set $\varphi(n) = 1 - t_2(n-1)/2$. In this case $n_{\rm opt}$ decreases as a result of the corresponding decrease in t_0. Note that the splitting optimization method being proposed can be used for the moments (1) by means of repeating the realization of the group of trajectories $\{\Omega_k\}$ independently n times.

2. OPTIMIZATION OF ESTIMATION OF AVERAGED DISTRIBUTIONS

First let $\varphi(x;\sigma)$ be a one-dimensional (in x) functional characteristic of the problem. It is necessary to estimate the averaged function $f(x) = \mathrm{E}\varphi(x;\sigma)$ by a numerical simulation of the parameter σ and the corresponding trajectories Ω. The Parzen–Rosenblatt universal statistical kernel estimator [7] with a rectangular ("uniform") kernel may be efficient for this purpose (see also [8]). It is constructed on the basis of statistical estimation of functionals of the form

$$J_\Delta = \int f(x')I_\Delta(x')\,dx' = \mathrm{E}\int \varphi(x';\sigma)I_\Delta(x')\,dx',$$

where $I_\Delta(x')$ is the indicator function of the interval $\Delta = \left(x - \frac{\delta}{2}, x + \frac{\delta}{2}\right)$. Assume that the problem allows constructing a Bernoulli estimator of the functionals $J_\Delta(\sigma) = \int \varphi(x';\sigma) \times I_\Delta(x')\,dx'$ by calculating the number of trajectories Ω that entered the interval Δ and remained there. In problems of particle transfer theory $\varphi(x';\sigma)$ is the stochastic density of the distribution of particles at the points of their "death," for instance, as a result of leaving the medium without coming back again. By virtue of relation (1) we have the following statistical estimator:

$$J_\Delta = \mathrm{E}J_\Delta(\sigma) \approx \frac{n_\Delta}{N},$$

where n_Δ is the number of trajectories of the particles that entered the interval Δ in a sample $\{(\sigma_i, \Omega_i)\}$ $(i = 1, \ldots, N)$, since $\mathrm{E}_\sigma\mathrm{E}_\Omega n_\Delta = NJ_\Delta$.

The mean squared error of the estimate $f(x) \approx n_\Delta/(N\delta)$ is (see, for instance, [8])

$$\begin{aligned}\varepsilon^2(x;N,\delta) &= \mathrm{E}\left[f(x) - \frac{n_\Delta}{N\delta}\right]^2 = \mathrm{D}\left(\frac{n_\Delta}{N\delta}\right) + \left(f(x) - \frac{J_\Delta}{\delta}\right)^2 \\ &\approx \frac{f(x)}{N\delta} + (f''(x))^2\frac{\delta^4}{576} \qquad (5)\end{aligned}$$

with a relative error decreasing to zero as $N \to \infty$, $\delta \to 0$, and $N\delta \to \infty$. By minimizing (5) according to [8], we obtain

$$\delta_0^5(x) = \frac{144\,f(x)}{N(f''(x))^2}, \qquad \varepsilon^2(x;N,\delta_0) \approx \frac{5}{4}\frac{f(x)}{\delta_0(x)N} \asymp N^{-\frac{4}{5}}.$$

Note that in [9] to estimate $f(x)$ and $f''(x)$ a quadratic approximation was used for the function $f(x)$ in the interval $\Delta_0 \supset \Delta$, which is the best in the L_2 metric, with Legendre polynomials of orders 0, 1, and 2. In [9] (as in [10]), a "microgrouped" sample with a spacing $h \ll \varepsilon/\max_x|f'(x)|$ was used to optimize the kernel estimator. In this case the average number of operations in the algorithm practically does not depend on δ.

Here our goal is to minimize the cost of the estimator according to (5). This is done by using the splitting method with a parameter n in Section 1. It is reasonable to average relation (5) with respect to x, that is (by analogy with [8]), to minimize the quantity

$$\varepsilon^2(N,\delta)=\int \varepsilon^2(x;N,\delta)\,dx=\frac{d}{N\delta}+f_0\delta^4,$$

where $d=\int f(x)\,dx$, $f_0=\int(f''(x))^2\,dx/576$, with N replaced by Nn.

Theorem 1. *The minimum of*

$$S^*(n,\delta)=\varepsilon^2(Nn,\delta)T_n=\left(\frac{d}{Nn\delta}+f_0\delta^4\right)(t_0+nt_1)$$

is reached at

$$\delta=\delta^*=\left(\frac{t_1}{t_0}\,\frac{d}{16f_0N}\right)^{\frac{1}{5}},\qquad n=n^*=\left(\frac{t_0}{t_1}\,\frac{d}{f_0}\,\frac{1}{(\delta^*)^5N}\right)^{\frac{1}{2}}=4\,\frac{t_0}{t_1},$$

and

$$\varepsilon^2(Nn^*,\delta^*)=\frac{5}{16}\,\frac{t_1\,d}{t_0\,N\,\delta^*}\asymp N^{-\frac{4}{5}}.$$

Proof. Since it is assumed that t_0 and t_1 do not depend on δ, the quantity δ^*, on simple transformations according to (3), is obtained from the equation

$$\frac{\partial}{\partial\delta}\left(\sqrt{f_0\delta^4t_0}+\sqrt{\frac{t_1d}{N\delta}}\right)^2=0.$$

Next, n^* is obtained from (2) at $d_0=f_0(\delta^*)^4$, $d_1=d/(N\delta^*)$. □

Recall that for the analog of Bernoulli's estimator of the functional J_Δ we use $d=\int f(x)\,dx$. For the unbiased weight modification in all expressions beginning with (5), the symbol $f(x)$ according to [9] is replaced by $f_{w^2}(x)$ if the auxiliary weight of a particle w is limited, that is, $w\le C<+\infty$. Here $f_{w^2}(x)$ is the distribution density of the squared weight. Note that absorption may not be "put into play" in the problem of transfer of particles, and the auxiliary weight is $\exp(-\tau_c)$, where τ_c is the "optical" trajectory length with respect to the absorption coefficient (see, for instance, [2, 5]).

Now consider an exact formulation of the problem of optimizing the randomized kernel estimator. For this we need

Lemma 1. *Assume that the mean squared error of a statistical functional estimator with parameter β is $\mathrm{D}(\beta)N^{-\alpha}$, and the average number of operations to calculate the sampled value of the estimator is $t(\beta)$, where N is the sample size. Then the optimal (according to the criterion of computational cost) value of β is*

$$\arg\min_\beta \mathrm{D}^{1/\alpha}(\beta)t(\beta)=\arg\min_\beta \mathrm{D}(\beta)t^\alpha(\beta).$$

Proof. By definition (see [4]) the computational cost is the mean number S of computational operations required to achieve a given error ε. According to the conditions of the lemma $\varepsilon^2=\mathrm{D}(\beta)N^{-\alpha}$, whence $N=\mathrm{D}^{1/\alpha}(\beta)\varepsilon^{-2/\alpha}$ and

$$S(\beta)=\mathrm{D}^{1/\alpha}(\beta)t(\beta)\varepsilon^{-2/\alpha}.$$

□

Thus, we have the following theorem:

Theorem 2. *The cost of the randomized kernel estimator of the function $f(x)$ is determined asymptotically over N by the parameters n_0^* and δ_0^* minimizing the quantity*

$$\varepsilon^2(Nn,\delta)(t_0+nt_1)^{4/5}.$$

In this case the asymptotics of $\varepsilon^2(Nn_0^*,\delta_0^*) \asymp N^{-4/5}$ is preserved.

The problem of minimization presented in Theorem 2 can be solved numerically. As a first approximation $n_0^* \approx n^*$, $\delta_0^* \approx \delta^*$, since

$$S_0^*(n,\delta) \le S^*(n,\delta)t_0^{-\frac{1}{5}}$$

and $(t_0+nt_1)^{\frac{1}{5}}$ is a weakly varying function of the argument n.

Additional studies have shown that $S(1,\delta_0)/S(n^*,\delta^*) = 0.8(1+t_0/t_1)$, and the use of n_0^* and δ_0^* can increase this estimate only up to $1+t_0/t_1$ (see the example in Section 1).

Note that if $T_n = t_0+nt_1$ nonlinearly depends on n (the case considered in Section 1), n^* and, hence, the ratio t_0/t_1, can be corrected by numerically optimizing the splitting algorithm for the functional $J = \int f(x)\,dx$ on the basis of relations (4).

Consider a generalization of the results obtained for the multidimensional case, that is, for $x \in R^m$. As in the one-dimensional case, for the corresponding kernel estimator with a "hypercubic" kernel (with side length δ), according to [11] we have the following expressions:

$$\varepsilon_m^2(x,N,\delta) \approx \frac{f(x)}{N\delta^m} + F_m(x)\delta^4, \quad \varepsilon^2(N,\delta) = \frac{d}{N\delta^m} + f_0^{(m)}\delta^4,$$

where $F_m(x) = \Big(\sum\limits_{i=1}^{m} f_i''(x)\Big)^2 \big/ 576$, $f_0^{(m)} = \int F_m(x)\,dx$.

Hence, we have

$$\delta_0^{m+4}(x) = \frac{mf(x)}{4NF_m(x)}, \qquad \varepsilon_m^2(x;N,\delta_0) = \delta_0^{-m}\frac{f(x)}{N}\,\frac{4+m}{4} \asymp N^{-\frac{4}{4+m}}.$$

By analogy with the above, for the algorithm with splitting we have

$$\delta_m^* = \left(\frac{m^2}{16}\,\frac{t_1 d}{t_0 f_0^{(m)}}\,\frac{1}{N}\right)^{\frac{1}{m+4}}, \quad n^* = \frac{4}{m}\,\frac{t_0}{t_1}, \quad \varepsilon^2(Nn^*,\delta^*) = \frac{t_1 dm(4+m)}{16t_0N(\delta^*)^m} \asymp N^{-\frac{4}{4+m}},$$

and asymptotically exact optimization is obtained by minimizing the quantity

$$\varepsilon_m^2(Nn,\delta)(t_0+nt_1)^{4/(4+m)}.$$

This generalization corresponds to the case of vector coordinates $x = (x^{(1)},\dots,x^{(m)})$ "on the same scale." Otherwise, as usual, some scaling $\delta^{(k)} = c_k\delta$ $(k=1,\dots,m)$ should be used. In this case, in the relations obtained $f(x)$ is replaced by $f(x)\big/\prod\limits_{k=1}^{m} c_k$, and $f_i''(x)$ is replaced by $c_i^2 f''(x)$. Here $S(1,\delta_0)/S(n^*,\delta^*) = 4(4+m)^{-1}(1+t_0/t_1)$.

Finally note that according to (1) the randomized projection method proposed by Chentsov for numerical statistical modeling [12] can be extended to the estimation of averaged distributions. However, calculations show that this method can be efficient only in the case of a sufficiently smooth one-dimensional function or its ratio to some additional probability density (see, for instance, [13]). In this case any multidimensional generalization is rather difficult.

3. ESTIMATION OF NONLINEAR FUNCTIONAL MOMENTS

3.1. Here we consider nonlinear random functionals of the form

$$L(\sigma) = \Phi(J(\sigma)), \tag{6}$$

where Φ is a nonlinear function, and $J(\sigma)$ is the linear functional considered in Sections 1 and 2.

Randomized algorithms for estimating the moments $\mathrm{E}L^i(\sigma)$ $(i = 1, \ldots, m)$ can be constructed by analogy with [14] (in which the function $\Phi(J) = \sqrt[n]{J}$ was considered) by using the power series approximation

$$\Phi^i(J(\sigma)) \approx \Phi_n^{(i)}(J(\sigma)) = \sum_{s=1}^{n} a_s^{(i)} \big(J(\sigma) - \hat{J}\big)^s, \tag{7}$$

where

$$a_s^{(i)} = \frac{1}{s!} \frac{d^s \Phi^i(x)}{dx^s} \bigg|_{x=\hat{J}},$$

and $\hat{J}$ is a preliminarily constructed sufficiently exact statistical estimator of the quantity $\mathrm{E}J$. Note that here the estimator $\hat{J}$ is considered as a deterministic one. This does not violate the convergence over n in (7) and allows obtaining the variance of the statistical estimator of $\mathrm{E}\Phi_n^{(i)}$ in the ordinary way. From (7) we obtain

$$\mathrm{E}L^i(\sigma) \approx \sum_{s=1}^{n} a_s^{(i)} \mathrm{E}\big(J(\sigma) - \hat{J}\big)^s.$$

Randomized estimators $\zeta_n^{(s)}$ of the moments $\mathrm{E}(J(\sigma) - \hat{J})^s$ are constructed according to (1) by using a basic series of conditionally independent trajectories $\{\Omega_k\}$ $(k = 1, \ldots, n)$, with $\xi_k(\Omega_k; \sigma)$ in (1) replaced by $\hat{\xi}(\Omega_k; \sigma) = \xi_k(\Omega_k; \sigma) - \hat{J}$. In this case it is reasonable to use all the various sequences of the order s obtained from the basic series.

Let $Q_s = \big\{F_1^{(s)}, \ldots, F_q^{(s)}\big\} = \{\{l_1, \ldots, l_s\}_j\}$ be the set of various series of indices, that is, $F_j^{(s)} \subset \{1, \ldots, n\}$. The number of $F_j^{(s)}$ in Q_s is equal to the number of combinations $q = C_n^s$. Hence, we can use the following unbiased estimator of the moment $\mathrm{E}(J(\sigma) - \hat{J})^s$:

$$\zeta_n^{(s)} = \frac{1}{C_n^s} \sum_{j=1}^{C_n^s} \prod_{k \in F_j^{(s)}} \hat{\xi}(\Omega_k, \sigma). \tag{8}$$

Thus, we obtain the estimator

$$\mathrm{E}L^i(\sigma) \approx \mathrm{E} \sum_{s=1}^{n} a_s^{(i)} \zeta_n^{(s)}, \tag{9}$$

which can be realized by double randomization. As shown in Section 1, first a realization of the parameters σ is constructed. Then a series of trajectories $\{\Omega_k\}$ $(k = 1, \ldots, n)$ is simulated, and an elementary sample value of the statistical estimator of (9) is calculated. Averaging of these values and their squares yields a statistical estimator of the moment $\mathrm{E}L^i(\sigma)$ and its root-mean-square error.

3.2. Although the terms in (8) are not independent, the use of this expression (as shown by calculations in [14]) can considerably decrease the statistical error of the resulting estimator in contrast to using only one set, $\Omega_1, \ldots, \Omega_s$. This can be explained in the following way. It is easy to see that if $J(\sigma) = \mathrm{E}_\Omega \xi(\Omega;\sigma) \equiv J$, at $\hat{J} = J$, the terms in the sum of (8) are pairwise conditionally uncorrelated. Therefore, under the conditions when small perturbations (assumed in [14]) can be applied, that is, when $J(\sigma) \approx J \,\forall\, \sigma$, we have

$$\mathrm{D}\big(\zeta_n^{(s)} \,\big|\, \sigma\big) \approx \frac{1}{C_n^s} \mathrm{D}\Bigg(\prod_{k \in F_1^{(s)}} \hat{\xi}(\Omega_k, \sigma) \Bigg).$$

Note that in the equality

$$\mathrm{D}\big(\zeta_n^{(s)}\big) = \mathrm{DE}\big(\zeta_n^{(s)} \,\big|\, \sigma\big) + \mathrm{ED}\big(\zeta_n^{(s)} \,\big|\, \sigma\big),$$

under the conditions of applicability of small perturbations the second term usually dominates.

Therefore, taking into account the above discussion, to implement the estimator (9) it is reasonable to simulate $(n+1)$ trajectories, that is, use the estimator of $\Phi_{n+1}^{(i)}$ without the last term, for which the sum in (8) is unity.

3.3. In [14] Lotova constructed an efficient algorithm for a successive calculation of the quantities $\zeta_n^{(1)}, \ldots, \zeta_n^{(s)}$ for a sequence of the partial sums:

$$\begin{aligned} S_k^{(1)} &= \sum_{j=k}^{n} \hat{\xi}(\Omega_j, \sigma), \qquad S_k^{(2)} = \sum_{j=k}^{n-1} \hat{\xi}(\Omega_j, \sigma) S_{j+1}^{(1)}, \\ S_k^{(s)} &= \sum_{j=i}^{n+1-s} \hat{\xi}(\Omega_j, \sigma) S_{j+1}^{(s-1)}, \quad k = 1, \ldots, (n+1-s). \end{aligned}$$

In [14] mathematical induction was used to show that

$$S_1^{(s)} = \sum_{j=1}^{C_n^s} \prod_{k \in F_j^{(s)}} \hat{\xi}(\Omega_k, \sigma), \qquad \zeta_n^{(s)} = S_1^{(s)} / C_n^s.$$

Note that the trajectories $\Omega_k, \ldots, \Omega_n$ are used to calculate $S_k^{(n)}$.

A logically simpler algorithm consists in a successive construction of products from (8) in the following way: as soon as a trajectory Ω_i is simulated, the sequence of products of the order $s+1$ is supplemented by the products of the order s multiplied by $\hat{\xi}(\Omega_i, \sigma)$ $(s = 1, \ldots, i-1)$. Finally, as soon as the basic series $\{\Omega_k\}$ is simulated, all obtained sequences (in contrast to the algorithm from [14]) are summed up independently. This greatly increases the cost of the simple algorithm. It is evident that the number of final additions in this algorithm is $\sum\limits_{s=0}^{n} C_n^s = 2^n$, and the number of multiplications is determined by the following relation:

$$\sum_{s=1}^{n} (s-1) C_n^s = \sum_{s=0}^{n} s C_n^s - 2^n + 1 = [(1+x)^n]'_{x=1} - 2^n + 1 = (n-2)2^{n-1} + 1.$$

4. ESTIMATION OF DISTRIBUTION FUNCTIONS OF NONLINEAR FUNCTIONALS

4.1. The probability $\mathrm{P}(\Phi(J(\sigma)) < x)$ cannot be estimated by using randomization. Therefore, here it is reasonable to use some global estimator of the distribution. However, it should be taken into account that practically (especially under the conditions of small perturbations) a Gaussian approximation of the distribution of the functional $J(\sigma)$ may be efficient with the use of a corresponding orthogonal decomposition. Hence, it is appropriate to use the relation

$$\mathrm{P}\left(\Phi(J(\sigma)) < x\right) = \mathrm{P}\left(J(\sigma) < \Phi^{-1}(x)\right),$$

and, for the distribution density $f(y)$ of the random quantity $J(\sigma)$, the approximation

$$f(y) \approx f_M(y) = f_{\mathrm{norm}}(y; a, d^2)\left(\psi_0 + \sum_{i=1}^{M} c_i \psi_i([y-a]/d)\right), \quad \psi_0 \equiv 1. \tag{10}$$

Here a and d^2 are estimates of the mean and variance of the random quantity $J(\sigma)$, $\{\psi_i\}$ are orthonormal Hermite polynomials with the weight

$$f_{\mathrm{norm}}(y; a, d^2) = \frac{\exp(-(y-a)^2/(2d^2))}{\sqrt{2\pi d^2}}.$$

In this case $c_i = \mathrm{E}\psi_i([J(\sigma) - a]/d)$. Therefore, it is reasonable to use (10) with previously estimated parameters. This is confirmed by numerical testing of estimates of the probability $\mathrm{P} = \mathrm{P}(J(\sigma) > 1)$ obtained on the basis of (10) for a model problem considered in [14].

It is difficult to use the estimator (10) due to its slow convergence over M at low smoothness of the density $f(y)$. A heuristic criterion for choosing M may be that of coincidence, within the required accuracy, of estimates of the sought-for probability for M and $M+1$.

It is also reasonable to determine the approximation order in (10) by using sufficiently accurate numerical-analytical estimates of the values of J for some simplified model of the problem. In [14], the diffusion approximation for homogenized realizations of the medium was used for this purpose. The calculations have shown that to retain the arithmetic accuracy and to estimate the variances $\mathrm{D}\tilde{c}_i$ the corresponding sampled values (and their squares) must be averaged; the estimates of the moments $\mathrm{E}J^s(\sigma)$ $(i = 1, \ldots, M)$ must not be used for this. In the case of statistical modeling of the process, the required result is obtained by randomized estimation of c_i, $i = 1, \ldots, M$, with M conditionally independent trajectories.

4.2. Paper [14] studies the fluctuations of the coefficient $k(\sigma)$ of particle multiplication in a medium with random density $\sigma(r)$, $r \in R^3$. The coefficient k is the leading eigenvalue of the corresponding positive integral operator: $kf = Kf$ (see, for instance, [15]). The standard Monte Carlo algorithms for the estimation of k, in fact, realize Kellogg's method (see, for instance, [15, 16]), that is, the limiting relation

$$k = \lim_{m\to\infty} \frac{(K^{m+1} f_0, h)}{(K^m f_0, h)}. \tag{11}$$

As shown in [14], such algorithms are not efficient for randomized estimation of the quantities $\mathrm{E}k(\sigma)$, $\mathrm{D}k(\sigma)$, and $\mathrm{P}(k(\sigma) > 1)$. Therefore, the following limiting relation was proposed in [14]:

$$k = \lim_{m\to\infty} k_m, \quad k_m = \sqrt[m]{(K^m f_0, h)}. \tag{12}$$

Paper [14] provides a proof of the limit (12), by analogy with the proof of the limit (11) presented in [15]. The value of m in the approximate formula $k \approx k_m$ is selected on the basis of preliminary calculations, possibly using a semianalytic estimator [14]. In (12) the operator K depends on σ, and one can assume

that $(K^m f_0, h) = J(\sigma)$, since to estimate the moments $\mathrm{E}(K^m f_0, h)^s$, one can use randomized Monte Carlo algorithms (see Section 1). Thus, relation (12) is reduced to the form (6). Hence, the above is a generalization and an explanation of the results of [14].

Note that [17] considered the various aspects of construction and investigation of algorithms to estimate some probabilistic moments of solving the equation of radiation transfer through a random medium. Realistic computational models were presented for exponentially correlated nonnegative homogeneous isotropic random fields of density $\sigma(r)$ of the medium. Their realizations are close to continuous ones, and the one-dimensional distributions are close to Gaussian ones. The realistic character of the models was confirmed by the illumination field for radiation passing through the medium calculated on this basis.

These models have the form

$$\sigma_n(r) = \sum_{i=1}^{n} \sigma_i^{(n)}(r),$$

where $\{\sigma_i^{(n)}\}$ are independent realizations of an "elementary mosaic" nonnegative random field with an exponential correlation function and one-dimensional "beta"-distribution. Boundedness of the field $\sigma_i^{(n)}$ allows simulating the free path length l of a radiation quantum by the "maximum cross section" (also called "delta-scattering") method [18]. In this case the free path length in the corresponding realistic field $\sigma(r)$ is obtained by the formula $l = \min_i l_i$. As shown in [19], this method of modeling the free path length (including the "delta-scattering" method) follows directly from the invariance of Poisson random point flows with respect to the operations of combination and random "Bernoulli" thinning. Note that corresponding algorithms for the above realistic models of $\sigma(r)$ were developed in [6]. These models and algorithms turned out to be useful, in particular, for solving the practically important problem of averaging the equation of radiation transfer through a random medium [2].

ACKNOWLEDGMENTS

The author would like to thank his former PhD students G.Z. Lotova, A.Yu. Ambos, and others for active participation in the development and realization of randomized algorithms for the Monte Carlo method.

FUNDING

This work was performed within the framework of the State Order for the SB RAS Institute of Computational Mathematics and Mathematical Geophysics (project no. 0315-2016-0002) and was partly supported by the Russian Foundation for Basic Research (project nos. 16-01-00530, 17-01-00823, and 18-01-00356).

REFERENCES

1. Mikhailov, G.A., *Optimization of Weighted Monte Carlo Methods*, Springer-Verlag, 1992.
2. Ambos, A.Yu. and Mikhailov, G.A., Effective Averaging of Stochastic Radiative Models Based on Monte Carlo Simulation, *Comput. Math. Math. Phys.*, 2016, vol. 56, no. 5, pp. 881–893.
3. Mikhailov, G.A., Efficient Algorithms of the Monte Carlo Method for Computing the Correlation Characteristics of Conditional Mathematical Expectations, *USSR Comput. Math. Math. Phys.*, 1977, vol. 17, no. 1, pp. 244–247.
4. Mikhailov, G.A. and Voitishek, A.V., *Chislennoe statisticheskoe modelirovanie. Metody Monte-Karlo* (Numerical Statistical Modeling: Monte Carlo Methods), Moscow: Akademiya, 2006.
5. Marchuk, G.I., Mikhailov, G.A., Nazaraliev, M.A., Darbinjan, R.A., Kargin, B.A., and Elepov, B.S., *The Monte Carlo Methods in Atmospheric Optics*, Berlin: Springer, 1980.
6. Ambos, A.Yu., Numerical Models of Mosaic Homogeneous Isotropic Random Fields and Problems of Radiative Transfer, *Num. An. Appl.*, 2016, vol. 9, no. 1, pp. 12–23.
7. Parsen, E., On Estimation of a Probability Density Function and Mode, *Ann. Math. Stat.*, 1962, no. 35, pp. 1065–1076.

8. Borovkov, A.A., *Matematicheskaya statistika* (Mathematical Statistics), Novosibirsk: Institute of Mathematics SB RAS, 1997.
9. Mikhailov, G.A., Prigarin, S.M., and Rozhenko, S.A., Comparative Analysis of Vector Algorithms for Statistical Modeling of Radiative Transfer Process, *Russ. J. Num. An. Math. Model.*, 2018, vol. 33, no. 4, pp. 220–229.
10. Lotova, G.Z., Monte Carlo Algorithms for Calculation of Diffusive Characteristics of an Electron Avalanche in Gases, *Russ. J. Num. An. Math. Model.*, 2011, vol. 31, no. 6, pp. 369–377.
11. Epanechnikov, V.A., Non-Parametric Estimation of a Multivariate Probability Density, *Theory Prob. Appl.*, 1969, vol. 14, no. 1, pp. 153–158.
12. Chentsov, N.N., *Statisticheskie reshayushchie pravila i optimal'nye vyvody* (Statistical Decision Rules and Optimal Inference), Moscow: Nauka, 1972.
13. Mikhailov, G.A., Tracheva, N.V., and Ukhinov, S.A., Randomized Projection Method for Estimating Angular Distributions of Polarized Radiation Based on Numerical Statistical Modeling, *Comput. Math. Math. Phys.*, 2016, vol. 56, no. 9, pp. 1540–1550.
14. Mikhailov, G.A. and Lotova, G.Z., Monte Carlo Methods for Estimating the Probability Distributions of Criticality Parameters of Particle Transport in a Randomly Perturbed Medium, *Comput. Math. Math. Phys.*, 2018, vol. 58, no. 11, pp. 1828–1837.
15. Vladimirov, V.S., On Application of the Monte Carlo Method to Find the Least Characteristic Value and the Corresponding Eigenfunction of a Linear Integral Equation, *Teor. Ver. Prim.*, 1956, vol. 1, no. 1, pp. 113–130.
16. Ermakov, S.M. and Mikhailov, G.A., *Statisticheskoe modelirovanie* (Statistical Modeling), Moscow: Nauka, 1982.
17. Ambos, A.Yu. and Mikhailov, G.A., Monte Carlo Estimation of Functional Characteristics of Field Intensity of Radiation Passing through a Random Medium, *Num. An. Appl.*, 2018, vol. 11, no. 4, pp. 279–292.
18. Woodcock, E., Murphy, T., Hemmings, P., and Longworth, S., Techniques Used in the GEM Code for Monte Carlo Neutronics Calculations in Reactors and Other Systems of Complex Geometry, *Proc. Conf. on Applications of Computing Methods to Reactor Problems*, 1965, p. 557.
19. Averina, T.A. and Mikhailov, G.A., Algorithms for Exact and Approximate Statistical Simulation of Poisson Ensembles, *Comput. Math. Math. Phys.*, 2010, vol. 50, no. 6, pp. 951–962.
20. Belyaev, Yu.K., *Veroyatnost' i matematicheskaya statistika. Entsiklopediya* (Probability and Mathematical Statistics: Encyclopedia), Moscow: BRE, 1999.