ISSN 1063-4541, Vestnik St. Petersburg University, Mathematics, 2021, Vol. 54, No. 1, pp. 28–38. © Pleiades Publishing, Ltd., 2021.
Russian Text © The Author(s), 2021, published in Vestnik Sankt-Peterburgskogo Universiteta: Matematika, Mekhanika, Astronomiya, 2021, Vol. 54, No. 1, pp. 37–48.

MATHEMATICS

The Monte Carlo Method for Solving Large Systems of Linear Ordinary Differential Equations

S. M. Ermakov[a],* and M. G. Smilovitskiy[a],**

[a] *St. Petersburg State University, St. Petersburg, 199034 Russia*

* *e-mail: sergej.ermakov@gmail.com*

** *e-mail: smmxgr@gmail.com*

Received June 3, 2020; revised July 27, 2020; accepted September 17, 2020

Abstract—The Monte Carlo method to solve the Cauchy problem for large systems of linear differential equations is proposed in this paper. Firstly, a quick overview of previously obtained results from applying the approach towards the Fredholm-type integral equations is made. In the main part of the paper, the method is applied towards a linear ODE system that is transformed into an equivalent system of the Volterra-type integral equations, which makes it possible to remove the limitations due to the conditions of convergence of the majorant series. The following key theorems are stated. Theorem 1 provides the necessary compliance conditions that should be imposed upon the transition propability and initial distribution densities that initiate the corresponding Markov chain, for which equality between the mathematical expectation of the estimate and the functional of interest would hold. Theorem 2 formulates the equation that governs the estimate's variance. Theorem 3 states the Markov chain parameters that minimize the variance of the estimate of the functional. Proofs are given for all three theorems. In the practical part of this paper, the proposed method is used to solve a linear ODE system that describes a closed queueing system of ten conventional machines and seven conventional service persons. The solutions are obtained for systems with both constant and time-dependent matrices of coefficients, where the machine breakdown intensity is time dependent. In addition, the solutions obtained by the Monte Carlo and Runge–Kutta methods are compared. The results are presented in the corresponding tables.

Keywords: Monte Carlo method, ODE systems, integral equation, queueing problems, optimal density, unbiased estimate, statistical modeling

DOI: 10.1134/S1063454121010064

1. INTRODUCTION

The Monte Carlo method applied to solve integral equations and large systems of linear algebraic equations is described in sufficient detail in the literature, for example, [1, 2] and [3, 4]. However, this is not the case for solving the Cauchy problems for large systems of ordinary differential equations, although certain classes of linear systems are of considerable interest for applying the Monte Carlo method. As an example, we consider the queueing problems that can be described by systems with a very large or even infinite number of equations. In the simplest case of Poisson systems, it is easy to obtain the result when $t \to \infty$, and write the solution in analytical form for the transient mode. It is convenient to use the direct simulation modeling for many problems. In all cases, however, there are some problems, whose solution by these methods is quite time-consuming and unfeasible. Thus, for the transient mode, even in the simplest case, the solution is expressed by an infinite series of values of the Bessel functions, and applying the simulation modeling to solve the problems of calculating small probabilities can be prohibitively time consuming.

These considerations are sufficient to consider the Monte Carlo methods other than simulation to solve the following system:

$$\begin{gathered}\frac{d}{dt}Y(t) = A(t)Y(t) + F(t), \quad Y(0) = Y_0, \\ A(t) = \left\|a_{ij}(t)\right\|_{i,j=1}^{n}, \quad Y(t) = \left\|y_1(t),\ldots,y_n(t)\right\|^T, \quad F(t) = \left\|f_1(t),\ldots,f_n(t)\right\|^T,\end{gathered} \tag{1}$$

for the values of t from the interval $[0, T]$.

Some approaches to the construction of algorithms using the Monte Carlo method were discussed earlier in [5, 6]. In [5], stochastic differential equations were solved, and in [6], the general case of a nonlinear problem was considered.

In this paper, the linear case is considered, albeit from a general perspective; the issues of minimizing the variance of the simplest estimate are amply investigated. Numerical examples from the queueing theory are considered.

There are many ways to reduce problem (1) to the integral form, and the Monte Carlo methods for solving integral equations are well developed. For this reason, we briefly recall the simplest facts concerning the Monte Carlo method application to solve integral equations of the second kind.

2. SOLUTION OF THE FREDHOLM INTEGRAL EQUATIONS BY THE MONTE CARLO METHOD

Let the following equation with respect to $\varphi(x)$ be set:

$$\varphi(x) = \int k(x, y)\varphi(y)\mu(dy) + f(x); \tag{2}$$

it holds for carrier $\mathbf{X}$ of probability measure μ and for the set $\mathbf{H}$ of functions $h(x)$ such that the integral $(\varphi, h) = \int \varphi(x)h(x)\mu(dx)$ exists. Let the method of successive approximations converge (we mean the normwise convergence in some normalized space, see. ([2], p. 272–275)):

$$\tilde{\varphi}_{m+1}(x) = \int |k(x, y)|\,\tilde{\varphi}_m(y)\mu(dy) + |f(x)|, \quad \tilde{\varphi}_0 = |f(x)|; \quad m = 0,1,\ldots. \tag{3}$$

Then the following representation is true:

$$(\varphi, h) = (f, h) + \sum_{l=1}^{\infty} \int \mu_0 \ldots \int \mu_l h_0, k_{0,1} \ldots k_{l-1,l} f_l; \tag{4}$$

here we designate $\mu_r = \mu(dx_r)$, $h_r = h(x_r), f_r = f(x_r)$, $k_{r-1,\,r} = k(x_{r-1}, x_r)$, $r = 0, 1, \ldots, l$.

We choose the Markov chain with the initial distribution $p^0(x)$ and transition probability density $p(x, y)$, satisfying the condition

$$\int p(x, y)\mu(dy) = 1 - g(x), \quad 0 \le g(x) < 1 \tag{5}$$

and the compliance condition

$$\begin{aligned} p^0(x) > 0, &\quad \text{if} \quad h(x) \ne 0, \\ g(x) > 0, &\quad \text{if} \quad f(x) \ne 0, \\ p(x, y) > 0, &\quad \text{if} \quad k(x, y) \ne 0. \end{aligned} \tag{6}$$

Hereinafter, we imply $x, y \in \mathbf{X}$.

We simulate the Markov chain and calculate the following function along its trajectories $x_0 \to x_1 \to \ldots \to x_\tau$:

$$J_\tau(x_0, \ldots, x_\tau) = \frac{h(x_0)k(x_0, x_1)\ldots k(x_{\tau-1}x_\tau)f(x_\tau)}{p^0(x_0)p(x_0, x_1)\ldots p(x_{\tau-1}x_\tau)g(x_\tau)}. \tag{7}$$

It is assumed that almost all trajectories are finite. This is true, at least, when $g(x) > 0$.

It is easy to show (see ([1], pp. 94–102)) that when the convergence conditions (3) and compliance conditions (6) are satisfied the following equality is true:

$$\mathbf{E}J_\tau(x_0, \ldots, x_\tau) = (h, \varphi). \tag{8}$$

To do this, it is sufficient to take into consideration that the probability density of a random trajectory is $p^0(x_0)p(x_0, x_1)$, ..., $p(x_{\tau-1}, x_\tau)g(x_\tau)$, and use the definition of mathematical expectation.

There are some other estimates of the process' trajectories (5); however, estimate (7) is one of the simplest in terms of computation; it is very easy to obtain the analytical expression of its variance and to define the Markov chain satisfying (3) and (6), for which the variance has the minimum value.

This optimal Markov chain is homogeneous and can be set by

$$\begin{aligned} &\text{initial distribution} \quad \tilde{p}^0(x) = \frac{|h(x)| \cdot \overline{\varphi}(x)}{(|h|, |\overline{\varphi}|)}, \quad \text{and} \\ &\text{transition probability density} \quad \tilde{p}(x,y) = \frac{|k(x,y)| \cdot \overline{\varphi}(y)}{\overline{\varphi}(x)}, \end{aligned} \tag{9}$$

where $\overline{\varphi}$ is limit (3) at $m \to \infty$ in (3).

It is important to note for the following discussion that, in the case of nonnegative h, f, and k, the optimal value of the estimate variance (8) turns out to be zero. Although expressions (9) determining the optimal Markov chain depend on the exact value of $\overline{\varphi}$, they allow us to draw some conclusions concerning the choice of a Markov chain when solving real problems. Thus, we see that the factor $|k(x, y)|$ should be included in the expression for the density $p(x, y)$.

3. SOLVING THE CAUCHY PROBLEM FOR A SYSTEM OF ORDINARY DIFFERENTIAL EQUATIONS BY THE MONTE CARLO METHOD

Let us return to Eq. (1). We assume that the elements of the matrix $A(t)$ and vector $\tilde{F}(t)$ are bounded functions of t on the real axis. Hence, the fulfillment of the Lipschitz conditions for the right-hand side of (1) and, consequently, the existence and uniqueness of the solution on any finite interval $[0, T]$ follow. Then we can rewrite Eq. (1) in the form

$$Y(t) = Y(0) + \int_0^t F(\tau)d\tau + \int_0^t A(\tau)Y(\tau)d\tau \tag{10}$$

or

$$Y(t) = \int_0^T A(\theta)E(t-\theta)Y(\theta)d\theta + \tilde{F}(t), \tag{11}$$

where $F(t) = Y_0 + \int_0^t F(\tau)d\tau$, $E(x) = \begin{cases} 1, & \text{if} \quad x \geq 0, \\ 0, & \text{if} \quad x < 0, \end{cases}$ $t \in [0, T]$.

By analogy, we can choose a Markov chain that satisfies the compliance conditions

$$\begin{aligned} p^0(i,t) > 0, \quad &\text{if} \quad h_i(t) \neq 0, \\ g(i,t) > 0, \quad &\text{if} \quad f_i(t) \neq 0, \\ p(i,t;j,\theta) > 0, \quad &\text{if} \quad a_{i,j}(\theta) \neq 0, \end{aligned} \tag{12}$$

and construct an analogue of estimate (7) of functional $\int_0^T H(\tau)Y(\tau)d\tau$ for a fixed vector $H(t) = (h_1(t), \ldots, h_n(t))$. Later we note that Eq. (10) is a Volterra-type equation, and the majorant process

$$\overline{Y}_{m+1}(t) = \int_0^T |A(\theta)| E(t-\theta)\overline{Y}_m(\theta)d\theta + |\tilde{F}(t)| \tag{13}$$

always converges. This actually follows from the Picard theorem but can be verified directly.

We define the scalar product of two n-dimensional vectors as $(A, B) = a_1b_1 + \ldots + a_nb_n$. When constructing an unbiased estimate (H, Y), the only requirement for the Markov chain remains meeting the compliance conditions (12). If these conditions are met, the algorithm for solving the problem is to model the trajectory of the chain

$$i_0, t_0 \to i_1, t_1 \to \ldots \to i_\tau, t_\tau$$

and to calculate the estimate

$$\hat{J}_\tau = \frac{h_{i_0}(t_0)a_{i_0,i_1}(t_1)\ldots a_{i_{\tau-1},i_\tau}(t_\tau)\tilde{f}_{i_\tau}(t_\tau)}{p^0_{i_0}(t_0)p(i_0,t_0;i_1,t_1)\ldots p(i_{\tau-1},t_{\tau-1};i_\tau,t_\tau)g(i_\tau,t_\tau)}, \tag{14}$$

$t_0 = T$, for which the following equality is true:

$$\mathbf{E}\hat{J}_\tau = (H, Y).$$

The estimate $\hat{J}_\tau$ is zero outside the simplex $t_0 \geq t_1 \geq \ldots \geq t_\tau$. The trajectory takes place inside this simplex.

These preliminary considerations lead us to the following results:

Theorem 1. *Let a random Markov process with initial distribution and transition probability densities be given*,

$$\begin{gathered}\mathbf{p}^0(t_0) = \overline{p}_i^0(t_0) : \{p_i^0(\theta_0) = p^0(i_0, \theta_0)\}; \quad i_0 = \overline{1,n}; \quad t_0 = T, \\ \mathbf{p}(t) = \left\|p_{i,j}(t)\right\|_{i,j=1}^n : \{p_{i,j}(\theta) = p(i_k, \theta_k; i_{k+1}, \theta_{k+1})\}; \quad k = \overline{0,\tau}; \quad t \in [0,T],\end{gathered} \tag{15}$$

respectively, *for which the compliance condition* (12) *is met and the equality*

$$\sum_{j=1}^n \int_0^t p(i,t;j,\theta)d\theta = 1 - g(i,t), \quad 0 \leq g(i,t) \leq 1 \tag{16}$$

holds.

Then for estimate (14), *the following equality holds*:

$$\mathbf{E}\hat{J}_\tau = (H, Y), \tag{17}$$

where $H(t)$ is the set vector and $Y(t)$ is the solution of Eq. (1).

Proof. In accordance with the Picard theorem ([7], p. 53), we have

$$\begin{aligned}(H,Y)(t) &= (H,\tilde{F})(t) + \sum_{\tau=1}^\infty \int_0^t d\theta_1 \ldots \int_0^{\theta_{\tau-1}} d\theta_\tau H(\theta_1)A(\theta_2)\ldots A(\theta_\tau)\tilde{F}(\theta_\tau) \\ &= \sum_{\tau=0}^\infty \sum_{i_0=0}^n \ldots \sum_{i_\tau=0}^n \int_0^t d\theta_0 \ldots \int_0^{\theta_{\tau-1}} d\theta_\tau h_{i_0}(\theta_0)a_{i_0,i_1}(\theta_1)\ldots a_{i_{\tau-1},i_\tau}(\theta_\tau)\tilde{f}_{i_\tau}(\theta_\tau).\end{aligned}$$

Having a given Markov chain, we obtain N trajectories i_0, $t_0 \to i_1$, $t_1 \to \ldots \to i_\tau$, t_τ with the density

$$p^0(i_0,t_0)p(i_0,t_0;i_1,t_1)\ldots p(i_{\tau-1},t_{\tau-1};i_\tau,t_\tau)g(i_\tau,t_\tau). \tag{18}$$

The set of these densities determines the probability measure $P(d\theta)$ on the set of trajectories. According to condition (16), almost all trajectories are finite, so the following equality holds:

$$\sum_{\tau=0}^\infty \sum_{i_0=0}^n \ldots \sum_{i_\tau=0}^n \int_0^t d\theta_0 \ldots \int_0^{\theta_{\tau-1}} d\theta_\tau p^0(i_0,\theta_0)p(i_0,t_0;i_1,\theta_1)\ldots p(i_{\tau-1},t_{\tau-1};i_\tau,\theta_\tau)g(i_\tau,t_\tau) = 1.$$

Using the definition of the mathematical expectation, under compliance conditions (12) for estimate (14), we obtain

$$\begin{aligned}&\sum_{\tau=0}^\infty \sum_{i_0=0}^n \ldots \sum_{i_\tau=0}^n \int_0^t d\theta_0 \ldots \int_0^{\theta_{\tau-1}} d\theta_\tau \hat{J}_\tau p^0(i_0,\theta_0)p(i_0,t_0;i_1,\theta_1)\ldots p(i_{\tau-1},t_{\tau-1};i_\tau,\theta_\tau)g(i_\tau,t_\tau) \\ &\quad = \sum_{\tau=0}^\infty \sum_{i_0=0}^n \ldots \sum_{i_\tau=0}^n \int_0^t d\theta_0 \ldots \int_0^{\theta_{\tau-1}} d\theta_\tau h_{i_0}(\theta_0)a_{i_0,i_1}(\theta_1)\ldots a_{i_{\tau-1},i_\tau}(\theta_\tau)\tilde{f}_{i_\tau}(\theta_\tau).\end{aligned}$$

These are the following obvious differences from the classic case discussed at the beginning of the paper:

(1) the fulfillment of condition (16);

(2) the automatic fulfillment of the majorant condition (13).

When the conditions of Theorem 1 are fulfilled, the following theorem is also true:

Theorem 2. *The variance of estimate* (14) *is finite and equal to*

$$\left(\frac{H_i^2}{\mathbf{p}_i^0}, \Xi\right) - (H,Y)^2, \tag{19}$$

where $\Xi(t)$ *is the solution of the equation* $\Xi(t) = \int_0^T \frac{A_{ij}^2(\theta)E^2(t-\theta)}{\mathbf{p}_{ij}(\theta)}\Xi(\theta)d\theta + \frac{\tilde{F}_i^2(t)}{\mathbf{g}_i(t)}$, *and notations* A_{ij}^2, $\tilde{F}_i$, $\mathbf{p}_i^0$, $\mathbf{p}_{ij}$, *and* $\mathbf{g}_i$ *imply component-by-component operations with matrices and vectors* (*Hadamard operations*).

Proof. We use the standard expression for the variance:

$$D\hat{J}_\tau = \mathbf{E}\hat{J}_\tau^2 - [\mathbf{E}\hat{J}_\tau]^2.$$

It follows from Theorem 1 that $[\mathbf{E}\hat{J}_\tau]^2 = (H, Y)^2$. Let us consider expression $\mathbf{E}\hat{J}_\tau^2$. By raising $\hat{J}_\tau$ to the square we obtain the following expression:

$$\mathbf{E}\hat{J}_\tau^2(t) = \sum_{\tau=0}^{\infty}\sum_{i_0=0}^{n}\dots\sum_{i_\tau=0}^{n}\int_0^t d\theta_0\dots\int_0^{\theta_{\tau-1}} d\theta_\tau \frac{h_{i_0}^2(\theta_0)a_{i_0,i_1}^2(\theta_1)\dots a_{i_{\tau-1},i_\tau}^2(\theta_\tau)\tilde{f}_{i_\tau}^2(\theta_\tau)}{p^0(i_0,\theta_0)p(i_0,t_0;i_1,\theta_1)\dots p(i_{\tau-1},t_{\tau-1};i_\tau,\theta_\tau)g(i_\tau,\theta_\tau)} \tag{20}$$

for the scalar product $\left(\frac{H_i^2}{\mathbf{p}_i^0},\Xi\right)$, where $\Xi(t) = \int_0^T \frac{A_{ij}^2(\theta)E^2(t-\theta)}{\mathbf{p}_{ij}(\theta)}\Xi(\theta)d\theta + \frac{\tilde{F}_i^2(t)}{\mathbf{g}_i(t)}$, or the equivalent system of differential equations

$$\frac{d}{dt}\Xi(t) = \frac{A_{ij}^2}{\mathbf{p}_{ij}}(t)\Xi(t) + \frac{F_i^2}{\mathbf{g}_i}(t), \quad \Xi(0) = \Xi_0.$$

According to the Picard theorem, the solution to this system of equations exists and is unique:

$$\left(\frac{H_i^2}{\mathbf{p}_i^0},\Xi\right)(t) = \sum_{\tau=1}^{\infty}\int_0^t d\theta_0\dots\int_0^{\theta_{\tau-1}} d\theta_\tau \frac{H_i^2(\theta_0)}{\mathbf{p}_i^0(\theta_0)}\frac{A_{ij}^2(\theta_1)}{\mathbf{p}_{ij}(\theta_1)}\dots\frac{A_{ij}^2(\theta_\tau)}{\mathbf{p}_{ij}(\theta_\tau)}\frac{\tilde{F}_i^2(\theta_\tau)}{\mathbf{g}_i(\theta_\tau)}.$$

Therefore, $\left(\frac{H_i^2}{\mathbf{p}_i^0},\Xi\right) = \mathbf{E}\hat{J}_\tau^2$.

Clearly, the variance of the estimate is always finite, and it is possible to specify such parameters of the Markov chain that ensure the minimum value of the variance of the estimate.

Theorem 3. *The variance of estimate* $\hat{J}_\tau$ *has the minimum value when the following parameters of the Markov chain are chosen*:

$$\begin{aligned} q_{\text{opt}}^0(i,t) &= \frac{|h_i(t)|\,\overline{y}_i(t)}{(|H|,\overline{Y})}; \\ q_{\text{opt}}(i,t;j,\theta) &= \frac{|a_{i,j}(\theta)|\,E(t-\theta)\overline{y}_j(\theta)}{\overline{y}_i(t)}, \end{aligned} \tag{21}$$

where $\overline{Y}$ *is the solution of equation* $\overline{Y}(t) = \int_0^T |A(\theta)|\,E(t-\theta)\overline{Y}_m(\theta)d\theta + |\tilde{F}(t)|$ *and* y_i *is the* i*th component of* $\overline{Y}$.

Proof. Let us consider estimate (14). As far as we have seen, when using this estimate for each Markov chain that meets the compliance conditions, we have

$$(H,Y)(t) = \mathbf{E}\hat{J}_\tau(t) = \int_0^t \hat{J}_\tau(\omega_{\tau,\theta})P(d\theta),$$

where P is the measure on the Markov chain's trajectory, such that the trajectory's distribution density with respect to the measure $\mu^{\tau+1}$ is expression (18). Suppose that we want to introduce another measure $Q(d\theta)$ such that the Radon–Nikodym derivative $\frac{dP}{dQ}$ exists. For this, measure P should be absolutely continuous with respect to measure Q. In other words, it must have a structure similar to P: trajectory $\omega_{\tau,\theta}$ has the density $q_\tau(i_o, t_0; \dots; i_\tau, t_\tau)$ with respect to $\mu^{\tau+1}$, and q_τ should be greater than zero for those trajectories $\omega_{\tau,\theta}$ for which density expression (18) takes strictly positive values.

According to the significant sampling theorem ([1], p. 106), we can write the expression for the optimal density as follows:

$$q_{\tau,\mathrm{opt}}(\omega_{\tau,\theta}) = C\left|\hat{J}_\tau\right| p^0(i,t)p(i_0,t_0;i_1,t_1)\dots p(i_{\tau-1}t_{\tau-1};i_\tau t_\tau),$$

which is similar to

$$q_{\tau,\mathrm{opt}}(\omega_{\tau,\theta}) = C\left|h_{i_0}(t_0)a_{i_0,i_1}(t_1)\dots a_{i_{\tau-1},i_\tau}(t_\tau)\tilde{f}_{i_\tau}(t_\tau)\right|,$$

where C is the normalization constant,

$$C^{-1} = \sum_{\tau=0}^{\infty}\sum_{i_0=0}^{n}\dots\sum_{i_\tau=0}^{n}\int_0^t d\theta_0\dots\int_0^{\theta_{\tau-1}} d\theta_k\left|h_{i_0}(\theta_0)a_{i_0,i_1}(\theta_1)\dots a_{i_{\tau-1},i_\tau}(\theta_\tau)\tilde{f}_{i_\tau}(\theta_\tau)\right|,$$

$$C^{-1} = (|H|,\bar{Y});$$

and $\bar{Y}$ satisfies Eq. (13).

Note that there does not necessarily have to be a Markov chain induced by Q. However, in the case under consideration, it seems possible to specify a Markov chain inducing this optimal measure. It follows from Eq. (13) that

$$\sum_{j=1}^{n}\int_0^T |a_{i,j}(\theta)|\,E(t-\theta)\bar{y}_j(\theta)d\theta = \bar{y}_i(t) - |\tilde{f}_i(t)|, \quad \frac{|\tilde{f}_i(t)|}{\bar{y}_i(t)} \le 1,$$

$$\sum_{j=1}^{n}\int_0^T \frac{|a_{i,j}(\theta)|\,E(t-\theta)\bar{y}_j(\theta)}{\bar{y}_i(t)}d\theta = 1 - \frac{|\tilde{f}_i(t)|}{\bar{y}_i(t)}.$$

Assuming $g_{\mathrm{opt}}(i,t) = \dfrac{|\tilde{f}_i(t)|}{\bar{y}_i(t)}$, we obtain the expression

$$\sum_{j=1}^{n}\int_0^t q_{\mathrm{opt}}(i,t;j,\theta)d\theta = 1 - g_{\mathrm{opt}}(i,t).$$

In other words, condition (16) is satisfied for the newly induced measure Q. The density expression takes the form

$$q_{\tau,\mathrm{opt}}(\theta_\tau) = \frac{|h_{i_0}(t_0)||a_{i_0,i_1}(t_1)|\dots|a_{i_{\tau-1},i_\tau}(t_\tau)||\tilde{f}_{i_\tau}(t_\tau)|}{(|H|,\bar{Y})}.$$

It is clear from Eq. (13) that $q^0_{\mathrm{opt}}(i,t)$ is a nonnegative function, and it is normalized (i.e., it is a density), and function $q_{\mathrm{opt}}(i,t;j,\theta)$ is nonnegative. Therefore, the compliance conditions (12) are satisfied for measure Q.

Using the Cauchy–Bunyakovsky inequality, we find the lower estimate for $\mathbf{E}\hat{J}^2_\tau$:

$$\int_0^t\left(\hat{J}_\tau\frac{dP}{dQ}(\theta)\right)^2 Q(d\theta) \ge \left(\int_0^t\left|\hat{J}_\tau\frac{dP}{dQ}(\theta)\right|Q(d\theta)\right)^2 = (H,Y)^2(t).$$

Substituting $q_{\tau,\,\mathrm{opt}}(\omega_{\tau,\theta})$ into the expression for $\mathbf{E}\hat{J}^2_\tau$, we obtain

$$\mathbf{E}\hat{J}^2_\tau(t) = \int_0^t \hat{J}^2_\tau\left(\frac{dP}{dQ}(\theta)\right)^2 Q(d\theta) = \int_0^t \hat{J}^2_\tau\cdot\frac{1}{C^2|\hat{J}^2_\tau|}\cdot C\left|\hat{J}_\tau\right| p_\tau(\omega_{\tau,\theta})$$

$$= \int_0^t \frac{|\hat{J}_\tau|}{C}P(d\theta) = (|H|,\bar{Y})(t)\cdot\int_0^t\left|\hat{J}_\tau\right|P(d\theta) = (|H|,\bar{Y})^2(t)$$

and ensure that Theorem 3 is true.

The results obtained above allow us to deduce the following corollary:

Corollary. *If $a_{i,j}(t) \ge 0$ and $f_i(t) \ge 0$ for all i, j, and t, then the optimal value of the variance of the estimate $\hat{J}_\tau$ is zero.*

Proof. In this case, it is obvious that if the specified conditions are met, then

$$(H,Y)^2 = (|H|,\overline{Y})^2,$$

and in expression (19), we obtain

$$D\hat{J}_\tau = (|H|,\overline{Y})^2 - (H,Y)^2 = 0.$$

Note that for the validity of the results similar to those from Theorem 3, the dependence of the absorption probability on t is important. We can explain this by the following simple example. Assume $\hat{F}(t) = 0$, and $A(t) = \|a_{i,j}\|_{i,j=1}^n$ is independent of t. Then the solution can be written in the closed form:

$$Y(t) = \exp(At) = \sum_{k=0}^{\infty} \frac{A^k t^k}{k!}. \tag{22}$$

The scalar product (Y, H) can be calculated using the Monte Carlo method by modeling a homogeneous Markov chain $\overline{p}^0$, P, for which compliance conditions (12) are met, and calculating the estimate

$$\hat{J}_\tau = \frac{h_{i_0} a_{i_0,i_1} \dots a_{i_{\tau-1},i_\tau} y_{i_\tau}^0}{p_{i_0}^0 p_{i_0,i_1} \dots p_{i_{\tau-1},i_\tau} g_{i_\tau} \cdot \tau!}. \tag{23}$$

It is easy to verify (see, for example, [6]), that at $a_{i,j} \ge 0$, $y_{i_\tau}^0 \ge 0$ and $h_{i_0} \ge 0$ we cannot choose a chain $\overline{p}^0; P$ that converts the variance to zero.

Below, some examples of using the proposed algorithm to solve the equations that arise in queueing theory are given.

4. CONSTANT MATRIX OF COEFFICIENTS

Here we present an illustrative example. Let us consider a queueing theory equation that describes a closed system of $M = 10$ conventional machines that break down with the intensity λ, and $S = 7$ conventional service persons, who repair the machines with intensity μ. Using ([8], p. 102), we obtain the following system of equations:

$$\begin{aligned}
\dot{y}_0(t) &= -10\lambda y_0(t) + \mu/7 y_1(t),\\
\dot{y}_1(t) &= -[9\lambda + \mu/7] y_1(t) + 10\lambda y_0(t) + 2\mu/7 y_2(t),\\
\dot{y}_2(t) &= -[8\lambda + 2\mu/7] y_2(t) + 9\lambda y_1(t) + 3\mu/7 y_3(t),\\
\dot{y}_3(t) &= -[7\lambda + 3\mu/7] y_3(t) + 8\lambda y_2(t) + 4\mu/7 y_4(t),\\
\dot{y}_4(t) &= -[6\lambda + 4\mu/7] y_4(t) + 7\lambda y_3(t) + 5\mu/7 y_5(t),\\
\dot{y}_5(t) &= -[5\lambda + 5\mu/7] y_5(t) + 6\lambda y_4(t) + 6\mu/7 y_6(t),\\
\dot{y}_6(t) &= -[4\lambda + 6\mu/7] y_6(t) + 5\lambda y_5(t) + \mu/7 y_7(t),\\
\dot{y}_7(t) &= -[3\lambda + \mu] y_7(t) + 4\lambda y_6(t) + \mu y_8(t),\\
\dot{y}_8(t) &= -[2\lambda + \mu] y_8(t) + 3\lambda y_7(t) + \mu y_9(t),\\
\dot{y}_9(t) &= -[\lambda + \mu] y_9(t) + 2\lambda y_8(t) + \mu y_{10}(t),\\
\dot{y}_{10}(t) &= -\mu y_{10}(t) + \lambda y_9(t).
\end{aligned} \tag{24}$$

Within this paper, we have considered a specific example with the following parameters: $\mu = 0.0202$ and $\lambda = 0.02$, and $\mu = 0.0202$ and $\lambda = 0.01$. This combination of parameters gives us the load factor values (see ([8], p. 104)) $\Psi = 0.99$ and $\Psi = 0.49$, respectively. When the load factor $\Psi = 0.99$, the simulation methods work poorly. Solving for second value of the load factor demonstrates the viability of the proposed method, while also comparing it with Runge–Kutta method's benchmark, under more conven-

Table 1. Comparison of solutions obtained by the Monte Carlo and Runge–Kutta methods at $T = 1.5$, $\Psi = 0.99$, and $N = 1800000$

	Monte Carlo	Runge–Kutta
y_0	6.584041e-06	1.425017e-05
y_1	6.602467e-03	6.672860e-03
y_2	7.843779e-01	7.819781e-01
y_3	1.900002e-01	1.898743e-01
y_4	2.019594e-02	2.018617e-02
y_5	1.276377e-03	1.226557e-03
y_6	4.864802e-05	4.658389e-05
y_7	3.336787e-07	1.132339e-06
y_8	1.867346e-08	1.721394e-08
y_9	9.384152e-11	1.495906e-10
y_{10}	$\leq \hat{s}_n$	5.691357e-13

Table 2. Comparison of the confidence interval of the Monte Carlo method and the absolute error between the solutions obtained by the Monte Carlo and Runge–Kutta methods at $T = 1.5$, $\Psi = 0.99$, and $N = 1800000$

	$\hat{s}_n$	χ
y_0	8.884365e-06	7.666127e-06
y_1	3.728903e-04	7.039296e-05
y_2	4.111465e-03	2.399780e-03
y_3	1.706013e-03	1.258374e-04
y_4	3.551892e-04	9.779131e-06
y_5	5.111231e-05	4.981978e-05
y_6	4.754939e-06	2.064131e-06
y_7	1.866752e-07	–
y_8	1.055185e-08	–
y_9	8.037395e-11	–
y_{10}	1.036435e-16	–

tional conditions. Let us choose the initial position $y_2^0 = 1$; $y_i^0 = 0$, $\forall i \neq 2$; and $T = 1.5$; and set the number of trajectories to $N = 1800000$.

In order to compare the results of the calculations, we use the SciPy library of the Python programming language [9]. Let us use the routine scipy.integrate.ode("dop853") that implements the explicit Dormand–Prince algorithm of order 8(5, 3) for the Runge-Kutta method. This is an adaptive algorithm that automatically changes the integration step depending on whether the local error falls within the required error value. The error is calculated based on the formula $AbsTol + RelTol * |\hat{y}_{\text{step}}|$, where *AbsTol* is the absolute error, *RelTol* is the relative error, and $\hat{y}_{\text{step}}$ is the calculated solution at each step (see ([10], p. 188–204)). We set *AbsTol* = 1e-06 and *RelTol* = 1e-04.

In order to verify the accuracy of the Monte Carlo method, we construct a 99.7% confidence interval $[-\hat{s}_n, \hat{s}_n]$ for the estimate of $\hat{J}_\tau$ and check whether the difference between the obtained solution values falls into this interval. For illustration, we present the values of $\hat{s}_n = \frac{3\sigma}{\sqrt{N}}$ and the values of absolute actual error $\chi = |\hat{y}(t) - y^{RK}(t)|$ in the tables.

Table 3. Comparison of the solutions obtained by the Monte Carlo and Runge–Kutta methods at $T = 1.5$, $\Psi = 0.49$, and $N = 1800000$

	Monte Carlo	Runge–Kutta
y_0	1.400258e-05	1.630445e-05
y_1	7.620840e-03	7.575809e-03
y_2	8.787064e-01	8.804753e-01
y_3	1.058588e-01	1.061585e-01
y_4	5.537256e-03	5.601951e-03
y_5	1.644043e-04	1.689383e-04
y_6	2.892959e-06	3.184300e-06
y_7	4.260589e-08	3.841370e-08
y_8	$\leq \hat{s}_n$	2.898100e-10
y_9	$\leq \hat{s}_n$	1.249898e-12
y_{10}	$\leq \hat{s}_n$	2.359078e-15

Table 4. Comparison of the confidence interval of the Monte Carlo method and the absolute error between the solutions obtained by the Monte Carlo and Runge–Kutta methods at $T = 1.5$, $\Psi = 0.49$, and $N = 1800000$

	$\hat{s}_n$	χ
y_0	4.204214e-06	2.301872e-06
y_1	2.519817e-04	4.503130e-05
y_2	3.694861e-03	1.768928e-03
y_3	8.335724e-04	2.996921e-04
y_4	9.236051e-05	6.469456e-05
y_5	6.750112e-06	4.533990e-06
y_6	3.615292e-07	–
y_7	1.237473e-08	–
y_8	2.560555e-11	–
y_9	5.164668e-14	–
y_{10}	4.225728e-24	–

The calculation results are shown in Tables 1–4. When constructing the tables, we decided not to round up the results of the calculations obtained by the Runge–Kutta method but to exclude from the calculation of χ those $y^{RK}(t)$ whose significands were less than or equal in order to the *AbsTol* value. In these cases, there are dashes in the χ columns of the tables. Apparently, the values obtained by the Monte Carlo method are less than or equal to $\hat{s}_n$ in some states, which may be attributed to their closeness to zero.

5. VARIABLE MATRIX OF COEFFICIENTS

Here we demonstrate the applicability of the proposed method in solving the system of equations with a time-dependent matrix of coefficients. We use system (24) and make the parameter λ dependent on t: $\lambda(t) = \lambda_0/(1 + t^2)$. We retain the initial parameter $\lambda_0 = 0.02$. The algorithm for calculating the general form functional (H, Y) repeats the algorithm described above for the case of a constant coefficient matrix, but with the need to take into consideration the changes that occur with time in the matrix of coefficients $A(t)$.

Table 5. Comparison of the solutions obtained by the Monte Carlo and Runge–Kutta methods at $T = 1.5$, $\lambda = \lambda_0/(1 + t^2)$, $\lambda_0 = 0.02$, and $N = 2200000$

	Monte Carlo	Runge–Kutta
y_0	1.571291e-05	1.569973e-05
y_1	7.285691e-03	7.297974e-03
y_2	8.508264e-01	8.486886e-01
y_3	1.342014e-01	1.343154e-01
y_4	9.442907e-03	9.304705e-03
y_5	3.745529e-04	3.683801e-04
y_6	9.152182e-06	9.115712e-06
y_7	1.262589e-07	1.443672e-07
y_8	1.206253e-09	1.430467e-09
y_9	1.160412e-11	8.104518e-12
y_{10}	1.902090e-16	2.009527e-14

Table 6. Comparison of the confidence interval of the Monte Carlo method and the absolute error between the solutions obtained by the Monte Carlo and Runge–Kutta methods at $T = 1.5$, $\lambda = \lambda_0/(1 + t^2)$, $\lambda_0 = 0.02$, and $N = 2200000$

	$\hat{s}_n$	χ
y_0	1.327831e-07	1.318437e-08
y_1	3.981266e-05	1.228340e-05
y_2	3.025420e-03	2.137891e-03
y_3	1.408177e-03	1.139761e-04
y_4	2.221353e-04	1.382022e-04
y_5	1.686217e-05	6.172786e-06
y_6	8.709463e-07	–
y_7	2.493599e-08	–
y_8	5.125696e-10	–
y_9	7.929071e-12	–
y_{10}	8.923702e-17	–

The estimate $\hat{J}_\tau$ takes form (14) if we keep the same parameters of the transition probability and initial distribution densities that induce the Markov chain: $\mathbf{p}^0 = \overline{p}_i^0$ for the initial distribution and $\mathbf{p}(\theta)$: $\left\{ p(i, \theta_k; j, \theta_{k+1}) = \dfrac{p_{i,j}}{t_i} \right\}$ for the transition probability density. The calculation is performed with parameters $\mu = 0.0202$, $\lambda_0 = 0.02$, $T = 1.5$, $y_2^0 = 1$, $y_i^0 = 0$, $\forall i \neq 2$, and $N = 2200000$. The results of the calculations are shown in Tables 5 and 6.

Note that the well-known methods for reducing the variance of the estimates, such as the significant sampling method, trajectory branching method, and quadrature formulas with random nodes, were not used by the authors in their calculations. These methods would obviously reduce the required computation time. However, the results presented here reasonably testify to the workability of the theory described in the paper.

REFERENCES

1. C. M. Ermakov, *The Monte-Carlo Method in Computational Mathematics* (Nevskii Dialekt, St. Petersburg, 2009) [in Russian].

2. C. M. Ermakov, *Monte-Carlo Method and Related Topics* (Nauka, Moscow, 1975) [in Russian].
3. C. M. Ermakov and A. S. Sipin, *Monte-Carlo Method and Parametrical Divisibility of Algorithms* (S.-Peterb. Gos. Univ., St. Petersburg, 2014) [in Russian].
4. C. M. Ermakov and G. A. Mikhailov, *Statistical Modeling* (Nauka, Moscow, 1982) [in Russian].
5. S. Ermakov and A. Pogosian, "On solving stochastic differential equations," Monte Carlo Methods Appl. **25**, 155–161 (2019). https://doi.org/10.1515/mcma-2019-2038
6. S. M. Ermakov and T. M. Tovstik, "Monte Carlo method for solving ODE systems," Vestn. St. Petersburg Univ.: Math. **52**, 272–280 (2019). https://doi.org/10.1134/S1063454119030087
7. Yu. N. Bibikov, *A Course in Ordinary Differential Equations* (Lan', St. Petersburg, 2011) [in Russian].
8. A. Kaufmann and R. Cruon, *Los Fenómenos de Espera: Teoría y Aplicaciones* (Continental, México, 1964; Mir, Moscow, 1965).
9. P. Virtanen, R. Gommers, T. E. Oliphant, et al., "SciPy 1.0: Fundamental algorithms for scientific computing in Python," Nat. Methods **17**, 261–272 (2020). https://doi.org/10.1038/s41592-019-0686-2
10. E. Hairer, *Solving Ordinary Differential Equations I: Nonstiff Problems* (Springer-Verlag, Berlin, 1993).

Translated by N. Semenova