

УДК 519.24

DOI: 10.17223/19988605/47/5

B.S. Dobronets, O.A. Popova

COMPUTATIONAL ASPECTS OF PROBABILISTIC EXTENSIONS

In this article we propose a new approach to computing of functions with random arguments. Approach based on the idea of dimension reduction by to calculating some integrals and the application of numerical probability analysis. We apply one of the basic concepts of numerical probabilistic analysis as the probabilistic extension to computing a function with random arguments. To implement this technique, a new method based on parallel recursive calculations is proposed. Numerical examples are presented demonstrating the effectiveness of the proposed approach.

Keywords: computational probabilistic analysis; probabilistic extensions; non-Monte Carlo methods; random boundary value problem.

Problems of modelling can be reduced to the numerical analysis of functions [1].

$$z = f(x_1, x_2, \dots, x_n).$$

Let (x_1, \dots, x_n) be a system of continuous random variables with joint probability density function $p(x_1, \dots, x_n)$ and random variable z is a function $f(x_1, \dots, x_n)$. Consider main methods to find the distribution of random variable z .

These are to use the cumulative distribution function

$$F(y) = P(f(x_1, \dots, x_n) \leq y)$$

and the probability $z \leq y$ is determined by the formula

$$P(z \leq y) = \int_{\Omega_y} p(x_1, \dots, x_n) dx_1 \dots dx_n, \tag{1}$$

where $\Omega_y = \{(x_1, \dots, x_n) | f(x_1, \dots, x_n) \leq y\}$. Further, the probability density function f of random variable z is defined as the derivative [2]

$$f(y) = \frac{dF(y)}{dy}.$$

In the case of the monotone function $f(x)$, the distribution density f of the random variable z is defined as [2] as follows:

$$f(y) = f(f^{-1}(y)) |(f^{-1})'(y)|, \tag{2}$$

where f^{-1} is the inverse function of f .

Note that the calculation of the integral (1) and the use of the formula (2) in some cases is rather difficult. Moreover, one of the reasons for the emergence of the Monte Carlo method was the need to calculate integrals of the form (1) in the case of large values of n . Thus, the Monte Carlo method is often the only way to calculate the probability density function f of a random variable z . Monte Carlo method is a powerful approach, but it has some serious shortcomings, first of all, this is an extremely low rate of convergence.

Non-Monte Carlo methods have been developed since the 1960s. A major non-Monte Carlo approach is interval analysis. However, interval analysis computing only to the boundaries of random processes, without examining their internal distributions.

An important special case is operations on independent random variables. In the papers [3–6], various approaches for numerical operations on densities of random variables are considered.

In our work, we develop a technique that uses Computational Probabilistic Analysis (CPA) to solve various problems with stochastic data uncertainty [7–9].

The basis of computational probabilistic analysis is numerical operations on probability density functions of the random values. These are operations “+”, “−”, “.”, “/”, “↑”, “max”, “min”, as well as binary relations “≤”, “≥” and some others. The numerical operations of the piecewise polynomial function arithmetic constitute the major component of CPA. The use of CPA for these problems is more effective than the Monte Carlo method in a thousand times.

Using the arithmetic of probability density functions and probabilistic extensions, we can construct numerical methods that enable us solving systems of linear and nonlinear algebraic equations with random parameter [7, 8].

We will use piecewise polynomial models to represent probability density functions [10–12]:

- piecewise constant functions (histograms);
- piecewise linear functions (frequency polygons);
- piecewise polynomial functions (splines).

1. Probabilistic extensions

One of the most important problems that NPA deals with is to construct probability density functions of random variables.

Let z be a function $f(x_1, \dots, x_n)$

$$z = f(x_1, \dots, x_n).$$

Definition. By probabilistic extension $f(\xi, \mathbf{x}_1, \dots, \mathbf{x}_n)$ of the function f , we mean a probability density function z of the random variable z

$$z(\xi) = f(\xi, \mathbf{x}_1, \dots, \mathbf{x}_n).$$

Definition. Support of the probability density functions f will be called the set

$$\text{supp}(f) = \{x \mid f(x) > 0\}.$$

One possible way to estimate the probability density z of a random variable z

$$z = f(x_1, \dots, x_n). \quad (3)$$

is the Monte Carlo method [13]. For these purposes a random vector (x_1^i, \dots, x_n^i) with joint probability density function $p(x_1, \dots, x_n)$ is generated. Further calculated $z^i = f(x_1^i, \dots, x_n^i), i = 1, \dots, N$. Using the histogram method for z^i , we can construct an estimate of the probability density function z .

We study the properties of the probability density z . Build the grid

$$\{z_0, z_1, \dots, z_k\} \in \text{supp}(z)$$

and calculate the number n_l of sample points falling in each segment z_{l-1}, z_l . It is known that histograms converge to the probability density function of the mean integral value

$$p_l = \frac{n_l}{N} \approx \int_{z_{l-1}}^{z_l} z(\xi) d\xi.$$

Build the grid $\{t_0, t_1, \dots, t_m\} \subset \text{supp}(x_1)$ and calculate the number m_j of sample points falling in each segment t_{j-1}, t_j .

$$p_j^1 = \frac{m_j}{N}.$$

Consider only those random vectors (x_1^i, \dots, x_n^i) for such $(x_1^i \in (t_{j-1}, t_j))$. Number such vectors is equal exactly m_j . If v_{jl} is the number of sample points falling in each segment z_{l-1}, z_l , then

$$n_l = \sum_j^m v_{jl}$$

and

$$p_l = \frac{n_l}{N} = \sum_j^m \frac{m_j}{N} \frac{v_{jl}}{m_j}.$$

Going to the limit at $N \rightarrow \infty$

$$x_1(t) \approx \frac{m_j}{N}.$$

Further, let for any $t \in \text{supp } x_1$ we can construct a probabilistic extension $f(\cdot, t, \mathbf{x}_2, \dots, \mathbf{x}_n)$

$$\frac{v_{jl}}{m_j} \approx f(\xi, t, \mathbf{x}_2, \dots, \mathbf{x}_n).$$

Increasing the dimension of the grids and going to the limit, we obtain

$$f(\xi) = \int_{x_1}^{\bar{x}_1} \mathbf{x}_1(t) f(\xi, t, \mathbf{x}_2, \dots, \mathbf{x}_n) dt. \quad (4)$$

Thus proven Theorem.

Theorem 1. Let $f(\xi, \mathbf{x}_1, \dots, \mathbf{x}_n)$ be probabilistic extensions of function $f(x_1, x_2, \dots, x_n)$ and for each real t function $f(\xi, t, \mathbf{x}_2, \dots, \mathbf{x}_n)$ be probabilistic extensions of the function $f(t, x_2, \dots, x_n)$. Then

$$f(\xi, \mathbf{x}_1, \dots, \mathbf{x}_n) = \int_{x_1}^{\bar{x}_1} \mathbf{x}_1(t) f(\xi, t, \mathbf{x}_2, \dots, \mathbf{x}_n) dt. \quad (5)$$

Corollary. Theorem 1 infers the possibility of recursive computations for the general form of probability extensions and reduction to the calculation of the one-dimensional case.

Let us consider the computing of the integral (5). For simplicity, we represent (5) as a numerical quadrature

$$\int_{x_1}^{\bar{x}_1} \mathbf{x}_1(t) f(\xi, t, \mathbf{x}_2, \dots, \mathbf{x}_n) dt \approx \sum_{l=1}^m \gamma_l \mathbf{x}_1(t_l) f(\xi, t_l, \mathbf{x}_2, \dots, \mathbf{x}_n).$$

Further, for the computing $f(\xi, t_l, \mathbf{x}_2, \dots, \mathbf{x}_n)$ we can also use numerical quadratures and so on. In general, it is NP-hard problem with actual parallelization.

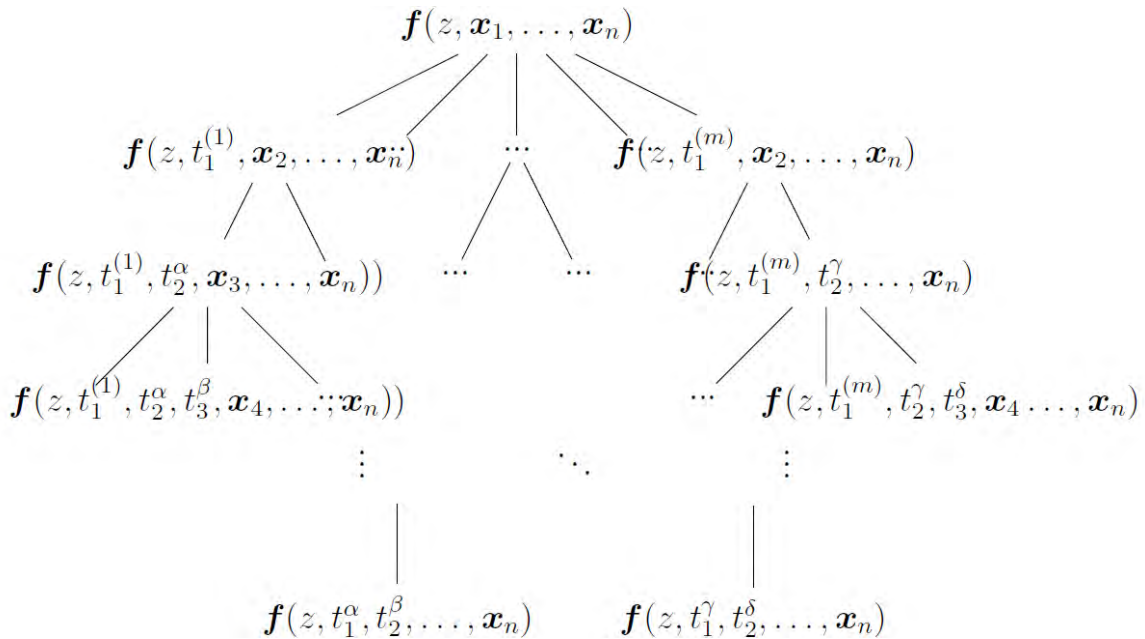


Fig. 1. The tree of the parallel recursive programming

In Fig. 1 we are shown the tree of the parallel recursive organization of the computational process. Thus, on the lower layer, it is necessary to computing the probabilistic extensions only for one variable. Note that all computations on each layer are independent and can be computed simultaneously.

2. One-dimensional case

Consider the procedure for computing the probabilistic extensions for the one-dimensional case. Let there be given a functional dependence

$$z = f(x),$$

where x is a random variable. Let \mathbf{x} be the probability density function of a random variable x with support $[\underline{x}, \bar{x}]$. Further $\{r_i(z) \in [\underline{x}, \bar{x}] | i = 1, \dots, n\}$ are the roots of the equation $z = f(x)$.

Following the main method (1) it is easy to construct a generalization of the expression (2). We can represent probabilistic extensions $f(\cdot, \mathbf{x})$ of function $f(x)$ in the form

$$f(\xi, \mathbf{x}) = \sum_{i=1}^n \frac{\mathbf{x}(r_i(\xi))}{|f'(r_i(\xi))|}.$$

Example 1. As an example, consider the construction of a probabilistic extensions function

$$f = ax^2 + bx, a \geq 0, b \geq 0,$$

x is random variable distributed on $[0, 2]$ by a triangular law

$$\mathbf{x}(\xi) = \begin{cases} \xi & \text{if } \xi \in [0, 1], \\ 2 - \xi & \text{if } \xi \in [1, 2]. \end{cases}$$

Further

$$r_1(z) = -\frac{\sqrt{4az + b^2} + b}{2a}, r_2(z) = \frac{\sqrt{4az + b^2} - b}{2a}.$$

are roots $z = f(r_i)$. Choose a positive root $r(\xi) = r_2(\xi) = \frac{\sqrt{4a\xi + b^2} - b}{2a}$, and

$$[ax^2 + bx]' = 2ax + b = \sqrt{4az + b^2}.$$

Finally

$$f(\xi, \mathbf{x}) = \mathbf{x}(r(\xi)) / \sqrt{4a\xi + b^2}.$$

Put $a = 1$ and $b = 0$

$$f(\xi, \mathbf{x}) = \begin{cases} 1/2 & \text{if } \xi \in [0, 1], \\ 1/\sqrt{\xi} - 1/2 & \text{if } \xi \in [1, 4]. \end{cases}$$

2.1. Numerical approach

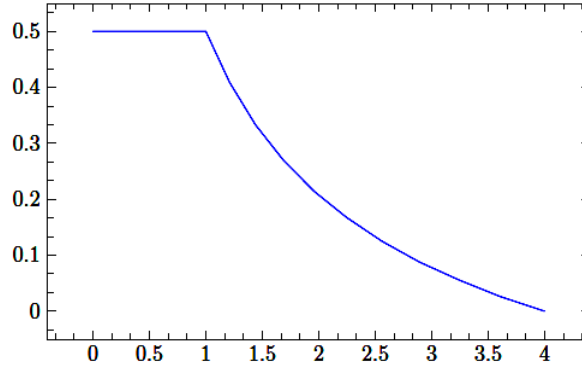
Consider a numerical approach to construct a probabilistic extensions f of function $f(x)$. For these purposes, we construct in the support $[\underline{x}, \bar{x}]$ of the probability density function \mathbf{x} a grid $\{\xi_i | \xi_i \in [\underline{x}, \bar{x}], i = 0, 1, 2, \dots, n\}$ and compute $\{z_i = f(\xi_i), i = 0, 1, 2, \dots, n\}$. Next we set

$$f_z(z_i) = \frac{\mathbf{x}(\xi_i)}{|f'(\xi_i)|}$$

and using $(z_i, f_z(z_i))$ we construct a piecewise polynomial interpolation.

So, for Example 1. the support is $[0, 2]$, $\xi_i = i/10, i = 0, 1, \dots, 20$. With $a = 1, b = 0$ we get

$$(z_i, f_z(z_i)) = \left((i/10)^2, \frac{\mathbf{x}(i/10)}{i/5} \right), \quad i = 0, 1, \dots, 20.$$


 Fig. 2. Probabilistic extensions $f(\cdot, \mathbf{x})$

The Fig. 2 shows the probability density function $f(\cdot, \mathbf{x})$ from Example 1. Unlike the general approach, there is no need to find the roots of the equation.

3. Two dimensional case

Let (x, y) be a system of continuous random variables with joint probability density function $p(x, y)$ and the random variable z is a function $f(x, y)$

$$z = f(x, y).$$

Need to find probability density function the random variable z . Define

$$\Omega_z = \{(x, y) \mid z > f(x, y)\}$$

and cdf F_z the random variable z

$$F_z = \int_{\Omega_z} p(x, y) dx dy$$

probability density function of z

$$f_z = \frac{d}{dz} F_z.$$

Consequently

$$\begin{aligned} \frac{d}{dz} F_z &= \frac{d}{dz} \int_{\Omega_z} p(x, y) dx dy = \\ &= \lim_{dz \rightarrow 0} \frac{\int_{\Omega_{z+dz}} p(x, y) dx dy - \int_{\Omega_z} p(x, y) dx dy}{dz} = \\ &= \lim_{dz \rightarrow 0} \frac{\int_{\Omega_{z+dz} - \Omega_z} p(x, y) dx dy}{dz}. \end{aligned}$$

Assuming that $dx = 0$ then

$$dz = f'_y dy$$

and

$$dS = dx dy = dx dz / |f'_y|.$$

Finally get

$$f_z(z) = \int_{\Gamma_z} \frac{p(x, y)}{|f'_y(x, y)|} dx,$$

where

$$\Gamma_z = \{(x, y) \mid z = f(x, y)\} = \{(x, y(x))\}.$$

Let

$$f_y(\xi, t) = \frac{p(t, y(\xi))}{|f'_y(\xi, y(t))|}$$

be probabilistic extensions of $f(t, y)$. Thus, the calculation of the probabilistic extension f_z is reduced to the calculation of the integral of probabilistic extensions $f(t, y)$.

Example 2. Consider the construction of a probabilistic extension for the function

$$z = x^2 y + xy^2.$$

Let t be real, put $y = t$ and a probabilistic extension for the function

$$z_x = x^2 t + xt^2.$$

is

$$z_x(\xi, t) = \mathbf{x}(\eta(\xi)) / \sqrt{4t\xi + t^4},$$

where $\eta(\xi) = (\sqrt{4t\xi + t^4} - b) / (2t)$.

Thus, the probabilistic extension for z can be represented as

$$z(\xi) = \int \mathbf{y}(t)_{z_x(\xi, t)} dt = \int \mathbf{y}(t) \mathbf{x}(\eta(\xi)) / \sqrt{4t\xi + t^4} dt. \quad (6)$$

4. Random Boundary value Problem

Consider using probabilistic extensions to calculate solutions Random Boundary value Problem [14]

$$Lu \equiv -pu'' + qu = f(x), \quad x \in (0, 1), \quad (7)$$

with boundary conditions

$$u(0) = 0, \quad u(1) = 0.$$

where $p > 0$, $q \geq 0$, p , q are independent random variables.

Let $\omega_h = \{x_i = ih, i = 1, 2, \dots, N-1, h = 1/N\}$ be grid and

$$L^h u_i^h = -p \frac{u_{i-1} - 2u_i + u_{i+1}}{h^2} + qu_i = f(x_i), \quad i = 1, 2, \dots, N-1.$$

is difference scheme. $[\underline{p}, \bar{p}]$, $[\underline{q}, \bar{q}]$ are supports of p , q . Generate grids

$$\omega_p = \{p_0 = \underline{p} < p_1 < \dots < p_K = \bar{p}\} \text{ and } \omega_q = \{q_0 = \underline{q} < q_1 < \dots < q_L = \bar{q}\}.$$

Solve numerically *KL* problems

$$-p_k \frac{u_{i-1} - 2u_i + u_{i+1}}{h^2} + q_l u_i = f(x_i), \quad i = 1, 2, \dots, N-1.$$

Thus we get an array of solutions $u_{kl} = u_i(p_k, q_l)$. Consider building probability density functions for u_i . For these purposes, we construct Hermitian cubic splines $s_l(p)$, $l = 0, 1, \dots, 10$ using the values $u_i(p_k, q_l)$. In fig. 3 they are shown in solid line. Further for some ξ we find the roots p_l

$$s_l(p_l) = \xi, \quad l = 0, 1, \dots, 10.$$

The values of $u_i(\xi)$ are computed using numerical quadratures, for example, Simpson quadratures

$$u_i(\xi) = h \sum_{k=0}^{10} \gamma_k q(q_k) p(p_k) / s'(p_k).$$

In fig. 3 line of integration mark = 'o'. Note, Simpson quadratures and cubic splines have $O(h^4)$ accuracy.

Numerical experiments with $K, L = 10$ showed good agreement with the Monte Carlo method with the number of samples $\sim 10^6$. Thus, in this example, the proposed method turned out to be more efficient than Monte Carlo $\sim 10^4$ times.

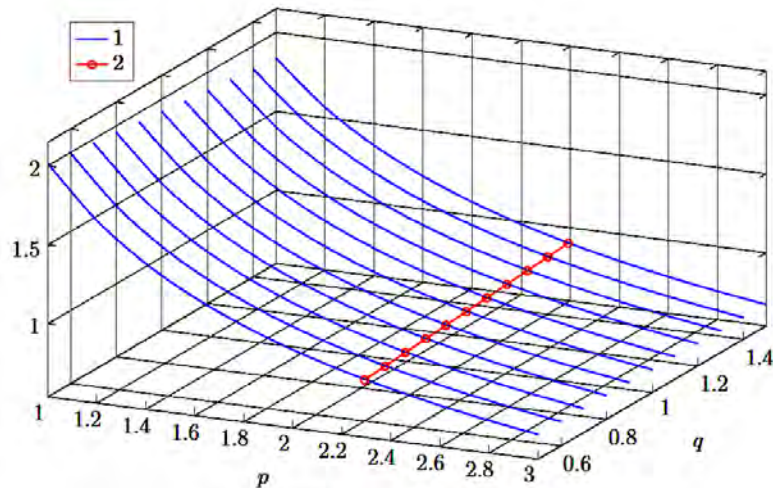


Fig. 3. Construction of probabilistic extensions for u_i , 1 are splines, 2 is a line of integration

Remark. The main computational costs are spent on building the set $u_{ikl} \sim O(KLN)$. Computational costs building of the probability extension of u_i is $O(m)$. Therefore, once you have u_{ikl} , you can compute relatively quickly u_i for different p, q .

Conclusion

The proposed approach makes it possible to solve the problem of computing the probability density function in the modeling processes with random input data. For these purposes we propose using a parallel-recursive organization of the computational process. Thus, the important problem of computing probability extensions can be solved within parallel recursive programming. This opens multifold possibilities for studying various models with random input data. Fast and accurate calculations are based on the properties of numerical arithmetic procedures over piecewise polynomial models developed within the framework of computational probabilistic analysis.

REFERENCES

1. Rocquigny, E. (2012) *Modelling Under Risk and Uncertainty: An Introduction to Statistical, Phenomenological and Computational Methods*. Wiley Series in Probability and Statistics. John Wiley & Sons.
2. Springer, M.D. (1979) *The Algebra of Random Variables*. New York; Chichester; Brisbane: John Wiley & Sons.
3. Gerasimov, V.A., Dobronets, B.S. & Shustrov M.Yu. (1991) Numerical operations of histogram arithmetic and their applications. *Automation and Remote Control*. 52(2). pp. 208–212.
4. Williamson, R. & Downs, T. (1990) Probabilistic arithmetic i: numerical methods for calculating convolutions and dependency bounds. *International Journal of Approximate Reasoning*. 4(2). pp. 89–158. DOI: 10.1016/0888-613X(90)90022-T
5. Li, W. & Hym, J. (2004) Computer arithmetic for probability distribution variables. *Reliability Engineering and System Safety*. 85(1–3). pp. 191–209. DOI: 10.1016/j.res.2004.03.012
6. Jaroszewicz, S. & Korzeń, M. (2012) Arithmetic operations on independent random variables: a numerical approach. *SIAM J. Sci. Comput.* 34(3). pp. A1241–A1265. DOI: 10.1137/110839680
7. Dobronets, B.S. & Popova, O.A. (2014) Numerical probabilistic analysis under aleatory and epistemic uncertainty. *Reliable Computing*. 19. pp. 274–289.
8. Dobronets, B. & Popova, O. (2016) Numerical Probabilistic Approach for Optimization Problems. *Scientific Computing, Computer Arithmetic, and Validated Numerics. Lecture Notes in Computer Science*. Vol. 9553. Springer International Publishing, Cham. pp. 43–53.
9. Dobronets, B.S. & Popova, O.A. (2017) Improving the accuracy of the probability density function estimation. *Journal of Siberian Federal University, Mathematics and Physics*. 10(1). pp. 16–21. DOI: 10.17516/1997-1397-2017-10-1-16-21
10. Dobronets, B.S. & Popova, O.A. (2018) Piecewise Polynomial Aggregation as Preprocessing for Data Numerical Modeling. *IOP Conf. Series: Journal of Physics: Conf. Series*. 1015. DOI:10.1088/1742-6596/1015/3/032028

11. Dobronets, B.S. & Popova, O.A. (2018) Improving reliability of aggregation, numerical simulation and analysis of complex systems by empirical data. *IOP Conf. Series: Materials Science and Engineering*. 354. DOI:10.1088/1757-899X/354/1/012006
12. Popova, O.A. (2019) Using Richardson extrapolation to improve the accuracy of processing and analyzing empirical data. *Measurement Techniques*. 2. pp. 18–22.
13. Mikhailov, G.A. & Voitishchik, A.V. (2018) *Statisticheskoe modelirovanie: metod Monte-Karlo* [Statistical modeling. Monte Carlo methods]. Moscow: Yurait.
14. Soong, T.T. (1973) *Random Differential Equations in Science and Engineering*. New York and London: Academic Press.

Received: December 14, 2018

Dobronets B.S., Popova O.A. (2019) COMPUTATIONAL ASPECTS OF PROBABILISTIC EXTENSIONS. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie vychislitel'naya tekhnika i informatika* [Tomsk State University Journal of Control and Computer Science]. 47. pp. 41–48

DOI: 10.17223/19988605/47/5

Добронетц Б.С., Попова О.А. ВЫЧИСЛИТЕЛЬНЫЕ АСПЕКТЫ ВЕРОЯТНОСТНЫХ РАСШИРЕНИЙ. *Вестник Томского государственного университета. Управление, вычислительная техника и информатика*. 2019. № 47. С. 41–48

В статье предлагается новый подход к вычислению функций со случайными аргументами. Подход основан на идее уменьшения размерности функции путем вычисления определенных интегралов и использования вычислительного вероятностного анализа. Применяется одно из основных понятий вычислительного вероятностного анализа – вероятностное расширение для вычисления функции со случайными аргументами. Для реализации этого метода предлагается способ, основанный на параллельных рекурсивных вычислениях. Приведены численные примеры, демонстрирующие эффективность предложенного подхода.

Ключевые слова: вычислительный вероятностный анализ; вероятностные расширения; не Монте-Карло методы; случайные краевые задачи.

DOBRONETS Boris Stanislavovich (Doktor of Physics and Mathematics, Professor of the Department of Systems of artificial intelligence Institute of Space and Information Technologies, Siberian Federal University, Krasnoyarsk, Russian Federation).
E-mail: BDobronets@yandex.ru

POPOVA Olga Arcadevna (Candidate of Technical Sciences, Associate Professor of the Department of Systems of artificial intelligence Institute of Space and Information Technologies, Siberian Federal University, Krasnoyarsk, Russian Federation).
E-mail: OlgaArc@yandex.ru