Smooth approximation of probability and quantile functions: vector generalization and its applications

R Torishnyi* and V Sobol

Department of Probability Theory and Computer Modeling, Moscow Aviation Institute (National Research University), 4 Volokolamskoe highway, 125993, Moscow, Russian Federation

*E-mail: rotorishnyj@mai.education

Abstract. In this paper, we provide an approximation method for probability function and its derivatives, which allows using the first order numerical algorithms in stochastic optimization problems with objectives of that type. The approximation is based on the replacement of the indicator function with a smooth differentiable approximation – the sigmoid function. We prove the convergence of the approximation to the original function and the convergence of their derivatives to the derivatives of the original ones. This approximation method is highly universal and can be applied in other problems besides stochastic optimization – the approximation of the kernel of the probability measure, considered in the present article as an example, and the confidence absorbing set approximations.

1. Introduction

Many theoretical and practical problems involve some kind of uncertainty, which must be considered to find a proper solution. The optimal method to handle these undetermined factors depends on the application domain and required properties of the system. One of the common ways to describe the model with such uncertainties is to involve them in the mathematical model as stochastic processes or random variables. This approach is applicable in various areas, not only those that are close to math, such as economics and management but even in chemistry [1] and biology [2]. If the problem can be represented as the mathematical programming problem, then the addition of random variables turns the objective and constraints to be random functions. Since that we need a special way to compare the solutions to choose the optimal one, as we cannot compare objectives directly. One of the advanced ways is to consider the optimization problems in which the objective or constraints are in the form of a probability function or a quantile function. This leads to stochastic optimization problems with probability and quantile functions. For instance, ensuring an acceptable level of reliability of some systems can be described as optimization with probability function as an objective, and the problem of minimizing loss level at a given level of reliability can be described as optimization with quantile function as an objective. Stochastic optimization problems are generally more complex than usual optimization problems, and the level of additional complexity depends on the type of the problem. Nevertheless, the wide application area and possible benefits from such sophisticated models could lead to more effective models and strategies.

We should note that researches on optimization problems with more complex criteria have been developed recently, and in some cases, the solution to these problems can be found. For example, problems, where the optimization criterion is complex nonlinear risk functional, can be solved using
statistical estimates [3]. Also, if the risk function in a two-stage optimization problem can be represented as a combination of its marginal and conditional forms, then this problem can be reduced to a simpler one [4]. In the present paper, however, we consider only the probability function.

Solution methods for certain types of stochastic optimization problems have been already developed. A quantile optimization problem with discrete distribution can be in some cases reduced to a mixed integer programming problem [5]. Under certain conditions, some of the two-step optimization problems can be reduced to a mixed integer linear programming problem [6]. Another algorithm, based on sample average estimates, was used to solve the two-stage optimization problem with the quantile criterion in [7]. These methods are proven to work for certain types of problems, but unfortunately can not be simply generalized.

A different approach to the solution of stochastic optimization problems is to calculate the gradients of the objectives. Theoretically, it leads to the solution of considered problems using gradient optimization methods. The first time an expression for probability function gradient was obtained by Raik [8] in form of a surface Riemann integral. Then, an expression for probability function gradient in form of a Lebesgue integral over a surface was obtained and used by Kibzun and Tretyakov [9]. Different expressions for probability function derivatives using the Lebesgue integral transformation were obtained by Marti [10,11]. Then it was shown by Uryasev [12] that under some requirements the surface integral presented in the expression of probability function gradient can be replaced with the volume integral, or at least with the sum of volume and surface integrals [13]. In general, direct calculation of probability and quantile function gradient is related to the calculation of the complex integrals over the surface which in many cases can not be easily and explicitly determined.

Many later works related to this issue are dedicated to the boundaries of the probability function derivatives. An expression for the lower boundary of probability function gradient norm in a linear optimization problem with probability constraints was obtained by Henrion in [14]. Also, an expression for the upper boundary of probability function subdifferentials was obtained by Henrion et al in [15]. Both of these works are focused on the Gaussian distribution of stochastic variables. Using weak derivatives and sampling mechanisms, Pflug et al [16] obtained an estimation for the probability function derivative. The approximation of quantile value, as well as an estimate for probability function derivatives, were obtained by Yu et al [17] for Gaussian, exponential, and chi-square distributions using extensions of the Taylor series. Method of gradient value estimation for Weibull distribution and method of approximate solutions of differential equations were developed by Okagbue et al in [18]. Also, asymptotic formulas for first and second order derivatives of probability function were proposed by Garniera et al [19] for Gaussian distribution, where the value of the derivative is estimated using the Monte-Carlo method. Methods of approximation proposed in these papers are restrained by the type of the distribution and specific conditions, or rely on other stochastic mechanisms.

It should be mentioned that recently a certain generalization of the probability function was developed. A buffered probability of exceedance was described in [20] and formulas for its differentials and subdifferentials were found in [21], but these formulas and methods do not apply to the standard probability function.

In this paper, we propose another method of approximation of the probability functions and their derivatives for a random vector with absolutely continuous distribution and continuous loss function. The main idea of the approximation is to replace the indicator function with its continuous differentiable approximation. This method is not computationally expensive due to the integration over volume in formulas for the probability function value and its derivatives. This allows us to unify the solution process using well-known optimization methods, such as the gradient descent or the gradient projection method. The second theoretical contribution of the present paper is the new algorithm of approximation of the probability measure’s kernel. The boundary of the kernel is determined as the isocuant of the special probability function. This problem is presented in section 2.3. Given the formulas of the probability function’s derivatives, we obtain the boundary of the kernel as the solution of the differential equation.
2. Methodology

2.1. An overview of the one-dimensional random variable case
Consider complete probability space \((\Omega, F, P)\) and absolutely continuous random variable \(X: \Omega \rightarrow \mathbb{R}\) with probability density function \(f(x)\) on that space. Suppose the smooth loss function \(\Phi(u,x): U \times \mathbb{R} \rightarrow \mathbb{R}\), where \(U \subseteq \mathbb{R}^m\) is a non-empty set of possible control vectors \(u\), is given. Also, suppose that the maximum allowable level of loss \(\varphi\) is fixed.

The probability function is defined as
\[
P_{\varphi}(u) = P\{\Phi(u,X) \leq \varphi\},
\]
and the problem associated with this criterion function is defined as
\[
P_{\varphi}(u) \rightarrow \max_u.
\]

The quantile function is defined as
\[
\varphi_{\alpha}(u) = \left[\Phi(u,x)\right]_u = \min\{\varphi : P_{\varphi}(u) \geq \alpha\},
\]
and the problem associated with this criterion function is defined as
\[
\varphi_{\alpha}(u) \rightarrow \min_u. \tag{2.1}
\]

The probability function can be represented as
\[
P_{\varphi}(u) = \mathbb{E}[I\{\Phi(u,X) \leq \varphi\}] = \int_{-\infty}^{+\infty} I\{\Phi(u,x) \leq \varphi\} f(x) dx = \int_{-\infty}^{+\infty} \Theta(\varphi - \Phi(u,x)) f(x) dx,
\]
where \(I(\cdot)\) and \(\Theta(\cdot)\) are indicator function and Heaviside function respectively:
\[
I\{\Phi(u,x) \leq \varphi\} = \begin{cases} 1, & \Phi(u,x) \leq \varphi \\ 0, & \Phi(u,x) > \varphi \end{cases}, \quad \Theta(y) = \begin{cases} 1, & y \geq 0 \\ 0, & y < 0 \end{cases}.
\]

The main idea of approximation, described in detail in [22], is to replace the Heaviside function in equation (2.2) with its differentiable smooth approximation – the sigmoid function:
\[
S_{\varphi}(y) = (1 + e^{-\varphi y})^{-1} \tag{2.3}
\]
where parameter \(\vartheta\) represents the steepness of the sigmoid and usually is a large positive number. Hence the probability function approximation formula is
\[
P_{\varphi}^\vartheta(u) = \int_{-\infty}^{+\infty} S_{\varphi}(\varphi - \Phi(u,x)) f(x) dx, \tag{2.4}
\]
which further leads to expressions for the approximation of the probability function partials with respect to control vector component \(u_i\) and level of loss:
\[
\frac{\partial P_{\varphi}^\vartheta(u)}{\partial u_i} = \int_{-\infty}^{+\infty} \vartheta S_{\varphi}(\varphi - \Phi(u,x))[S_{\varphi}(\varphi - \Phi(u,x)) - 1] \Phi_i'(u,x) f(x) dx,
\]
\[
\frac{\partial P_{\varphi}^\vartheta(u)}{\partial \varphi} = \int_{-\infty}^{+\infty} \vartheta S_{\varphi}(\varphi - \Phi(u,x))[1 - S_{\varphi}(\varphi - \Phi(u,x))] f(x) dx.
\]

It was proven in [22] that if loss function is strictly piecewise monotonic in addition to the requirements above, the following expressions are true for each \(u_i\) as \(\vartheta \rightarrow \infty\):\[
S_{\varphi}(y) \xrightarrow{a.s.} \Theta(y), \quad P_{\varphi}^\vartheta(u) \rightarrow P_{\varphi}(u), \quad \frac{\partial P_{\varphi}^\vartheta(u)}{\partial u_i} \rightarrow \frac{\partial P_{\varphi}(u)}{\partial u_i}, \quad \frac{\partial P_{\varphi}^\vartheta(u)}{\partial \varphi} \rightarrow \frac{\partial P_{\varphi}(u)}{\partial \varphi}. \tag{2.5}
\]

2.2. Two-dimensional random vector case and its generalization
Consider random vector \((X, Y)^T\) having joint probability density function \(f(x, y)\) and the loss function \(\Phi(u, X, Y)\) is given in addition to previous requirements. Then the probability function can be represented by analogy with the equation (2.2)

\[
P_{\psi}(u) = P[\Phi(u, X, Y) \leq \varphi] = E[1\{\Phi(u, X, Y) \leq \varphi\}] = \int \int \Theta(\varphi - \Phi(u, x, y)) f(x, y) dy dx,
\]

where \(G_x = \text{supp}(X)\), \(G_y = \text{supp}(Y)\), and its approximation formula is similar to equation (2.4)

\[
P_{\psi}^\theta(u) = \int \int S_\theta(\varphi - \Phi(u, x, y)) f(x, y) dy dx.
\]

To prove the convergence of approximation to the original function, consider the conditional probability density function of \(X\) given the value \(y_0\) of \(Y\)

\[
f_{X|Y}(x | y_0) = \frac{f(x, y_0)}{f_Y(y_0)},
\]

where \(f_Y(\cdot)\) is the marginal probability density function of \(Y\). The conditional probability function is defined and represented by analogy with the equation (2.6) as

\[
P_{\psi}^\theta(u | Y = y_0) = P[\Phi(u, X, y_0) \leq \varphi] = \int \Theta(\varphi - \Phi(u, x, y_0)) f_{X|Y}(x | y_0) dx,
\]

and its approximation formula is similar to equation (2.7)

\[
P_{\psi}^{\theta}(u | Y = y_0) = \int \int S_{\theta}(\varphi - \Phi(u, x, y_0)) f_{X|Y}(x | y_0) dx.
\]

As we considering probability density functions of one-dimensional random variables, then statements analogical to those in the equation (2.5), are applicable for these functions, e.g. as \(\theta \to \infty\)

\[
P_{\psi}^\theta(u | Y = y_0) \to P_{\psi}(u | Y = y_0),
\]

\[
\frac{\partial P_{\psi}^\theta(u | Y = y_0)}{\partial \varphi} \to \frac{\partial P_{\psi}(u | Y = y_0)}{\partial \varphi},
\]

\[
\frac{\partial P_{\psi}^\theta(u | Y = y_0)}{\partial u_i} \to \frac{\partial P_{\psi}(u | Y = y_0)}{\partial u_i}.
\]

Consider the approximation of the probability function. According to the law of total probability:

\[
P_{\psi}(u) = \int \int P_{\psi}(u | Y = y_0)f_Y(y_0) dy_0, \quad P_{\psi}^\theta(u) = \int \int P_{\psi}^\theta(u | Y = y_0)f_Y(y_0) dy_0.
\]

The factor \(P_{\psi}^\theta(u | Y = y_0)\) of the integrand is dominated by 1 in the following sense:

\[
\left| P_{\psi}^\theta(u | Y = y_0) \right| \leq 1.
\]

Considering \(P_{\psi}^\theta(u | Y = y_0)\) as a sequence with index \(\theta \in \mathbb{N}\), formulas in the equation (2.11), the fact of convergence in the equation (2.8), and the equation (2.12) itself, it appears that according to Lebesgue's dominated convergence theorem, the following statement takes place:

\[
P_{\psi}^\theta(u) \to P_{\psi}(u) \quad \text{as} \quad \theta \to \infty.
\]

The convergence of partial derivatives approximation to original partial derivatives is proved similarly.

For the case of partials with respect to the loss level, we use similar expressions to those in the equation (2.11), the equation (2.9), and the fact following from [22] that

\[
\left| \frac{\partial P_{\psi}^\theta(u | Y = y_0)}{\partial \varphi} \right| \leq \frac{\theta}{4}.
\]

Thus, according to Lebesgue's dominated convergence theorem:
\[
\frac{\partial P_\varphi^\theta(u)}{\partial \varphi} \to \frac{\partial P_\varphi(u)}{\partial \varphi} \quad \text{as } \theta \to \infty.
\]

For the case of partials with respect to the control vector components \(u_i\), we use similar expressions to those in the equation (2.11) and the equation (2.9) itself, but we demand that
\[
\forall i = 1, m \exists K_i \in \mathbb{R}, K_i < \infty : \left| \frac{\partial \Phi(u,x,y)}{\partial u_i} \right| \leq K_i,
\]
and according to [22], this condition is equivalent to:
\[
\forall i = 1, m \exists K_i \in \mathbb{R}, K_i < \infty : \left| \frac{\partial P_\varphi^\theta(u | Y = y_i)}{\partial u_i} \right| \leq \frac{\theta}{4} K_i.
\]

Thus, according to Lebesgue’s dominated convergence theorem:
\[
\frac{\partial P_\varphi^\theta(u)}{\partial u_i} \to \frac{\partial P_\varphi(u)}{\partial u_i} \quad \text{as } \theta \to \infty.
\]

Formulas for quantile function approximation and its partials approximation follow from one-dimensional analogs and equations of probability function partials similarly to those shown in [22].

This approach of using the lower-dimensional conditional probability density function to prove the convergence of the approximation to the original functions in higher dimensions can be generalized to a case of dimensions higher than 2. Changes in the approximation algorithm for higher dimensions will affect only additional conditions of the loss function corresponding with a considered dimension.

2.3. An application example: approximation of probability measure kernel

The present chapter describes a method of approximate probability measure kernel construction through the construction of its boundaries.

First, consider the basic definitions used below. Let \((\Omega, F, P)\) be a complete probability space.

The set \(S \in F\) is called alpha-confident if \(P(S) \geq \alpha\). Particularly, if the probability measure \(P\) is induced by a random variable \(X\), then the set \(S\) is called alpha-confident if \(P(X \in S) \geq \alpha\).

The intersection of all convex and closed alpha-confident sets related to one probability measure is called a probability measure kernel, usually designated as \(K_\alpha\). Finding a probability measure kernel can be beneficial as it generally allows us to calculate a lower boundary of quantile directly or can be used in various solution methods [23]. Probability measure kernel \(K_\alpha\) is called regular if every closed half-space containing \(K_\alpha\) is alpha-confident.

Consider the half-spaces containing regular probability measure kernel \(K_\alpha\). As by definition each one of them is alpha-confident, then each one of them contains a family of sets with a measure equal to \(\alpha\). If \(\alpha > 0.5\), then all of these sets in the family have a non-empty intersection. If we choose a set from a family for every alpha-confident half-space, then these sets also have a non-empty intersection containing the kernel itself. The hyperplane bounding the alpha-confident half-space can be shifted in such a way that it is tangent to the boundary of the chosen set. Therefore, the boundary of the regular probability measure kernel can be approximated as a boundary of a set, which is an intersection of half-spaces with the measure \(\alpha\).

Consider a random vector \(X = [X_1, X_2]^T\) with a probability density function \(f(x_1, x_2)\). Assume that the median vector of \(X\) coincide with the point \((0,0)\). Also, assume that \(\alpha > 0.5\) and a regular probability measure kernel \(K_\alpha\) exists and contains the point \((0,0)\).

The probability function is defined as a probability of random vector \(X\) falling into half-plane determined by inequality \(c_1 X_1 + c_2 X_2 \leq 1\):
\[
P(c) = P\{c_1 X_1 + c_2 X_2 \leq 1\}.
\]

It can be represented as
\[ P(c) = \iint_{c_1x_1 + c_2x_2 \leq 1} f(x_1, x_2) \, dx_1 \, dx_2 = \iint_{G_X} I(c, x_1, x_2) f(x_1, x_2) \, dx_1 \, dx_2, \quad (2.13) \]

where \( G_X = \text{supp}(X) \) and \( I(c, x_1, x_2) \) is an indicator function:

\[
I(c, x_1, x_2) = \begin{cases} 
1, & c_1 x_1 + c_2 x_2 \leq 1 \\
0, & c_1 x_1 + c_2 x_2 > 1. 
\end{cases} \quad (2.14)
\]

According to the approximation method, the goal is to find an expression for the solution of the equation \( P(c) = \alpha \). The surface of the probability function level corresponds to a set of vectors \( c = [c_1, c_2] \), at which the value of the probability function is constant. Thus, if we consider the probability function as a function with respect to components of the vector \( c \), then the condition of a constant level of the probability function is equal to condition

\[
\frac{\partial P(c)}{\partial c_1} + \frac{\partial P(c)}{\partial c_2} = 0.
\]

Furthermore, it is equivalent to:

\[
\frac{dc_2}{dc_1} = -\frac{\partial P(c)}{\partial c_1} \left( \frac{\partial P(c)}{\partial c_2} \right)^{-1}. \quad (2.15)
\]

The equation (2.15) can be used only to check if the point, which satisfies the condition \( P(c) \geq \alpha \), belongs to the border of the desired set. Considering that the probability measure kernel includes the point \((0,0)\), to build boundary formulas that account for the needed direction we transition to polar coordinates:

\[
c_1 = r \cos t, \quad c_2 = r \sin t. \quad (2.16)
\]

Another problem that lies in the equation (2.15) is that partials on the right-hand side are difficult to calculate. To solve this issue we use the replacement of indicator function by sigmoid defined earlier in the equation (2.3), e.g. we transition to probability function approximation:

\[
\hat{P}(c) = \iint_{G_X} S(1-c_1 x_1 - c_2 x_2) f(x_1, x_2) \, dx_1 \, dx_2. \quad (2.17)
\]

Considering the equation (2.16):

\[
\hat{P}(r,t) = \iint_{G_X} S(1-r \cos t - x_2 r \sin t) f(x_1, x_2) \, dx_1 \, dx_2. \quad (2.17)
\]

Probability function partials in the equation (2.15) become partials with respect to \( r \) and \( t \):

\[
\frac{\partial \hat{P}(r,t)}{\partial r} = -\iint_{G_X} \theta(x_1 \cos t + x_2 \sin t) S'(1-r \cos t - x_2 r \sin t) f(x_1, x_2) \, dx_1 \, dx_2, \quad (2.18)
\]

\[
\frac{\partial \hat{P}(r,t)}{\partial t} = -\iint_{G_X} \theta r(x_1 \cos t - x_2 \sin t) S'(1-r \cos t - x_2 r \sin t) f(x_1, x_2) \, dx_1 \, dx_2, \quad (2.19)
\]

and the equation (2.15) transitions to:

\[
\frac{dr}{dt} = -\frac{\partial \hat{P}(r,t)}{\partial t} \left( \frac{\partial \hat{P}(r,t)}{\partial r} \right)^{-1}. \quad (2.20)
\]

Now we describe the approximation algorithm for the kernel boundary:

1. Set the initial approximation of the variables \( t_0 = 0, \quad r_0 = \left( [X_1]_r \right)^{-1} \) and step value \( \Delta t \);
2. Calculate partial derivatives using numerical integration according to the equation (2.18) and equation (2.19) for current values \( t_k \) and \( r_k \);
3. Calculate the value of the derivative \( dr/dt \) according to the equation (2.20) and calculate the approximate value \( r_{k+1} \) using this value (for example, using the Euler method);
4. Set a new current value \( t_{k+1} = t_k + \Delta t \);
5. Repeat steps 2-4 until $t_{k+1} \leq 2\pi$.

The result of the algorithm is a piecewise determined function $r(t)$, which describes the external boundary of the probability measure kernel. After inverse transformation, this function represents a relation between components $c_1$ and $c_2$, determining a borderline for each alpha-confident half-plane.

The probability measure kernel approximation algorithm and visualization of obtained results were implemented using the Python programming language.

3. Results and discussion
Consider the results of the approximation algorithm for different two-dimensional distributions.

For a random vector $X$ having a uniform distribution with given algorithm parameters

$$X \sim \text{U}([-1,1] \times [-1,1]), \quad \alpha = 0.8, \quad \theta = 10, \quad \Delta t = 0.01, \quad t_0 = 0, \quad r_0 = 1.667,$$

the result of the algorithm is shown in figure 1. The approximate value of $r_{k+1}$ calculates using the standard Euler method. For a random vector $X$ having a Gaussian distribution with given algorithm parameters

$$X \sim \text{N}(0,0,\begin{pmatrix} 2 & 0.7 \\ 0.7 & 1 \end{pmatrix}), \quad \alpha = 0.8, \quad \theta = 10, \quad \Delta t = 0.01, \quad t_0 = 0, \quad r_0 = 0.8402,$$

the result of the algorithm is shown in figure 2.

According to figure 1, the boundary of the probability measure kernel in the case of the uniform distribution appears to be composed of four hyperbolic line segments. This result is well correlated to the results obtained in other papers dedicated to kernel construction and approximation, for example, with the results of [24]. According to figure 2, the boundary of the probability measure kernel in the case of the Gaussian distribution appears to be an ellipsoid, and this fact also correlates well with the results of other papers [24].

The result of the algorithm for an exponentially distributed random vector $X = [X_1, X_2]$ with algorithm parameters stated below is shown in figure 3.

$$X_1 \sim \text{E}(1), \quad X_2 \sim \text{E}(0.5), \quad \alpha = 0.8, \quad \theta = 10, \quad \Delta t = 0.01, \quad t_0 = 0, \quad r_0 = 1.6409.$$
4. Practical applications

In the present section, we discuss some practical applications of the results. A straightforward application of the described approximations was presented in [22]. The task was to find the minimal possible area of an airstrip that guarantees the successful landing of an aircraft with a high enough probability. We used two-dimensional Gaussian distribution to model the components of the wind speed vector, and a simplified model to determine the random bias of the aircraft from the desired landing point due to the wind. The components of the bias vector are random and dependent. Although all results in [22] were derived in a one-dimensional case, this example shown the applicability of proposed approximations in at least a two-dimensional case.

Another possible application is the construction of the confidence absorbing sets as in [25]. The confidence absorbing set is understood as the set of initial positions for which at a terminal time instant a system will not leave an admissible domain with a given probability. An algorithm for constructing an approximation of the confidence absorbing set can be based on the proposed algorithm of constructing kernel approximation. Under some assumptions, the problem also transforms to the problem of finding the isoquant of the probability function, where the initial position is treated as the input of the probability function. The results in [25] are used to construct the set of all possible wind speed vectors, measured at the departure time, which guarantees with a given probability, that the wind at the estimated arrival time would not exceed the allowable limits. Under the given model, the distribution of future wind speed components depends on the values obtained at the initial point and shifts in speed and direction are considered to have Gaussian distribution. The application of the smooth approximation of the probability function to the problem of construction of the confidence absorbing sets is a subject of further research.

Given approximations in the multidimensional case can be applied to the problem of the design of the water supply system in a desert region, described in [26]. The system includes a water desalination plant, solar cells that provide energy for this plant, and a reservoir for freshwater storage. The price of such a system depends on the area of solar cells and the volume of the reservoir. The stochastic aspect of the problem is random solar activity. The task is to create the cheapest system that meets the requirements in freshwater with a high enough probability. This task will also be considered in our future research. And the last application we discuss here is the application of probability measure’s kernel to the problem of quantile minimization. As shown in [26] the quantile minimization problem can be transformed into a minimax problem in the case of a bilinear loss function. In this minimax task, we minimize the loss function with respect to the control vector and maximize it with respect to the values.
of the random vector under the given control vector (inner maximization). It was proven that the inner maximum must be obtained within the probability measure kernel. So the probability measure kernel can be used for the solution of quantile minimization problems with linear objectives (as in some economic tasks) and with bilinear objectives (e.g. in gas pooling problem), and some nonlinear objectives where a bilinear function can be used as a good approximation.

5. Conclusion
In the present paper, we generalized the previous results on the smooth approximation of the probability and quantile functions to the case of a multidimensional random vector. The method demands only the limited loss function partials in addition to the requirements for a one-dimensional random variable. The convergence of approximations to their original functions for a two-dimensional random vector was shown. All formulas given have the form of the integral over volume or the form of the unconditional expectation, so they can be effectively calculated.

Based on the probability function approximations we provide a new way to determine the boundary of the kernel of probability measure as the solution to a differential equation. This problem is important both theoretically and practically since the probability measure’s kernel can be used to solve the stochastic optimization problems with an objective in the form of the quantile function and linear or bilinear loss function. The same approach as we used in the probability measure’s kernel construction can be applied to the problem of constructing the confidence absorbing sets, as we construct an isoquant of the specific probability function.

We also discuss possible applications of the smooth approximation in applied stochastic programming tasks and discuss further research.

Acknowledgments
The reported study was funded by RFBR according to research project No 20-31-90035.

References
[1] Lipowsky R 2000 Molecular Motors and Stochastic Models Stochastic Processes in Physics, Chemistry, and Biology eds J A Freund and T Pöschel (Berlin: Springer-Verlag Berlin Heidelberg) pp 21-31 doi: 10.1007/3-540-45396-2_2
[2] Marchisio M A 2018 Stochastic Modeling Introduction to Synthetic Biology. Learning Materials in Biosciences (Singapore: Springer Singapore) pp 39-52 doi: 10.1007/978-981-10-8752-3_4
[3] Dentcheva D, Penev S and Ruszczynski A 2017 Statistical estimation of composite risk functionals and risk optimization problems. Ann. Inst. Stat. Math. 69 737 doi: 10.1007/s10463-016-0559-8
[4] Dentcheva D and Ruszczynski A 2019 Risk forms: representation, disintegration, and application to partially observable two-stage systems. Math. Program. 181 297 doi: 10.1007/s10107-019-01376-1
[5] Kibzun A, Naumov A and Norkin V 2013 On reducing a quantile optimization problem with discrete distribution to a mixed integer programming problem. Automat. Rem. Contr. 74 951 doi: 10.1134/S0005117913060064
[6] Kibzun A and Ignatov A 2016 Reduction of the two-step problem of stochastic optimal control with bilinear model to the problem of mixed integer linear programming. Automat. Rem. Contr. 77 2175 doi: 10.1134/S0005117916120079
[7] Ivanov S and Kibzun A 2018 Sample average approximation in a two-stage stochastic linear program with quantile criterion. Proc. Steklov Inst. Math. 303 115 doi: 10.1134/S0081543818090122
[8] Raik E 1975 The differentiability in the parameter of the probability function and optimization of the probability function via the stochastic pseudogradient method. Proc. of Academy of Sciences of the Estonian SSR. Physics. Mathematics 24(1) 3
[9] Kibzun A and Tretyakov G 1997 On the smoothness of criteria function in quantile optimization.
10

Automat. Rem. Contr. 58(9) 1459

[10] Marti K 1994 Approximations and Derivatives of Probability Functions Approximation, Probability, and Related Fields eds G Anastassiou and S T Rachev (Boston: Springer, Boston, MA) pp 367-377 doi: 10.1007/978-1-4615-2494-6_28

[11] Marti K 1996 Differentiation formulas for probability functions: The transformation method. Math. Program. 75 201 doi: 10.1007/BF02592152

[12] Uryas’ev S 1995 Derivatives of probability functions and some applications. Ann. Oper. Res. 56 287 doi: 10.1007/BF02031712

[13] Uryas’ev S 1994 Derivatives of probability functions and integrals over sets given by inequalities. J. Comput. Appl. Math. 56(1-2) 197 doi: 10.1016/0377-0427(94)90388-3

[14] Henrion R 2012 Gradient estimates for Gaussian distribution functions: application to probabilistically constrained optimization problems. Numerical Algebra, Control and Optimization 2(4) 655 doi: 10.3934/naco.2012.2.655

[15] van Ackooij W and Henrion R 2017 (Sub-)Gradient formulae for probability functions of random inequality systems under Gaussian distribution. SIAM/ASA J. Uncertainty Quantification 5(3) 63 doi: 10.1137/16M1061308

[16] Pflug G and Weißhaupt H 2005 Probability gradient estimation by set-valued calculus and applications in network design. SIAM J. Optimiz. 15(3) 898 doi: 10.1137/S1052623403431639

[17] Yu C and Zelterman D 2017 A general approximation to quantiles. Commun. Stat. Theory. Methods. 46(19) 9834 doi: 10.1080/03610926.2016.1222433

[18] Okagbue H, Adamu M and Anake T 2018 Ordinary differential equations of the probability functions of the weibull distribution and their application in ecology. Int. J. Eng. Future. Tech. 15 57

[19] Garnier J, Omrane A and Rouxchdy Y 2009 Asymptotic formulas for the derivatives of probability functions and their Monte Carlo estimations. Eur. J. Oper. Res. 198(3) 848 doi: 10.1016/j.ejor.2008.09.026

[20] Mafusalov A and Uryasev S 2018 Buffered probability of exceedance: mathematical properties and optimization. SIAM J. Optimiz. 28(2) 1077 doi: 10.1137/15M1042644

[21] Zhang T, Uryasev S and Guan Y 2019 Derivatives and subderivatives of buffered probability of exceedance. Oper. Res. Lett. 47(2) 130 doi: 10.1016/j.orl.2019.02.002

[22] Sobol V and Torishnyi R 2020 On smooth approximation of probabilistic criteria in stochastic programming problems. SPIIRAS Proceedings 19(1) 181 doi:10.15622/sp.2020.19.1.7

[23] Vasil’eva S and Kan Y 2019 Approximation of probabilistic constraints in stochastic programming problems with a probability measure kernel. Automat. Rem. Contr. 80 2005 doi: 10.1134/S0005117919110055

[24] Vasil’eva S and Kan Y 2018 A visualization algorithm for the plane probability measure kernel Informatics and Applications 12(2) 60 doi: 10.14357/19922264180209

[25] Kibzun A, Ivanov S and Stepanova A 2020 Construction of confidence absorbing set for analysis of static stochastic systems. Automat. Rem. Contr. 81 589 doi: 10.1134/S0005117920040025

[26] Kibzun A and Kan Y 1996 Stochastic programming problems with probability and quantile functions (London: John Wiley & Sons) p 301