A global optimization approach to maximization of the probability function

O Khamisol
Melentiev Energy System Institute, 130 Lermontov street, Irkutsk, 664033, Russian Federation

E-mail: khamisov@isem.irk.ru

Abstract. We suggest an approach that reduces the problem of maximizing the probability function to a deterministic global optimization problem. The reduction technique is described and discussed. A numerical example illustrating efficiency of the final algorithm is provided.

1. Introduction

In our paper, we consider the following optimization problem:

\[
\max \{ \nu(x) = \mathbb{P}(\xi : f(x, \xi) \leq t) \}, \quad x \in X,
\]

(1.1)

where \(X \subseteq \mathbb{R}^n\) is a convex compact set, \(\xi \in \Xi \subseteq \mathbb{R}^q\) is a random vector, a scalar \(t\) is fixed, \(\mathbb{P}\) denotes a probability measure. The properties of function \(f : \mathbb{R}^n \times \mathbb{R}^q \rightarrow \mathbb{R}\) will be given below.

In stochastic programming, problem (1.1)-(1.2) is known as the probability function maximization problem. Function \(\nu(x)\) in (1.1) is called a probability function. Function \(f(x, \xi)\) is called a loss or cost function. The inequality in (1.1) reflects the fact that we must not exceed the given loss level \(t\). The value of decision vector \(x\) is determined under stochastic uncertainty conditions generated by random vector \(\xi\). The maximization in (1.1) corresponds to finding the highest probability that the loss is less or equal to level \(t\).

Problem (1.1)-(1.2) has quite a number of applications, such as the stock exchange paradox problem, the supply optimization and shopkeeper’s problems under stochastic uncertainty of some parameters, and many others (see [1, 2] and references therein).

A good review on problems with probability functions is given in [3, 4]. For a fixed \(x\), the expression in (1.1) is a (cumulative) distribution function. Hence, value \(\nu(x)\) belongs to the interval \([0, 1]\) and therefore function \(\nu(x)\) is neither convex nor concave. Generalized concavity (convexity) properties of probability functions are investigated in [5-7]. These properties are based on properties of the loss function or the probability measure. A general analysis, ideas, and approaches related to the generalized concavity (convexity) properties, which are used in stochastic programming, can be found in [8]. The differentiability properties of probability functions and an estimation of the gradients are studied in [9-14].
In this paper, we suggest quite a general approach based on the technique of global optimization with nonlinear support functions [15] and do not assume any kind of smoothness. A stochastic approximation technique [16] is used to construct a deterministic approximation of problem (1.1)-(1.2). The deterministic approximation is a difficult global optimization problem with a nonconvex and not necessarily continuous objective function.

Below we assume that function \( f(x, \xi) \) satisfies the following property. There exists a function \( \phi(x, y, \xi) \) such that it is concave and continuous in \( x \) for fixed \( y \in X \) and \( \xi \in \Xi \) and

1. \( f(x, \xi) \geq \phi(x, y, \xi) \quad \forall x, y \in X, \forall \xi \in \Xi \); \hspace{1cm} (1.3)

2. \( f(y, \xi) = \phi(y, y, \xi) \quad \forall y \in X, \forall \xi \in \Xi \). \hspace{1cm} (1.4)

Function \( \phi(x, y, \xi) \) is called a concave minorant. The class of functions satisfying this property is rather wide. It includes, for example, all functions \( f(x, \xi) \) that are continuously differentiable in \( x \) for a fixed \( \xi \) or Lipschitz in \( x \) for a fixed \( \xi \) [15].

2. The algorithm
First of all, we will describe a reduction of problem (1.1)-(1.2) to an approximate deterministic problem. From the definition of the probability function, we have

\[
v(x) = \int_{\{\xi \in \Xi | f(x, \xi) \geq t\}} P(d\xi), \hspace{1cm} (2.1)
\]

By using the Heaviside function

\[
\chi(s) = \begin{cases} 0, & s < 0, \\ 1, & s \geq 0, \end{cases}
\]

the integral in (2.1) can be rewritten in the following form:

\[
v(x) = \int_{\Xi} \chi(t - f(x, \xi))P(d\xi) \hspace{1cm} (2.3)
\]

(see, for example, [3,4,16]). If the distribution function \( H(\xi) \) is available, then

\[
v(x) = \int_{\Xi} \chi(t - f(x, \xi))dH(\xi) \hspace{1cm} (2.4)
\]

If the probability density function \( h(\xi) \) is available, then

\[
v(x) = \int_{\Xi} \chi(t - f(x, \xi))h(\xi)d\xi. \hspace{1cm} (2.5)
\]

With a sample \( \xi_1, ..., \xi_m \), we can approximate function \( v \) by the following one:

\[
v_m(x) = \frac{1}{m} \sum_{i=1}^{m} [\chi(t - f(x, \xi_i))]. \hspace{1cm} (2.6)
\]

In quite a standard way, we formulate the approximate problem

\[
\max \{ v_m(x) : x \in X \}. \hspace{1cm} (2.7)
\]
Function $f$ with properties (1.3), (1.4) is shown ([15]) to be lower semicontinuous in $x$ for any fixed $\xi$. From this point on, we will assume that $f$ has even a stronger property: it is lower semicontinuous in both variables $(x, \xi)$. In this case, function $f$ is called a normal integrand. In [17], the following statement is proved. If $X$ is a nonempty compact set and $f$ is a normal integrand, then

$$\lim_{m \to \infty} \max \{ v_m(x) : x \in X \} = v^* \quad (\text{a.s.})$$

(2.8)

and every accumulation point $\bar{x}$ of sequence $x^m, m = 1, 2, \ldots$, where

$$x^m \in \text{Arg} \max \{ v_m(x) : x \in X \},$$

is a solution of problem (1.1)-(1.2) (a.s.), and $v^*$ is the optimal value of problem (1.1)-(1.2). The abbreviation a.s. means almost surely. Therefore, using a large sample, we can try to find an approximate solution of the probability function maximization problem.

Let us now consider the auxiliary problem (2.7). We will start with the following example.

Example 1. The loss function is $f(x, \xi) = \sin(2\xi^2 + x \xi)$, the random variable $\xi$ is normally distributed with an expectation 0 and a standard deviation $\frac{2}{3}$, $t = 0.75$, $X = [-5, 5]$.

For a sample of size $m = 50$, the plot of the corresponding function $v_{50}(x)$ is shown in figure 1.

![Figure 1](image_url)

**Figure 1.** Plot of function $v_{50}(x)$, solution $x^*_{50} = -0.974$, $v^*_{50} = 0.9$.

From figure 1 it is clearly seen that function $v_m(x)$ can be a multie xtremal function. Therefore, problem (2.7) is in general a hard global optimization problem.

We suggest to find the global maximum of $v_m(x)$ by means of convex majorants. Function $\chi(s)$ has a convex majorant function $\psi_s(s, r)$ determined by the following formula:

$$\chi(s) \leq \psi_s(s, r) = \begin{cases} \max\{0.1 - \frac{s}{r}\}, & r < 0, \\ 1, & r \geq 0. \end{cases}$$

(2.10)
Let $i$ be fixed, $s = t - f(x, \xi_i), r = t - f(y, \xi_i), x, y \in X$. Then, from (2.10), we have

$$
\chi(t - f(x, \xi_i)) \leq \max\{0, 1 - \frac{t - f(x, \xi_i)}{t - f(y, \xi_i)}\}, t - f(y, \xi_i) < 0, 1, t - f(y, \xi_i) \geq 0.
$$

(2.11)

Due to the property of function $f(x, \xi)$, there exists a continuous concave in $x$ function $\varphi(x, y, \xi)$ such that for any fixed $y \in X$ we have

$$
f(x, \xi_i) \geq \varphi(x, y, \xi_i), \forall x \in X,
$$

(2.12)

$$
f(y, \xi_i) = \varphi(y, y, \xi_i).
$$

(2.13)

If $t - f(y, \xi_i) < 0$, then it follows from (2.12) that

$$
1 - \frac{t - f(x, \xi_i)}{t - f(y, \xi_i)} \leq 1 - \frac{t - \varphi(x, y, \xi_i)}{t - f(y, \xi_i)}.
$$

(2.14)

Let us introduce function

$$
\phi(x, y, \xi_i) = \max\{0, 1 - \frac{t - \varphi(x, y, \xi_i)}{t - f(y, \xi_i)}\}.
$$

(2.15)

Note that $\phi(x, y, \xi_i)$ is convex in $x$. Then, from (2.11), (2.14), and (2.15) we have

$$
\chi(t - f(x, \xi_i)) \leq \left\{ \begin{array}{ll}
\phi(x, y, \xi_i), & t - f(y, \xi_i) < 0, \\
1, & t - f(y, \xi_i) \geq 0.
\end{array} \right.
$$

(2.16)

Let us form the following sets: $I^{-}(y) = \{i : t - f(y, \xi_i) < 0\}, I^{+}(y) = \{i : t - f(y, \xi_i) \geq 0\}, m(y) = |I^{+}(y)|$. It follows from (2.13) and (2.16) that function $v_{m}(x)$ in (2.7) has the following convex in $x$ majorant $\psi_{x}(x, y)$:

$$
v_{m}(x) \leq \psi_{x}(x, y) = \frac{m(y)}{m} + \frac{1}{m} \sum_{i \in I^{-}(y)} \phi(x, y, \xi_i).
$$

(2.17)

Now we can describe the algorithm for solving problem (2.7).

**Initialization.** Let $x^{0} \in X, \varepsilon > 0$. Set $k = 0, v_{m}^{0} = v_{m}(x^{0}), \varepsilon^{0} = \chi^{0}$.

**Iteration** $k(k \geq 0)$.

**k.1.** Solve the global optimization problem

$$
\max\{h_{k}(x) = \min_{0 \leq j \leq k} \psi_{x}(x, x^{j})\},
$$

(2.18)

$$
x \in X.
$$

(2.19)

Let $x^{k+1}$ be the solution of problem (2.18)-(2.19).

**k.2.** If $v_{m}(x^{k+1}) > v_{m}^{k}$, then $v_{m}^{k+1} = v_{m}(x^{k+1}), \varepsilon^{k+1} = \chi^{k+1}$. Otherwise, $v_{m}^{k+1} = v_{m}^{k}, \varepsilon^{k+1} = \varepsilon^{k}$.

**k.3.** If $h(x^{k+1}) - v_{m}^{k+1} \leq \varepsilon$, stop. $\varepsilon^{k+1}$ is an $\varepsilon$-optimal solution of problem (2.2).

**k.4.** Update $k = k + 1$ and goto k.1.
If functions \( h_k(x) \) form a family of equicontinuous and uniformly bounded on \( X \) functions, then every accumulation point of sequence \( \{x^k\} \) is the global minimum point of problem (2.7) [18]. Such a convergence condition is not very restrictive and often is met in practice. Problem (2.18)-(2.19) is an explicit global optimization problem of maximization of the minimum of a family of convex functions over a compact set. Such a problem can be reduced to a concave programming problem, and different concave programming methods can be applied [19].

Example 2 (Example 1 continued). Consider the problem introduced in example 1. Let us describe how a concave minorant satisfying (1.3), (1.4) for the given loss function \( f(x, \xi) = \sin(x\xi + \xi^2) \) can be constructed. For any fixed \( w \), function \( \sin(u) \) can be represented as

\[
\sin(u) = g(u) - h(u), \quad g(u) = \sin(u) + 0.5(u - w)^2, \quad h(u) = 0.5(u - w)^2. \tag{2.20}
\]

Both functions \( g(u) \) and \( h(u) \) are convex. Then a concave minorant is obtained by linearization of \( g(u) \) at \( w \),

\[
\sin(u) \geq g(w) + g'(w)(u - w) - 0.5(u - w)^2 = \sin(w) + \cos(w)(u - w) - 0.5(u - w)^2. \tag{2.21}
\]

The inequality in (2.21) holds due to the convexity of \( g(u) \). Note that the right-hand side function in (2.21) is concave in \( u \). Now, we apply (2.21) with substitution \( u = x\xi + \xi^2 \) and \( w = y\xi + \xi^2 \) to our loss function

\[
f(x, \xi) = \sin(x\xi + \xi^2) \geq \sin(y\xi + \xi^2) + \cos(y\xi + \xi^2)(x - y)\xi - 0.5(x - y)^2 \xi^2. \tag{2.22}
\]

Hence, we can use the function

\[
\phi(x, y, \xi) = \sin(y\xi + \xi^2) + \cos(y\xi + \xi^2)(x - y)\xi - 0.5(x - y)^2 \xi^2 \tag{2.23}
\]
as a concave (in \( x \)) minorant. Then we apply the above algorithm to globally maximize function \( v_{30}(x) \) over \( X = [-5,5] \). The algorithm stops after 35 iterations having obtained the global maximum of \( v_{30}(x) \) within the given tolerance \( \varepsilon = 0.001 \). In figure 2, the objective function \( v_{30}(x) \) and the upper approximating function \( h_{35}(x) \) are graphed in green and red correspondingly (see (2.18)).

![Figure 2](image)

Figure 2. Objective function \( v_{30}(x) \) (green), approximating function \( h_{35}(x) \) (red),

\[
v_{30}^{36} = 0.90061, \ z_{36}^{36} = -0.9736.
\]
3. Conclusion
In this paper, an approach based on a global optimization deterministic technique is used for solving the probability function maximization problem. We think that this approach can be applied to problems with quite a wide set of loss functions. For example, many functions can be expressed in the form (2.20). Representing a function as a difference of two convex functions is an effective tool in global optimization [19]. A question that is still under thorough investigation is the sample size $m$ in problem (2.7), and it is going to become the subject of our next work.

Acknowledgments
The investigation was supported by project III.17.3.1 of program of fundamental investigations SB RAS, No. AAAA-A17-117030310444-2.

References
[1] Kibzun A and Kan Yu 1996 Stochastic programming problems with probability and quantile functions (London: John Wiley & Sons)
[2] Norkin V 1993 The analysis and optimization of probability functions IIASA Working Paper, IIASA, Luxenburg, Austria
[3] Lepp R 2009 Approximation of extremum problems with probability functionals Floudas C.A., Pardalos P.M. (eds) Encyclopedia of Optimization. Springer, Boston, MA 78-83
[4] Lepp R 2009 Extremum problems with probability functions: kernel type solution methods Floudas C.A., Pardalos P.M. (eds) Encyclopedia of Optimization. Springer, Boston, MA 969-73
[5] Norkin V and Roenko N 1991 $\alpha$-concave functions and measures and their applications Cybernetics and Systems Analysis 27 77-88
[6] Kibzun A and Kan Yu 1996 Convexity properties of probability and quantile functions in optimization problems Automation and Remote Control 3 820-2
[7] Kibzun A and Matveev E 2010 Sufficient conditions for quasiconcavity of the probability function Automation and Remote Control 71 413-30
[8] Prekopa A 1995 Stochastic programming Series: Mathematics and Its Applications, Kluwer Academic Publishers, Dordrecht
[9] Marti K 1995 Differentiation of probability functions: the transformation method Computers Math. Applic. 30 361-82
[10] Uryasev S 1995 Derivatives of probability functions and some applications Annals of Operations Research 56 287-311
[11] Pflug G and Weiszpuft H 2005 Probability gradient estimation by set-valued calculus and applications in network design SIAM J. Optim. 15 (3) 898-914
[12] Garnier J, Omrane A and Rouchdy Y 2009 Asymptotic formulas for the derivatives of probability functions and their Monte-Carlo estimations European Journal of Operational Research 198 848-59
[13] Van Ackooij W, Aleksovska I and Munoz-Zuniga M 2017 (Sub-) Differentiability of probability functions with elliptical distributions SetValued Var. Anal. https://doi.org/10.1007/s11228-017-0454-3
[14] 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 63-87
[15] Khamisov O 1999 On optimization properties of functions with a concave minorant Journal of Global Optimization 14 79-101
[16] Lepp R 1983 Stochastic approximation type algorithm for the maximization of the probability function Proc. Estonian Acad. Sci., Phys.Math. 32 150-6
[17] Ivanov S and Kibzun A 2018 On the convergence of sample approximations for stochastic programming problems with probabilistic criteria Automation and Remote Control 79(2)
216-28

[18] Bulatov V 1990 Methods for solving multi-extremal problems (global search) *Annals of Operations Research* 25 253-78

[19] Horst R and Tuy H 1996 *Global Optimization. Deterministic Approaches* (Berlin: Springer)