International Institute for Applied Systems Analysis
IIASA www.iiasa.ac.at

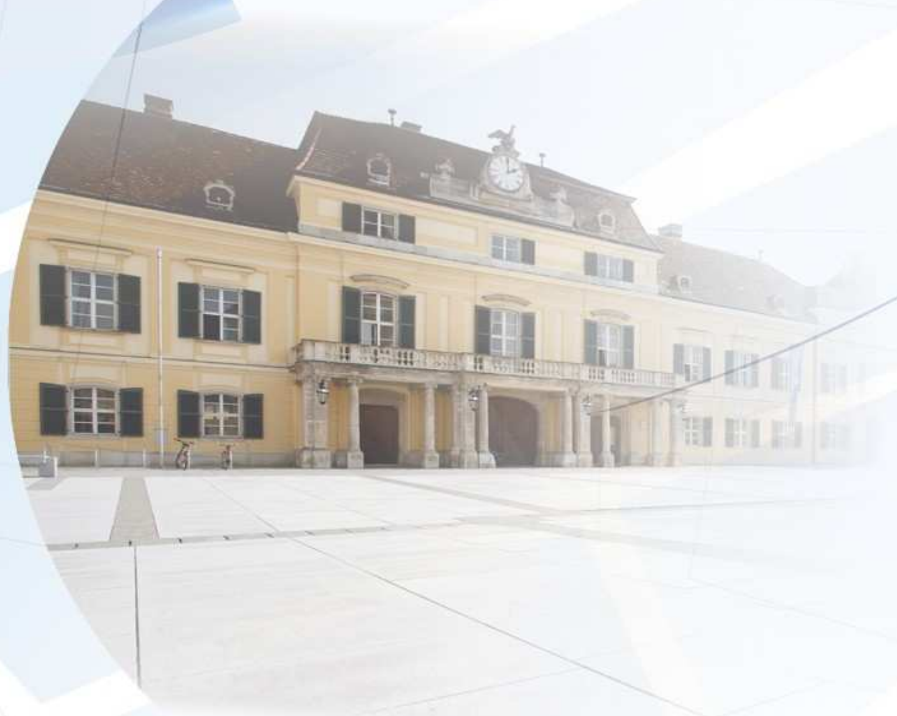
The Analysis and Optimization of Probability Functions

Norkin, V.I.

IIASA Working Paper

WP-93-006

January 1993



Working Paper

The analysis and optimization of probability functions

Vladimir I. Norkin

WP-93-6
January 1993

IIASA International Institute for Applied Systems Analysis □ A-2361 Laxenburg □ Austria
Telephone: +43 2236 715210 □ Telex: 079 137 iiasa a □ Telefax: +43 2236 71313

The analysis and optimization of probability functions

Vladimir I. Norkin

WP-93-6
January 1993

IIASA International Institute for Applied Systems Analysis ◻ A-2361 Laxenburg ◻ Austria
Telephone: +43 2236 715210 ◻ Telex: 079 137 iiasa a ◻ Telefax: +43 2236 71313

Abstract

A problem of probability function optimization is considered. This function represents probability that some random quantity depending on deterministic parameters does not exceed some given level. The problem is motivated by studies of safety domains and risk control problems in complex stochastic systems. For example, pollution control includes maximization of probability that some given levels of deposition at reception points are not exceeded. Optimization of probability function is performed over a given range of parameters. To solve the problem stochastic quasi-gradient method is applied under quasi-concavity assumption on functions and measures involved. Convergence and rate of convergence results are presented.

Keywords: risk, probability function, nonsmooth optimization, stochastic quasi-gradient method, quasi-concavity, α-concavity.

Contents

1 Introduction **1**

2 Properties of probability functions **4**
2.1 Notations 4
2.2 Continuity 4
2.3 Quasi-concave and α-concave functions 5
2.4 Quasi-concave and α-concave measures 6
2.5 Examples of α-concave functions and measures 7
2.6 Quasi-concavity and α-concavity of probability functions 9

3 Approximation of probability functions **9**
3.1 Smoothing of probability functions 9
3.2 Convergence of approximations 10
3.3 α-Concavity of approximations 12
3.4 Differentiability properties of approximations 13

4 Approximate optimization of the probability function **13**
4.1 Problem formulation 13
4.2 Stochastic quasi-gradient method 15
4.3 Convergence results 16
4.4 Rate of convergence 17

5 Conclusions **19**

The analysis and optimization of probability functions

Vladimir I. Norkin

1 Introduction

Probability function represents the probability that some random quantity, depending on controlled parameters, exceeds some given level. The study of probability functions properties and the development of appropriate optimization techniques was motivated by studies of safety domains and risk control problems in complex stochastic systems. As an example, problems of environment monitoring and control can be considered. In this case probability functions reflect the risks to exceed permitted levels of depositions at reception points.

In the present paper we discuss optimization problems for rather general (in particular nonsmooth) probability functions, which cannot be optimized by the existing methods. We also discuss connections between probability functions optimization and classical decision-making problems in inventory theory, two-stage planning, production planning under random supplies and others.

As a solution technique we propose special modifications of stochastic quasi-gradient methods (see Ermoliev [7]), for which convergence rate estimates are obtained.

Let us consider the problem of decision-making under stochastic uncertainty. Let vector x denote possible solutions (alternatives) from a feasible set X. Rational decision choice is made by taking into account their consequences. But these consequences often depend not only on the decision x but also on some random factors ω from some space Ω. The connections between solution x and its consequence y can be written in the form of functional dependence $y = \vec{f}(x,\omega)$, where the transformation $\vec{f} : X \times \Omega \longrightarrow Y$ is called a model (of a decision-making situation); a process of calculation of y for given x and given or statistically simulated ω is called a simulation process. The model can be described by algebraic relationships with random parameters, stochastic differential equations, Markov random processes and other controllable stochastic processes. Since parameters ω are uncertain or random, then with each solution x a corresponding vector-function of consequences $\vec{f}(x,\cdot)$ is associated. Generally, all consequences can be described by loss (expenses etc.) and gain (efficiency etc.). Moreover, to simplify the problem of decision-making, we shall assume that all consequences of decision x are characterized by a single "loss" or "gain" scalar function $f(x,\omega)$. Let us consider some examples of such functions from a number of economic applications.

Example 1.1 *(A choice of stores). Let it be necessary to prepare an inventory of n goods in quantities $(x_1,\ldots,x_n) = x$ for which there exists random demand $(\omega_1,\ldots,\omega_n) = \omega$. The lack of stored goods is penalized by coefficients $(c_1,\ldots,c_n) = c$ and expenses for keeping unsold goods are given by the vector $(d_1,\ldots,d_n) = d$. Then the loss function corresponding to solution x has the form:*

$$f(x,\omega) = \sum_{i=1}^{n}\{c_i \max(0,\omega_i - x_i) + d_i \max(0, x_i - \omega_i)\}.$$

Note that this loss function $f(x,\omega)$ is convex with respect to the pair of variables (x,ω).

Example 1.2 *(Supply optimization). Let some manufacturing system produce a product from some basic ingredients. Let us consider the production function of this system $f(x) = f(x_1,\dots,x_n)$ which expresses the output of the product if ingredients are taken in quantities $x_1,\dots,x_n$. A natural assumption accepted in mathematical economy is that production functions are concave in their variables. For example, the following production function*

$$f_1(x) = \min\{x_1/a_1,\dots,x_n/a_n\}$$

is concave, where $a_1,\dots,a_n$ are technological constants (numbers of ingredients necessary to produce a unit of resulting product). Cobb-Douglas production function

$$f_2(x) = Cx_1^{\alpha_1}\cdot\ldots\cdot x_n^{\alpha_n},$$

where $0 < \alpha_i \le 1$, $i = 1,\dots,n$, is logarithmic concave. Part of the ingredients in such production functions can be taken in deterministic numbers (solution) and others are determined by random supply. So, in general, one can assume that a production function f_i depends on a deterministic vector x and a random vector ω and f_i is quasi-concave in the pair of variables (x,ω).

Example 1.3 *(Shopkeeper's problem). Let a shopkeeper take from a store n kinds of goods in quantities $x_1,\dots,x_n$ for daily selling. Suppose there are (random) daily demands $\omega_1,\dots,\omega_n$ of these goods. The shopkeeper's goal is to maximize the following daily gain*

$$f(x,\omega) = \sum_{i=1}^{n} p_i \min(x_i,\omega_i),$$

where p_i is the unit price of the i-th good, $x = (x_1,\dots,x_n)$, $\omega = (\omega_1,\dots,\omega_n)$. Solution x must satisfy availability restrictions

$$0 \le x_i \le b_i,$$

and sale room restrictions

$$\sum_{i=1}^{n} c_i x_i \le d,$$

where c_i is a space taken by a unit of i-th commodity and d is the volume of the sale room. Let us observe that function $f(x,\omega)$ is jointly concave in (x,ω).

Example 1.4 *(Two-stage decision-making). Let some decision be made in two stages: at first a priori decision $x \in X$ is made, then some random factors $\omega \in \Omega$ are observed and finally some optimal correction y from the set $Y(x,\omega)$ is chosen. Suppose expenditures for decision x are given by a function $f_1(x)$ and expenditures for correction y under given x and ω are given by a function $f_2(x,y,\omega)$. Optimal correction $y^*(x,\omega)$ is chosen as a solution of the problem:*

$$\min_{y\in Y(x,\omega)} f_2(x,y,\omega).$$

Thus the consequences of decision x are described by the following random loss vector-function

$$\vec{f}(x) = \left\{ \begin{array}{l} f_1(x) \\ f_2(x,y^*(x,\omega),\omega) \end{array} \right\}.$$

Components $f_1(x)$ *and* $f_2(x, y^*(x,\omega), \omega)$ *of this function are related to different time intervals so they can be summed only with some discount multiplier* λ *determined by a decision-maker. Therefore total reduced expenditures for solution* x *are given by the expression*

$$f(x,\omega) = f_1(x) + \lambda f_2(x, y^*(x,\omega), \omega), \quad \lambda > 0.$$

If the functions f_1, f_2 *and the multivalued mapping* Y *are convex jointly in* (x, y, ω) *then* $f(x,\omega)$ *is also convex jointly in* (x,ω).

A decision-maker while considering possible solution x should take into account all possible values of loss-gain function $f(x,\cdot)$. Formally it means that the decision-maker's preferences are given in a functional space containing functions of ω. The decision-maker should decide which distribution $\{f(x,\cdot)\}_{x\in X}$ is the most preferable for him.

In general the choice of the most preferable distribution is a rather difficult problem. Even comparison of only two distributions can be a difficult task. So, the following approach seems to be natural from a practical point of view (see, for example, Keeney and Raiffa [2], Harvey [3]). The distributions are evaluated by one or a number of criteria and the decision-making problem is reduced to one- or multi-criteria stochastic optimization problem. In probability theory several characteristics to describe and compare random quantities such as $f(x,\cdot)$ were elaborated: mean value, variance, probability of not exceeding of a given level, ... and others. An enormous number of papers were devoted to optimization of mean values (see for example, Ermoliev and Wets [10]). Much less works deal with probability function optimization. For instance, Raik [34, 35] established sufficient conditions for probability function to be (semi) continuous and hence conditions for probability function optimization problem to have a solution. Prekopa [30, 31], made a principal step in the theory of probability function when he discovered its logarithmic concavity for logarithmic concave measures. Borrell [1], Brascamp and Lieb [2], Rinnot [37], Das Gupta [6], Tamm [48], Roenko [38], Norkin and Roenko [25, 26] obtained another general results on quasi-concavity of probability functions. Raik [36], Roenko [38], Uryas'ev [50, 51], Simon [44], Roenko and Norkin [25, 26] studied differentiability properties of probability functions. Szantai [47] proposed an efficient method for estimating of values and gradients of probability function. Rõemisch and Schultz [42], Salinetti [43] studied stability properties of stochastic programming problems with probabilistic constraints. Numerical methods for optimization of probability functions were proposed in Prekopa [30, 31, 32, 33], Raik [36], Yubi [53, 54], Tamm [49], Szantai [47], Lepp [19, 20], Roenko [38], Uryas'ev [50], Norkin [24], Kankova [13], Kibzun and Malyshev [16] Kibzun and Kurbakovski [15], Kovalenko and Nakonechniy [18]. There are also a number of papers devoted to numerical solution of probabilistic constrained programming (for references see Prekopa [32, 33]).

In the present paper we consider the problem of nonsmooth probability function optimization. While most of the existing methods assume differentiable probability functions, we apply the stochastic quasi-gradient approach which can handle the nondifferentiable case, too. We also exploit a special property of some probability functions - their α-concavity.

2 Properties of probability functions

2.1 Notations

Let us consider the following function

$$F_0(x) := \mathbf{P}\{f(x,\omega) \geq c\} = \int_{\{\omega | f(x,\omega) \geq c\}} \mathbf{P}(d\omega),$$

where $f : X \times \Omega \to \mathbb{R}^1$ is some (loss) function, $X \subseteq \mathbb{R}^n$ is a range for control parameters, $\Omega \subseteq \mathbb{R}^m$ is a range for random parameters, $c \in \mathbb{R}^1$ is some given level, $(\Omega, \Sigma, \mathbf{P})$ is a probability space, $\mathbb{R}^i (i = 1, 2, \ldots)$ denotes i-dimensional arithmetic vector space. We consider two representation forms for the probability function.

Let us introduce a multi-valued mapping

$$H(x) := \{\omega \in \Omega | f(x,\omega) \geq c\}$$

with the domain

$$D := \operatorname{dom} H := \{x \in X | \exists \omega(x) : f(x, \omega(x)) \geq c\}.$$

Then one can represent

$$F_0(x) = \mathbf{P}\{H(x)\} = \int_{H(x)} \mathbf{P}(d\omega).$$

Thus the probability function $F_0(x)$ is defined on the set $D = \operatorname{dom} H$. If for any $x \in X$ one has

$$\sup\{f(x,\omega) | \omega \in \Omega\} > c$$

then $D = X$. If $f(x,\omega)$ is continuous in ω and Ω is a compact set in $\mathbb{R}^m$ then

$$D = X \cap \{x | \, \varphi(x) := \max_{\omega \in \Omega} \{f(x,\omega) - c\} \geq 0\}.$$

If in addition the function $f(x,\omega)$ is concave jointly in (x,ω) then the function $\varphi(x)$ is also concave.

Another form of probability function is obtained by means of the following indicator function

$$\chi(t) = \begin{cases} 1, & t \geq 0, \\ 0, & t < 0. \end{cases}$$

Consider the function $F(x) := \int_\Omega \chi(f(x,\omega) - c) \mathbf{P}(d\omega)$.

Obviously, we have

$$F(x) = \begin{cases} F_0(x), & x \in D, \\ 0, & x \neq D, x \in X. \end{cases}$$

2.2 Continuity

The conditions for the probability function to be semicontinuous or continuous were established by Raik [34].

Theorem 2.1 *(Raik[34, 35]). If the function $f(x,\omega)$ is upper semicontinuous in x at a point x' for almost all ω then $F_0(x)$ is also upper semicontinuous at x'. If the function $f(x,\omega)$ is continuous in x at a point x' for almost all ω and*

$$\mathbf{P}\{f(x',\omega) = c\} = 0$$

then $F(x)$ is continuous at x'.

Remark 2.1 *Suppose that the measure* $\mathbf{P}$ *has a density on a nondegenerated convex set* $\Omega \subseteq \mathbb{R}^m$. *If function* $f(x,\omega)$ *is concave in* ω *and* $\sup_{\omega\in\Omega} f(x,\omega) > c$ *then* $\mathbf{P}\{f(x,\omega) = c\} = 0$ *as the measure of the boundary of a nondegenerated convex set* $\{\omega \in \Omega | f(x,\omega) \geq c\}$ *in* $\mathbb{R}^m$. *The same takes place if* $f(x,\omega)$ *is strictly concave in* ω.

2.3 Quasi-concave and α-concave functions

Concavity and quasi-concavity are very important properties of functions in the theory of extremal problems. For probability functions similar properties are formulated by means of the notion of α-concavity of functions and measures (see Borrell [1], Das Gupta [5], Roenko [38], Norkin and Roenko [25, 26]).

Definition 2.1 *A function* $F(x)$ *defined on a convex set* $X \subset \mathbb{R}^n$ *is called quasi-concave if for any* x_0, $x_1 \in X$ *and* $\lambda \in (0,1)$ *the following inequality holds*

$$F((1-\lambda)x_0 + \lambda x_1) \geq \min(F(x_0), F(x_1)).$$

Definition 2.2 *A nonnegative function* $F(x)$ *defined on a convex set* $X \subset \mathbb{R}^n$ *is called logarithmic concave if for any* x_0,$x_1 \in X$ *and* $\lambda \in (0,1)$, *we have*

$$F((1-\lambda)x_0 + \lambda x_1) \geq F(x_0)^{1-\lambda} F(x_1)^{\lambda}.$$

Definition 2.3 *A nonnegative function* $F(x)$ *defined on a convex set* $X \subset \mathbb{R}^n$ *is called* α*-concave* (α *is a real number parameter,* $\alpha \in -[\infty, +\infty]$), *if for any* x_0, $x_1 \in X$ *and* $\lambda \in (0,1)$ *one has*

$$F((1-\lambda)x_0 + \lambda x_1) \geq \begin{cases} ((1-\lambda)F(x_0)^\alpha + \lambda F(x_1)^\alpha)^{1/\alpha}, & \alpha \neq 0, \pm\infty, \\ F(x_0)^{1-\lambda} F(x_1)^{\lambda}, & \alpha = 0, \\ \min(F(x_0), F(x_1)), & \alpha = -\infty, \\ \max(F(x_0), F(x_1)), & \alpha = +\infty. \end{cases}$$

Here the following conventions are accepted: $\ln 0 = -\infty$, $0 \cdot (\pm\infty) = 0$, $0^0 = 1$, $\infty^{-|\alpha|} = 0$, $0^{-|\alpha|} = +\infty$, $+\infty^0 = 1$.

Obviously, α-concave functions are quasi-concave and 0-concave functions are logarithmic concave.

Proposition 2.1 *From the definitions it follows that* $F(x)$ *is* α*-concave on* X ($-\infty < \alpha < +\infty$) *iff* $F^\alpha(x)$ *is concave, (if* $\alpha > 0$) $\ln F(x)$ *is concave (if* $\alpha = 0$) *and* $F^\alpha(x)$ *is convex on* X *(if* $\alpha < 0$).

For example, function $f_1(x) = \max(0,x), x \in \mathbb{R}^1$, is logarithmic concave, function $f_2(x) = |x|^{-1}$, $x \in \mathbb{R}^1$, is (-1)-concave. The indicator function of a convex set in $\mathbb{R}^n$ is logarithmic concave in $\mathbb{R}^n$.

Let us mention some properties of α-concave functions (see Roenko [38], Roenko and Norkin [25, 26]).

Lemma 2.1 *If a function* $F(x)$ *is* α_1*-concave and* $\alpha_1 \geq \alpha_2 \geq -\infty$ *then* $F(x)$ *is also* α_2*-concave.*

Lemma 2.2 *If* F_1 *is* α_1*-concave,* F_2 *is* α_2*-concave on* X *and either* $\alpha_1 > 0$, $\alpha_2 > 0$ *or* $\alpha_1\alpha_2 < 0$ *and* $\alpha_1^{-1} + \alpha_2^{-1} < 0$ *then* $F_1(x) \cdot F_2(x)$ *is* α_0*-concave function, where* $\alpha_0 = (\alpha_1^{-1} + \alpha_2^{-1})^{-1}$.

Lemma 2.3 *Let $F_i(x)$, $i = 1,\dots,m$, be nonnegative concave functions defined on a convex set $X \subset \mathbb{R}^n$. Then the function $F(x) = \prod_{i=1}^m F_i(x)$ is $(\frac{1}{m})$-concave on X.*

Let us now consider differentiability properties of α-concave functions.

Definition 2.4 *The quantity*

$$f'(x;l) = \lim_{t\to+0} \frac{1}{t}[f(x+tl) - f(x)]$$

is called a derivative of the function f at the point $x \in \mathbb{R}^n$ in the direction $l \in \mathbb{R}^n$. The quantity

$$f^0(x;l) = \lim_{\substack{y\to x\\ t\to+0}} \frac{1}{t}[f(y+tl) - f(y)]$$

is called Clarke's [4] generalized derivative of f at x in the direction l. The set

$$\partial f(x) = \{g \in \mathbb{R}^n \mid \langle g, l\rangle \le f^0(x;l) \quad \forall l \in \mathbb{R}^n\}$$

is called Clarke's [4] subdifferential of f at x.

Function f is called (+)regular (regular) if $f^0(x;l) = f'(x;l)$ and (-)regular if $(-f)^0(x;l) = (-f)'(x;l)$.

Convex functions are (+)regular and concave ones are (-)regular.

Lemma 2.4 *If some function $f(x)$, $x \in X \subseteq \mathbb{R}^n$, is α-concave in an open set X and $F(x) > 0$ in X then $F(x)$ is locally Lipschitzian, directionally differentiable and (-)regular. Its Clarke's subdifferential is defined by the formula:*

$$\partial F(x) = \begin{cases} \frac{1}{\alpha} F(x)^{1-\alpha} \partial F(x)^\alpha, & \alpha \ne 0, \\ F(x) \cdot \partial \ln F(x), & \alpha = 0. \end{cases}$$

Proof. Represent

$$F(x) = \begin{cases} (F(x)^\alpha)^{1/\alpha}, & \alpha \ne 0, \\ \exp(\ln F(x)), & \alpha = 0. \end{cases}$$

Functions F^α ($\alpha > 0$) and $\ln F$ ($\alpha = 0$) are finite and concave and function $F^\alpha(\alpha < 0)$ is finite and convex. Finite convex functions are Lipschitzian and (+)regular, finite concave functions are Lipschitzian and (-)regular (see Clarke [4]). Thus Lipschitzian function $F(x)$ is represented as a composition of a regular convex or concave function and a continuously differentiable function. (-)Regularity of F now follows from Clarke [4], Theorem 2.3.9. Subdifferential of a compound Lipschitzian function can be calculated by means of differentiation chain rule (see Clarke [4], Theorem 2.3.9) from which the required subdifferentiation formula follows.

2.4 Quasi-concave and α-concave measures

Not only functions but also measures have some convexity properties.

Definition 2.5 *Nonnegative measure* **P** *defined on σ-algebra of Lebesgue measurable subsets of a convex set $\Omega \subseteq \mathbb{R}^m$ is called α-concave if for arbitrary measurable sets A_0, $A_1 \subset \Omega$ and for any number $\lambda \in (0,1)$ one has*

$$\underline{\mathbf{P}}(A_\lambda) \geq \begin{cases} ((1-\lambda)P(A_0)^\alpha + \lambda P(A_1)^\alpha)^{1/\alpha}, & \alpha \neq 0, \\ P(A_0)^{1-\lambda}P(A_1)^\lambda, & \alpha = 0, \\ \min(P(A_0), P(A_1)), & \alpha = -\infty, \\ \max(P(A_0), (A_1)), & \alpha = +\infty, \end{cases}$$

where

$$A_\lambda = (1-\lambda)A_0 + \lambda A_1 = \{(1-\lambda)a_0 + \lambda a_1 \mid a_0 \in A_0, a_1 \in A_1\},$$

$\underline{\mathbf{P}}(A_\lambda)$ *is a lower measure of A_λ.*

The uniform Lebesgue measure on a nondegenerated convex set $\Omega \subseteq \mathbb{R}^m$ is $\frac{1}{m}$-concave due to Brunn-Minkowski-Lusternik inequality, (see [3, 22, 21]).

A connection between α-concave measures and functions is given in the following theorem (see Borrell [1], Brascamp and Lieb [2], Das Gupta [6], Prekopa [30]-[32] ($\alpha = 0$) and for references Norkin and Roenko [26]).

Theorem 2.2 *Let Ω be an open convex set in $\mathbb{R}^{m_0}$ and let* **P** *be a positive measure on Ω. Suppose L is the smallest affine subspace in $\mathbb{R}^{m_0}$ containing Ω and let m denote the dimension of L. Then the measure* **P** *is α-concave ($-\infty \leq \alpha \leq 1/m$) if its density function p with respect to Lebesgue measure on L is α'-concave on Ω, where $\alpha' = \alpha/(1 - m\alpha)^{-1}$ ($-1/m \leq \alpha' \leq +\infty$).*

Corollary 2.1 *Let an integrable nonnegative function $p(\omega)$ be defined on a nondegenerated convex set $\Omega \subseteq \mathbb{R}^m$. Suppose $p(\omega)$ is α-concave ($-1/m \leq \alpha \leq +\infty$) and positive on the interior of Ω. Then measure* **P** *on Ω defined by the formula*

$$\mathbf{P}(K) = \int_K p(\omega)d\omega, \quad K \subseteq \Omega,$$

is α'-concave on Ω, where $\alpha' = \alpha/(1 + \alpha m)$.

Corollary 2.2 *If a measure* **P** *on $\mathbb{R}^m$ has a density function f such that $f^{-1/m}$ is convex then* **P** *is quasi-concave.*

2.5 Examples of α-concave functions and measures

It turns out that many classical probability distributions have α-concave density functions and thus are generated by α'-concave measures on the appropriate sets (see, for example, Roenko [38], Roenko and Norkin [25, 26]).

Example 2.1 *Consider the density function of nondegenerated multivariate normal distribution in $\mathbb{R}^n$:*

$$p_1(x) = \frac{1}{\sqrt{(2\pi)^n \det B}} \exp(-\frac{1}{2}(x-m)^T B^{-1}(x-m)),$$

where B^{-1} is a positive definite $n \times n$-matrix, m is n-dimensional vector. Since the function $\ln p_1(x)$ is concave then density $p_1(x)$ generates a logarithmic concave measure $\mathbf{P}_1$ on $\mathbb{R}^n$ (Prekopa [30]).

Example 2.2 *Consider the uniform distribution on a convex set $G \subset \mathbb{R}^n$ with density*

$$p_2(X) = \begin{cases} 1/V(G), & x \in G, \\ 0, & x \notin G, \end{cases}$$

where $V(G)$ denotes the Lebesgue measure of G. The function $p_2(x)$ is $(+\infty)$-concave, hence by Corollary 2.1 it generates a $\frac{1}{n}$-concave measure $\mathbf{P}_2$ on G.

Example 2.3 *Consider the density function of the multivariate β-distribution (Dirichlet's distribution) with parameters*

$$\{\alpha_1, \ldots, \alpha_n\} = \alpha,\ \alpha_i > 0,\ i = 1, \ldots, n :$$

$$p_3(x) = \begin{cases} \frac{\Gamma(\alpha_1+\alpha_2+\ldots+\alpha_n)}{\Gamma(\alpha_1)\ldots\Gamma(\alpha_n)} & x_1^{\alpha_1} \ldots x_{n-1}^{\alpha_{n-1}}(1 - x_1 - \ldots - x_{n-1})^{\alpha_n},\ \textit{if} \\ & x_1 \geq 0, \ldots, x_{n-1} \geq 0, 1 - x_1 - \ldots - x_{n-1} \geq 0, \\ 0 & \textit{otherwise} \end{cases}$$

where $\Gamma(\cdot)$ is the gamma-function. By Lemma 2.2 the function $p_3(x)$ is $(\alpha_1 + \ldots + \alpha_n)^{-1}$-concave on an open $(n-1)$-dimensional simplex $\{x \in \mathbb{R}^{n-1} | \sum_{i=1}^{n-1} x_i < 1,\ x_i > 0,\ i = 1, \ldots, n-1\}$; hence by Corollary 2.1 the corresponding measure $\mathbf{P}_3$ is $(\alpha_1 + \ldots + \alpha_n + n - 1)^{-1}$-concave on the set $\{x \in \mathbb{R}^{n-1} | \sum_{i=1}^{n-1} x_i \leq 1,\ x_i \geq 0, i = 1, \ldots, n-1\}$.

Example 2.4 . *Consider the density function of the l-dimensional Student's distribution with number parameter n, vector parameter m and matrix parameter T*

$$p_4(x) = \frac{\Gamma(\frac{l+n}{2})\sqrt{\det T}}{\Gamma(\frac{n}{2})\sqrt{(2\pi)^l}}[1 + \frac{1}{2}(x - m)^T T(x - m)]^{-(l+n)/2},$$

where T is a symmetric positive definite matrix. Function $p_4(x)$ is $\left(-\frac{2}{l+n}\right)$-concave, hence the corresponding measure $\mathbf{P}_4$ is $\left(-\frac{2}{n-l}\right)$-concave on $\mathbb{R}^l$.

Example 2.5 . *The multidimensional Pareto's distribution has the following density function with parameters $\alpha, \theta_1, \ldots, \theta_n > 0$:*

$$p_5(x) = \text{const}\,(\sum_{i=1}^{n} x_i/\theta_i - n + 1)^{-(\alpha+n)}, x_i \geq \theta_i, i = 1, \ldots n.$$

The function $p_5(x)$ is $\left(-\frac{1}{\alpha+n}\right)$-concave and hence the corresponding measure $\mathbf{P}_5$ is α^{-1}-concave on the set $\{x \in \mathbb{R}^n | x_i \geq \theta_i, i = 1, \ldots n\}$.

Example 2.6 . *The density function of the multidimensional F-distribution with parameters $n_0, n_1, \ldots, n_l$, $n = \sum_{i=1}^{l} n_i$ looks as follows:*

$$p_6(x) = \text{const}\left[\prod_{i=1}^{l} x_i^{n_i/2-1}\right]\left[n_0 + \sum_{i=1}^{l} n_i x_i\right]^{-n/2}, x_i \geq 0, i = 1, \ldots, l.$$

The function $\prod_{i=1}^{l} x_i^{n_i/2-1}$ is by Lemma 2.2 $\left(\frac{1}{2}\sum_{i=1}^{l} n_i - l\right)^{-1}$-concave and the function $\left[n_0 + \sum_{i=1}^{l} n_i x_i\right]^{-n/2}$ is $(-\frac{2}{n})$-concave. By Lemma 2.2 the density function $p_6(x)$ is $[-(n_0/2 + l)^{-1}]$-concave on the set $\{x \in \mathbb{R}^l \mid x_1 > 0, i = 1, \ldots l\}$ and the corresponding measure $\mathbf{P}_6$ is $(-\frac{2}{n})$-concave.

2.6 Quasi-concavity and α-concavity of probability functions

Now we shall formulate sufficient conditions for a probability function to be α-concave.

Theorem 2.3 *Let measure* $\mathbf{P}$ *be* α*-concave on a convex set* $\Omega \subseteq \mathbb{R}^m$*, function* $f : X \times \Omega \to \mathbb{R}^1$ *be quasi-concave on* $X \times \Omega$ *(see examples 1.1 -1.4), and let* X *be a convex set in* $\mathbb{R}^n$*. Then the probability function*

$$F(x) = \mathbf{P}\{f(x,\omega) \geq c\}$$

is α*-concave on the set*

$$D = \{x \in X \mid \exists \omega(x) \in \Omega : f(x, \omega(x)) \geq c\}.$$

Proof. Observe that the multivalued mapping

$$H(x) = \{\omega \in \Omega \mid f(x,\omega) \geq c\}$$

is convex on D, i.e. for arbitrary $x_0, x_1 \in D$ and $\lambda \in (0,1)$ we have

$$(1-\lambda)H(x_0) + \lambda H(x_1) \subseteq H((1-\lambda)x_0 + \lambda x_1).$$

Indeed, if $\omega_0 \in H(x_0)$, $\omega_1 \in H(x_1)$, $x_\lambda = (1-\lambda)x_0 + \lambda x_1$ and $\omega_\lambda = (1-\lambda)\omega_0 + \lambda\omega_1$, then $f(x_0,\omega_0) \geq c$, $f(x_1,\omega_1) \geq c$ and

$$f(x_\lambda, \omega_\lambda) \geq \min\left(f(x_0,\omega_0), f(x_1,\omega_1)\right) \geq c,$$

and thus $\omega_\lambda \in H(x_\lambda)$.

Now let $\alpha \neq 0, \pm\infty$, $x_0 \in D_0$, $x_1 \in D_0$, $x_\lambda = (1-\lambda)x_0 + \lambda x_1$ and $\lambda \in (0,1)$. Then, due to α-concavity of measure $\mathbf{P}$,

$$\begin{aligned} F(x_\lambda) = \mathbf{P}\{H(x_\lambda)\} &\geq \mathbf{P}\{(1-\lambda)H(x_0) + \lambda H(x_1)\} \geq \\ &\geq [(1-\lambda)\mathbf{P}\{H(x_0)\}^\alpha + \lambda P\{H(x_1)\}^\alpha]^{1/\alpha} \\ &\geq [(1-\lambda)F(x_0)^\alpha + \lambda F(x_1)^\alpha]^{1/\alpha}, \end{aligned}$$

which was required to prove. The proof for $\alpha = 0, \pm\infty$ is similar.

Similar statements were proved in Prekopa [30, 31] ($\alpha = 0$), Borrell [1], Brascamp and Lieb [2], Das Gupta[6], Wets [52] ($\alpha = -\infty$), Roenko [38], Norkin [24], Norkin and Roenko [25, 26].

Corollary 2.3 . *Under conditions of Theorem 2.3 the probability function* $F(x)$ *is continuous, differentiable in directions and Lipschitzian on the set* $\{x \in D \mid F(x) > 0\}$.

3 Approximation of probability functions

3.1 Smoothing of probability functions

In general, a probability function $F(x)$ can be discontinuous since its representation in the mathematical expectation form contains a discontinuous function $\chi(\cdot)$. For the same reason one cannot differentiate $F(x)$ by interchanging differentiation and integration operators in the expression defining $F(x)$. So, we replace the discontinuous function $\chi(\cdot)$ by some continous approximate function $\tilde{\chi}(\cdot)$ and in this way we obtain a continuous approximation for $F(x)$. For discontinuous functions Steklov-Sobolev's [46, 45] average approximations are very convenient (see Ermoliev, Norkin and Wets [8]).

Let $k(\tau), -\infty < \tau < +\infty$, be a nonnegative integrable function such that

$$\int_{-\infty}^{+\infty} k(\tau)d\tau = 1.$$

For convenience we shall assume that $k(\tau)$ is symmetric, i.e. $k(\tau) = k(-\tau)$. The density function $k(\tau)$ generates a measure $\mathbf{K}$ on $\mathbb{R}^1$ by the formula

$$\mathbf{K}(A) := \int_A k(\tau)d\tau, \quad A \in \mathbb{R}^1.$$

Denote its characteristic function by

$$\tilde{\chi}(t) := \int_{-\infty}^{t} k(\tau)d\tau = \int_{-\infty}^{t} \mathbf{K}(d\tau).$$

For the function

$$\chi(t) = \begin{cases} 1, & t \geq 0. \\ 0, & t < 0, \end{cases}$$

we consider Steklov-Sobolev's average functions

$$\tilde{\chi}_\varepsilon(t) := \int_{-\infty}^{+\infty} \chi(t+\varepsilon\tau)k(\tau)d\tau = \frac{1}{\varepsilon}\int_{-\infty}^{+\infty} \chi(\tau)k(\frac{\tau - t}{\varepsilon})d\tau, \quad \varepsilon > 0.$$

The following representation is true

$$\tilde{\chi}_\varepsilon(t) = \int_{-t/\varepsilon}^{+\infty} k(\tau)d\tau = \int_{-\infty}^{t/\varepsilon} k(-\tau)d\tau = \int_{-\infty}^{t/\varepsilon} k(\tau)d\tau = \tilde{\chi}(t/\varepsilon).$$

Now consider the following approximation for probability functions

$$F_\varepsilon(x) := \int_\Omega \tilde{\chi}_\varepsilon(f(x,\omega) - c)\mathbf{P}(d\omega) = \int_\Omega \tilde{\chi}((f(x,\omega) - c)/\varepsilon)\mathbf{P}(d\omega) =$$

$$= \frac{1}{\varepsilon}\int_\Omega \int_{-\infty}^{-c} k((\tau + f(x,\omega))/\varepsilon)d\tau.$$

Remark 3.1 *If $\omega_i, i = 1, 2, \ldots, n$, are i.i.d. observations of a random variable ω then $F_\varepsilon(x)$ can be approximated by its empirical estimate*

$$F_\varepsilon^n(x) := \frac{1}{\varepsilon n}\sum_{i=1}^{n}\int_{-\infty}^{-c} k((\tau + f(x,\omega_i))/\varepsilon)d\tau.$$

Such estimates for probability function $F(x)$ were constructed by Tamm [49] and Lepp [19]. They are similar to Parzen-Rosenblatt [28, 41] estimates for probability density.

3.2 Convergence of approximations

Let us study properties of approximations $F_\varepsilon(x)$ to probability function $F(x)$. The next lemma establishes conditions of point-wise convergence of $F_\varepsilon(x)$ to $F(x)$.

Lemma 3.1 *If for a given x*

$$\mathbf{P}\{f(x,\omega) = c\} = 0$$

then

$$\lim_{\varepsilon \to 0} F_\varepsilon(x) = F(x).$$

Proof. For a given x we have

$$\mathbf{P}\{\tilde{\chi}_\varepsilon(f(x,\omega)-c) \to \chi(f(x,\omega)-c)\} =$$

$$= 1 - \mathbf{P}\{\tilde{\chi}_\varepsilon(f(x,\omega)-c) \not\to \chi(f(x,\omega)-c)\} \leq$$

$$\leq 1 - \mathbf{P}\{f(x,\omega)=c\} = 1.$$

Besides $0 \leq \tilde{\chi}_\varepsilon(f(x,\omega)-c) \leq 1$, $0 \leq \chi(f(x,\omega)-c) \leq 1$. Therefore

$$\lim_{\varepsilon\to 0}(F(x) - F_\varepsilon(x)) = \int_\Omega \lim_{\varepsilon\to 0}[\chi(f(x,\omega)-c) - \tilde{\chi}_\varepsilon(f(x,\omega)-c)]\mathbf{P}(d\omega) = 0.$$

Now consider conditions under which approximations $F_\varepsilon(x)$ uniformly converge to $F(x)$.

Lemma 3.2 *(Norkin[24]). Let function $f(x,\omega)$ be continuous in x for almost all ω. Suppose that for any x and for all c' sufficiently close to c, one has either*

$$\{\omega \in \Omega \mid f(x,\omega) = c'\} = \emptyset$$

or

$$\mathbf{P}\{f(x,\omega) = c'\} = 0.$$

Then for any compact $X \subseteq D_0$ the functions $F_\varepsilon(x)$ uniformly converge in X to $F(x)$ under $\varepsilon \to 0$.

Proof. Let $\delta(\varepsilon) = \varepsilon^{1-\nu}$, where $0 < \nu < 1$. If $\varepsilon \to 0$ then $\delta(\varepsilon) \to 0$, $\chi(\delta(\varepsilon)/\varepsilon) \to 1$ and $\chi(-\delta(\varepsilon)/\varepsilon) \to 0$. We have the following estimate

$$\begin{aligned}
|F(x) - F_\varepsilon(x)| &\leq \int_\Omega |\chi(f(x,\omega)-c) - \tilde{\chi}_\varepsilon(f(x,\omega)-c)|\mathbf{P}(d\omega) \\
&\leq \chi(-\delta(\varepsilon)/\varepsilon)\int_{\{\omega | f(x,\omega)-c \leq -\delta(\varepsilon)\}} \mathbf{P}(d\omega) \\
&+ (1-\chi(\delta(\varepsilon)/\varepsilon)\int_{\{\omega | f(x,\omega)-c \geq \delta(\varepsilon)\}} \mathbf{P}(d\omega) \\
&+ \int_{\{\omega || f(x,\omega)-c| \leq -\delta(\varepsilon)\}} \mathbf{P}(d\omega) \\
&\leq \chi(-\delta(\varepsilon)/\varepsilon) + (1-\chi(\delta(\varepsilon)/\varepsilon) + \mathbf{P}\{|f(x,\omega)-c| \leq \delta(\varepsilon)\}.
\end{aligned}$$

It is sufficient to show that $\varphi_\delta(x) = \mathbf{P}\{|f(x,\omega)-c| \leq \delta\} \to 0$ uniformly in x in each compact $X \subseteq D$ if $\delta \to 0$. The function

$$\psi(x,\omega,\delta) = |f(x,\omega)-c| - \delta$$

is continuous in (x,δ) and measurable in ω. For sufficiently small $\delta \leq \delta_0$ one has

$$\mathbf{P}\{|f(x,\omega)-c| - \delta = 0\} = \mathbf{P}\{|f(x,\omega)-c| = \delta\} \leq$$

$$\leq \mathbf{P}\{f(x,\omega) = c+\delta\} + \mathbf{P}\{f(x,\omega) = c-\delta\} = 0,$$

where the convention $\mathbf{P}(\emptyset) = 0$ is accepted. By Theorem 2.1 function $\varphi_\delta(x)$ is continuous and hence uniformly continuous in $(x,\delta) \in X \times [-\delta_0, \delta_0]$. So, for an arbitrary σ there exists $\gamma(\sigma)$ such that $|\varphi_\delta(x)| = |\varphi_\delta(x) - \varphi_0(x)| < \sigma$ if $|\delta| < \gamma$. It means that uniformly in $x \in X$ $\mathbf{P}\{|f(x,\omega)-c| \leq \delta(\varepsilon)\} \to 0$ if $\varepsilon \to 0$, which was set out to prove.

3.3 α-Concavity of approximations

Under some conditions approximate functions $F_\varepsilon(x)$ not only uniformly converge to $F(x)$ but are α-concave with some α.

Lemma 3.3 *The following representation for $F_\varepsilon(x)$ is true*

$$F_\varepsilon(x) = \int_{H_\varepsilon(x)} \mathbf{K}(d\tau)\mathbf{P}(d\omega),$$

where

$$H_\varepsilon(x) = \{(\omega, \tau) \in \Omega \times \mathbb{R}^1 \mid f(x, \omega) - \varepsilon\tau \geq c\} \subseteq \mathbb{R}^{m+1}.$$

Proof. Indeed, we have

$$F_\varepsilon(x) = \int_\Omega \tilde{\chi}((f(x,\omega) - c)/\varepsilon)\mathbf{P}(d\omega) = \int_\Omega \int_{-\infty}^{(f(x,\omega)-c)/\varepsilon} k(\tau)d\tau\mathbf{P}(d\omega) =$$

$$= \int_\Omega \int_{-\infty}^{+\infty} \chi((f(x,\omega) - c)/\varepsilon - \tau)\mathbf{K}(d\tau)\mathbf{P}(d\omega) = \int_{H_\varepsilon(x)} \mathbf{K}(d\tau)\mathbf{P}(d\omega).$$

The following statement is obvious.

Lemma 3.4 *If the function $f(x,\omega)$ is jointly concave in (x, ω) (see examples 1.1 - 1.4) then the function $f(x,\omega) - \varepsilon\tau$ is jointly concave in (x, ω, τ) and hence the multivalued mapping $H_\varepsilon(x)$ is convex.*

Under the conditions of Lemma 3.3 the degree of concavity of $F_\varepsilon(x)$ is defined according to Theorem 2.3 by the degree of concavity of the product of measures **K** and **P**. In the following three lemmas this degree of concavity of the product of measures **K** and **P** is calculated for a number of particular cases.

Lemma 3.5 *If $k(\tau)$, $\tau \in \mathbb{R}^1$, and $p(\omega)$, $\omega \in \Omega \subseteq \mathbb{R}^m$, are logarithmic concave functions, i.e.* **K** *and* **P** *are logarithmic concave measures, then $k(\tau)p(\omega)$ is also a logarithmic concave density function and by Theorem 2.2 the corresponding measure is logarithmic concave.*

Now define a function

$$\underline{F_\varepsilon}(x) := \int_{H_\varepsilon(x)\cap(G\times[a,b])} \mathbf{K}(d\tau)\mathbf{P}(d\omega),$$

where $G \subseteq \Omega$, $-\infty \leq a < 0 < b \leq +\infty$. This function is a lower estimate for $F_\varepsilon(x)$:

$$0 \leq F_\varepsilon(x) - \underline{F_\varepsilon}(x) \leq 1 - \mathbf{P}(G) \cdot \mathbf{K}([a, b]).$$

Function $\underline{F_\varepsilon}(x)$ is defined on the set

$$\underline{D_\varepsilon} := \{x \in X \mid \exists\omega(x) \in G : f(x, \omega(x)) \geq c + \varepsilon a\}.$$

If $G = \Omega$ and $\mathbf{K}([a, b]) = 1$ then $F_\varepsilon(x) = \underline{F_\varepsilon}(x)$ on $D \subseteq \underline{D_\varepsilon}$.

Lemma 3.6 *If a function $p(\omega)$ is α-concave ($\alpha > 0$) on the interior of a convex set $G \subseteq \Omega \subseteq \mathbb{R}^m$ and a function $k(\tau)$ is β-concave ($\beta > 0$) on interval $(a, b) \subseteq \mathbb{R}^1$ then by Lemma 2.2 the function $k(\tau)p(\omega)$ is γ-concave ($\gamma = (\alpha^{-1} + \beta^{-1})^{-1}$) on the set* int $G \times (a, b)$ *and by Theorem 2.2 and Corollary 2.1 the corresponding measure is α'-concave ($\alpha' = \gamma(1 + \gamma m)^{-1}$) on $G \times [a, b]$.*

Lemma 3.7 *If a function $k(\tau)$ is constant on (a, b) and a function $p(\omega)$ is constant on G then the density function $k(\tau)p(\omega)$ is also constant on* int$G \times (a, b)$ *and by Theorem 2.2 and Corollary 2.1 the corresponding measure is $(1/(m + 1))$-concave on the set $G \times [a, b]$.*

3.4 Differentiability properties of approximations

Let us at first consider differentiability properties of the approximate functions

$$F_\varepsilon(x) = \int_\Omega \tilde{\chi}((f(x,\omega) - c)/\varepsilon)\mathbf{P}(d\omega),$$

where

$$\tilde{\chi}(t) = \int_{-\infty}^{t} k(\tau)d\tau.$$

Theorem 3.1 *Let a density function $k(\tau)$ be nonnegative, bounded and continuous, let function $f(x,\omega)$ be concave in $x \in X$ for every $\omega \in \Omega$ and let*

$$\sup\{\| g \| \mid g \in \partial f(y,\omega), \| y - x \| \leq \delta\} \leq L_\delta(x,\omega), \quad \delta > 0,$$

where $L_\delta(x,\omega)$ is integrable in ω for each $x \in X$. Then $F_\varepsilon(x)$ is a Lipschitzian $(-)$regular function and its subdifferential is given by the formula

$$\partial F_\varepsilon(x) = \frac{1}{\varepsilon}\int_\Omega k((f(x,\omega) - c)/\varepsilon) \cdot \partial f(x,\omega)\mathbf{P}(d\omega).$$

Proof. Under the theorem's conditions the function $\tilde{\chi}(t)$ is monotone and continuously differentiable. A concave function $f(x,\omega)$ is $(-)$regular (see Clarke [4], Theorem 2.3.6), so a compound function $\tilde{\chi}((f(x,\omega) - c)/\varepsilon)$ by Clarke [4], Theorem 2.3.9, is $(-)$regular and

$$\partial\tilde{\chi}((f(x,\omega) - c)/\varepsilon) = \frac{1}{\varepsilon}k((f(x,\omega) - c)/\varepsilon) \cdot \partial f(x,\omega).$$

Now by Clarke [4], Theorem 2.7.2, the mathematical expectation function $F_\varepsilon(x)$ is Lipschitzian $(-)$regular and

$$\partial F_\varepsilon(x) = \int_\Omega \partial\tilde{\chi}((f(x,\omega) - c)/\varepsilon)\mathbf{P}(d\omega),$$

what is required to be proved.

Corollary 3.1 *Suppose in addition to conditions of Theorem 3.1 that function $F_\varepsilon(x)$ is α-concave and*

$$\sup_{\omega\in\Omega} f(x,\omega) > c.$$

Then probability function $F(x)$ is also α-concave, the sequence of functions $F_\varepsilon(x)$, $\varepsilon \to +0$, uniformly converges to $F(x)$ on a compact X and the subdifferentials $\partial F_\varepsilon(x)$ converge to subdifferentials $\partial F(x)$ in the following sense: for any $\varepsilon \to +0$, $x^\varepsilon \to x$ and $g^\varepsilon \in \partial F_\varepsilon(x^\varepsilon)$, $g^\varepsilon \to g$, we have $g \in \partial F(x)$.

4 Approximate optimization of the probability function

4.1 Problem formulation

Consider a stochastic optimization problem of the following form:

$$F_0(x) = \mathbf{P}\{f(x,\omega) \geq c\} \to \max_{x\in X},$$

where the function $f(x,\omega)$ is concave in the pair of variables (x,ω) (see examples 1.1 - 1.4) X is a convex compact in $\mathbb{R}^n$, Ω is a nondegenerated convex set in $\mathbb{R}^m$, $c \in \mathbb{R}^1$, $(\Omega, \sum, \mathbf{P})$ is some probability space. Suppose that the measure $\mathbf{P}$ has a positive density function on the interior of Ω and the function $f(x,\omega)$ satisfies global Lipschitz condition in $x \in X$ with an integrable Lipschitz constant $L(\omega)$. The above problem includes the following implicit constraint

$$x \in \{x \mid \exists \omega(x) \in \Omega : f(x,\omega(x)) \geq c\}.$$

We represent this probability optimization problem in the mathematical expectation form:

$$F(x) = \int_\Omega \chi(f(x,\omega) - c)\mathbf{P}(d\omega) \to \max_{x \in X},$$

where

$$\chi(t) = \begin{cases} 1, & t \geq 0 \\ 0, & t < 0. \end{cases}$$

Function $F(x)$ is defined on X and

$$F(x) = \begin{cases} F_0(x), & x \in D \subseteq X, \\ 0, & x \notin D, x \in X, \end{cases}$$

$$D = \{x \in X \mid \exists \omega(x) \in \Omega : f(x,\omega(x)) \geq c\}.$$

Let F_0^* be the optimal value and X_0^* be the optimal set of this problem.

Define the function

$$\tilde{\chi}(t) = \int_{-\infty}^{t} k(\tau)d\tau,$$

where $k(\tau)$ is a continuous bounded nonnegative function such that $k(\tau) = 0$ for $\tau \leq a < 0$, $k(\tau) > 0$ if a $< \tau < b$ and $\tilde{\chi}(b) = 1, -\infty \leq a < 0 < b \leq +\infty$.

Now consider a family of approximate problems ($\varepsilon > 0$):

$$F_\varepsilon(x) := \int_\Omega \tilde{\chi}((f(x,\omega) - c)/\varepsilon)\mathbf{P}(d\omega) \to \max_{x \in D}.$$

Let F_ε^* be the optimal value and X_ε^* be the optimal set of this problem.

Under our assumptions, the function $F_\varepsilon(x)$ is by Theorem 3.1 Lipschitzian $(-)$regular and its subdifferential is given by the formula

$$\partial F_\varepsilon(x) = \frac{1}{\varepsilon}\int_\Omega k((f(x,\omega) - c)/\varepsilon) \cdot \partial f(x,\omega)\mathbf{P}(d\omega).$$

Theorem 4.1 *If in addition to the above assumptions*

$$\sup_{\omega \in \Omega} f(x,\omega) > c,$$

then $D = X$, *functions* $F_\varepsilon(x)$, $\varepsilon > 0$, *are continuous and positive on* X, *sequence* $\{F_\varepsilon(x),\ \varepsilon \to +0\}$ *uniformly converges to a continuous function* $F_0(x), F_\varepsilon^* \to F_0^*$ *and if* $x_\varepsilon^* \in X_\varepsilon^*, x_\varepsilon^* \to x^*$ *then* $x^* \in X_0^*$.

Proof. Obviously, $D = X$. For each $x \in X$ the set

$$H(x) = \{\omega \in \Omega | f(x,\omega) \geq c\}$$

has a nonempty interior, the function $\tilde{\chi}((f(x,\omega) - c)/\varepsilon)$ is greater than zero on this interior, so $F_\varepsilon(x) > 0$ on X. For c' sufficiently close to c, one must have

$$\mathbf{P}\{f(x,\omega) = c'\} = 0, \quad x \in X,$$

since the Lebesgue measure of the boundary of a convex set

$$\{\omega \in \Omega \mid f(x,\omega) \geq c'\} \subseteq \mathbb{R}^m, \quad x \in X,$$

is equal to zero. By Lemma 3.2 it follows that functions $F_\varepsilon(x), \varepsilon \to +0$, are uniformly convergent to $F(x)$ on X. This implies the required convergence of optimal values F_ε^* and sets X_ε^* to F_0^* and X_0^*.

Thus the original (possibly nonsmooth) probability maximization problem is reduced to solving a sequence of mathematical expectation maximization problems. Under our assumptions the functions $F_\varepsilon(x)$ are by Theorem 3.1 (−)regular. It follows (see Rockafellar [40], Kiriluk [17]) that they are weakly concave (see Nurminski [27]). So, for their maximization stochastic quasi-gradient methods are applicable (see Ermoliev [7], Nurminski [27]).

Another approach to optimization of function $F_\varepsilon(x)$ consists in replacing $F_\varepsilon(x)$ by its empirical estimates. Such approach to optimization of probabilities was first applied by Tamm [49] and Lepp [19].

And finally, an approximation of $F(x)$ by $F_\varepsilon(x)$ with $\varepsilon \to 0$ can be combined with optimization of $F_\varepsilon(x)$. Such approach combines ideas of stochastic quasi-gradient methods and nonstationary optimization and was developed in Ermoliev and Nurminski [9], Gaivoronski [11]. It was applied to optimization of probabilities by Lepp [20].

In the present paper we shall use the circumstance that functions $F(x)$ and $F_\varepsilon(x)$ can be α-concave (see Theorem 2.3 and Lemmas 3.3 - 3.7). For stochastic maximization of concave functions there exists an efficient method by Nemirovski and Yudin [23], which is a modification of stochastic quasi-gradient method by Ermoliev [7]. We are going to apply this method to stochastic optimization of α-concave (probability) functions.

4.2 Stochastic quasi-gradient method

Now suppose that $F_\varepsilon(x), \varepsilon \geq 0$, is an α-concave function (see Theorem 2.3 and Lemmas 3.3 - 3.7) and D is a convex set. We construct the following sequence of approximations ($x^0 = \bar{x}^0 \in D$):

$$x^{k+1} = \prod_D (x^k + \rho_k \cdot \xi^k),$$

$$\bar{x}^{k+1} = (1 - \sigma_{k+1})\bar{x}^k + \sigma_{k+1} x^{k+1}, \quad k = 0, 1, \ldots,$$

where stochastic quasi-gradient ξ^k of $F_\varepsilon(x)$ at point x^k is such that

$$\mathbf{E}\{\xi^k / x^0, \ldots, x^k\} \in \partial F_\varepsilon(x^k),$$

$$\mathbf{E}\{\| \xi^k \|^2 / x^0, \ldots, x^k\} \leq C(\varepsilon) < +\infty.$$

Next, $\prod_D$ is a projection operator on the set D, $\rho_k, k = 0, 1, \ldots$, are some positive numbers and

$$\sigma_k = \rho_k / \sum_{i=0}^{k} \rho_i.$$

In some cases the probability function $F_0(x)$ is differentiable and its stochastic quasi-gradient $\xi(x)$ can be constructed directly (see for example, Raik [35], Uryas'ev [50, 51], Szantai [47]).

In a nondifferentiable case one can take stochastic quasi-gradients of $F_\varepsilon(x), \varepsilon > 0$, at x^k in the form (see Theorem 3.1).

$$\xi^k = \xi(x^k, \omega^k),$$

where

$$\xi(x, \omega) = \frac{1}{\varepsilon} k((f(x, \omega) - c)/\varepsilon) \cdot g(x, \omega),$$
$$g(x, \omega) \in \partial_x f(x, \omega);$$

$g(x, \omega)$ is a measurable in (x, ω) selection of the multi-valued mapping $\partial_x f(x, \omega)$, and $\omega^k, k = 0, 1, \ldots$, are i.i.d. observations of problem random parameters ω.

If $f(x, \omega)$ is Lipschitzian in x with a square-integrable constant $L(\omega)$, the function $k(\tau)$ is continuous and bounded by constant K and $L^2 = \int_\Omega L^2(\omega) \mathbf{P}(d\omega)$, then

$$\mathbf{E}\{\| \xi^k \|^2 / x^0, \ldots, x^k\} \leq K^2 L^2 / \varepsilon^2 = C(\varepsilon) < +\infty.$$

4.3 Convergence results

Theorem 4.2 *Suppose that $F_\varepsilon(x)$ is a $(-)$regular (see Theorem 3.1) function, D is a closed convex set and*

$$\lim_{i \to +\infty} \rho_i = 0, \quad \sum_{i=0}^{\infty} \rho_i = +\infty, \quad \sum_{i=0}^{\infty} \rho_1^2 < +\infty.$$

Then for almost all trajectories $\{x^k\}$ generated by the above stochastic quasi-gradient method the following properties hold:

(i) all accumulation points of sequences $\{x^k\}$, $\{\bar{x}^k\}$ belong to the set

$$\arg \min_{x \in D} F_\varepsilon(x);$$

(ii) $\lim_{k \to \infty} F_\varepsilon(x^k) = \lim_{k \to \infty} F_\varepsilon(\bar{x}^k) = \min_{x \in D} F_\varepsilon(x)$.

Proof. It follows from general results by Rockafellar [40] that $(-)$regular functions are weakly concave by Nurminski [27] (for direct proof see Kiriluk [17]). So theorem's statements follow from general convergence results for stochastic quasi-gradient algorithm when applied to a weakly concave function (see Nurminski [27] (i), Ermoliev [7] (ii)). The statements for $\{\bar{x}^k\}$ easily follow from the ones for $\{x^k\}$.

Remark 4.1 *In the case of a compact set Ω and a concave $f(x, \omega)$, the implicit convex constraint*

$$\varphi(x) = \max\{f(x, \omega) - c | \omega \in \Omega\} \geq 0$$

can be taken into consideration not only through projection operation on D but directly in gradient procedure in the following way (see Polyak [29]):

$$x^{k+1} = \prod_X (x^k + \rho_k \cdot \eta^k),$$

$$\eta^k = \begin{cases} \xi^k, & \varphi(x^k) < 0, \\ g_\varphi(x^k), & \varphi(x^k) \geq 0, \end{cases}$$

where ξ^k is stochastic quasi-gradient of F_ε and $g_\varphi(x^k)$ is some generalized gradient of φ at x^k.

4.4 Rate of convergence

Lemma 4.1 *Suppose that the function $F_\varepsilon(x)$ is α-concave and*

$$F_\varepsilon(x) \geq \delta > 0, \quad x \in X.$$

Then the following estimates are true:

$$\begin{array}{ll}
\mathbf{E}F_\varepsilon^\alpha(\bar{x}^k) - F_\varepsilon^\alpha(x^*) \leq \frac{-\alpha \mathbf{E}\|x^0-x^*\|^2}{2\delta^{1-\alpha}\sum_{i=0}^k \rho_i} + \frac{-\alpha C(\varepsilon)\sum_{i=0}^k \rho_i^2}{2\delta^{1-\alpha}\sum_{i=0}^k \rho_i}, & \alpha < 0; \\
\ln F_\varepsilon(x^*) - \mathbf{E}\ln F_\varepsilon(\bar{x}^k) \leq \frac{\mathbf{E}\|x^0-x^*\|^2}{2\delta\sum_{i=0}^k \rho_i} + \frac{C(\varepsilon)\sum_{i=0}^k \rho_i^2}{2\delta\sum_{i=0}^k \rho_i}, & \alpha = 0; \\
F_\varepsilon^\alpha(x^*) - \mathbf{E}F_\varepsilon^\alpha(\bar{x}^k) \leq \frac{\alpha \mathbf{E}\|x^0-x^*\|^2}{2\delta^{1-\alpha}\sum_{i=0}^k \rho_i} + \frac{\alpha C(\varepsilon)\sum_{i=0}^k \rho_i^2}{2\delta^{1-\alpha}\sum_{i=0}^k \rho_i}, & 0 < \alpha < 1; \\
F_\varepsilon^\alpha(x^*) - \mathbf{E}F_\varepsilon^\alpha(\bar{x}^k) \leq \frac{\alpha \mathbf{E}\|x^0-x^*\|^2}{2\sum_{i=0}^k \rho_i} + \frac{\alpha C(\varepsilon)\sum_{i=0}^k \rho_i^2}{2\sum_{i=0}^k \rho_i}, & \alpha \geq 1;
\end{array}$$

where $\mathbf{E}$ *denotes the mathematical expectation operator over all trajectories* $\{\bar{x}^k\}$, *and the point* x^* *is such that*

$$F_\varepsilon(x^*) = \max_{x \in X} F_\varepsilon(x).$$

Proof. Consider the case of $0 < \alpha \leq 1$. For $\alpha \leq 0$ and $\alpha > 1$ the proof will be similar. The function $F_\varepsilon^\alpha(x)$ is Lipschitzian and concave, its subdifferential is calculated by the formula (see Clarke [4], Theorem 2.3.9)

$$\partial F_\varepsilon^\alpha(x) = \alpha F_\varepsilon(x)^{\alpha-1} \partial F_\varepsilon(x).$$

For any $g^i \in \partial F_\varepsilon(x^i)$ due to concavity of $F_\varepsilon^\alpha(x)$ the following inequality holds

$$\langle \alpha F_\varepsilon(x^i)^{\alpha-1} g^i, x^i - x^* \rangle \leq F_\varepsilon^\alpha(x^i) - F_\varepsilon^\alpha(x^*).$$

Denote $g^i = \mathbf{E}\{\xi^i/x^0, \ldots, x^i\}$. The following estimates are true

$$\begin{array}{rl}
& \| x^{i+1} - x^* \|^2 \leq \| x^i + \rho_i \xi^i - x^* \|^2 \\
= & \| x^i - x^* \|^2 + 2\rho_i \langle \xi^i, x^i - x^* \rangle + \rho_i^2 \| \xi^i \|^2 \\
= & \| x^i - x^* \|^2 + 2\rho_i \langle g^i, x^i - x^* \rangle + 2\rho_i \langle \xi^i - g^i, x^i - x^* \rangle + \rho_i^2 \| \xi^i \|^2 \\
= & \| x^i - x^* \|^2 + 2\rho_i \alpha^{-1} F_\varepsilon(x^i)^{1-\alpha} \langle \alpha F_\varepsilon(x^i)^{\alpha-1} g^i, x^i - x^* \rangle \\
& + 2\rho_i \langle \xi^i - g^i, x^i - x^* \rangle + \rho_i^2 \| \xi^i \|^2 \\
\leq & \| x^i - x^* \|^2 + 2\rho_i \alpha^{-1} F_\varepsilon(x^i)^{1-\alpha} (F_\varepsilon^\alpha(x^i) - F_\varepsilon^\alpha(x^*)) \\
& + 2\rho_i \langle \xi^i - g^i, x^i - x^* \rangle + \rho_i^2 \| \xi^i \|^2 \\
\leq & \| x^i - x^* \|^2 + 2\rho_i \alpha^{-1} \delta^{1-\alpha} (F_\varepsilon^\alpha(x^i) - F_\varepsilon^\alpha(x^*)) \\
& + 2\rho_i \langle \xi^i - g^i, x^i - x^* \rangle + \rho_i^2 \| \xi^i \|^2 .
\end{array}$$

Adding these inequalities from $i = 0$ up to $i = k$ we obtain

$$\| x^{k+1} - x^* \|^2 \leq \| x^0 - x^* \|^2 + 2\alpha^{-1}\delta^{1-\alpha} \left(\sum_{i=0}^k \rho_i F_\varepsilon^\alpha(x^i) - F_\varepsilon^\alpha(x^*) \sum_{i=0}^k \rho_i \right)$$

$$+ 2 \sum_{i=0}^k \rho_i \langle \xi^i - g^i, x^i - x^* \rangle + \sum_{i=0}^k \rho_i^2 \| \xi^i \|^2 .$$

Dividing this inequality by $\sum_{i=0}^{k} \rho_i$ and using the concavity property

$$F_\varepsilon^\alpha(\bar{x}^k) \geq \sum_{i=0}^{k} \rho_i F_\varepsilon^\alpha(x^i) / \sum_{i=0}^{k} \rho_i,$$

we obtain

$$0 \leq \| x^0 - x^* \|^2 / \sum_{i=0}^{k} \rho_i + 2\alpha^{-1}\delta^{1-\alpha}\left(F_\varepsilon^\alpha(\bar{x}^k) - F_\varepsilon^\alpha(x^*)\right) +$$

$$+2\frac{\sum_{i=0}^{k} \rho_i \langle \xi^i - g^i, x^i - x^* \rangle}{\sum_{i=0}^{k} \rho_i} + \frac{\sum_{i=0}^{k} \rho_i^2 \| \xi^i \|^2}{\sum_{i=0}^{k} \rho_i}.$$

Now taking a mathematical expectation of both sides of the inequality and using the estimate

$$\mathbf{E} \| \xi^i \|^2 = \mathbf{E}\{\mathbf{E}\{\| \xi^i \|^2 / x^0, \ldots, x^i\}\} \leq C(\varepsilon)$$

we get

$$0 \leq \frac{\mathbf{E} \| x^0 - x^* \|^2}{\sum_{i=0}^{k} \rho_i} + 2\alpha^{-1}\delta^{1-\alpha}\left(\mathbf{E}F_\varepsilon^\alpha(\bar{x}^k) - F_\varepsilon^\alpha(x^*)\right) + C(\varepsilon)\frac{\sum_{i=0}^{k} \rho_i^2}{\sum_{i=0}^{k} \rho_i}.$$

Finally we obtain

$$F_\varepsilon^\alpha(x^*) - \mathbf{E}F_\varepsilon^\alpha(\bar{x}^k) \leq \alpha\left(\mathbf{E} \| x^0 - x^* \|^2 + C(\varepsilon)\sum_{i=0}^{k} \rho_i^2\right)\left(2\delta^{1-\alpha}\sum_{i=0}^{k} \rho_i\right)^{-1}.$$

The proof for $\alpha \leq 0$ and $\alpha > 1$ is similar.

Theorem 4.3 *Under conditions of Lemma 4.1 the following estimates are true*

$$F_\varepsilon(x^*) - \mathbf{E}F_\varepsilon(\bar{x}^k) \leq \left(\frac{F_\varepsilon(x^*)}{\delta}\right)^{1-\alpha}\left(\frac{\mathbf{E}\|x^0-x^*\|^2 + C(\varepsilon)\sum_{i=0}^{k}\rho_i^2}{2\sum_{i=0}^{k}\rho_i}\right), \qquad \alpha \leq 1,$$

$$F_\varepsilon(x^*) - \mathbf{E}F_\varepsilon(\bar{x}^k) \leq \left(\frac{\mathbf{E}\|x^0-x^*\|^2 + C(\varepsilon)\sum_{i=0}^{k}\rho_i^2}{2\sum_{i=0}^{k}\rho_i}\right), \qquad \alpha \geq 1.$$

Proof. Let $\alpha < 0$. The function $y^\alpha, y > 0$, is convex, so due to Jensen's inequality

$$\mathbf{E}F_\varepsilon^\alpha(\bar{x}^k) - F_\varepsilon^\alpha(x^*) \geq (\mathbf{E}F_\varepsilon(\bar{x}^k))^\alpha - F_\varepsilon^\alpha(x^*).$$

Due to convexity of $y^\alpha, y > 0$, we have

$$(\mathbf{E}F_\varepsilon(\bar{x}^k))^\alpha \geq F_\varepsilon^\alpha(x^*) + \alpha F_\varepsilon^{\alpha-1}(x^*)(\mathbf{E}F_\varepsilon(\bar{x}^k) - F_\varepsilon(x^*))$$

hence

$$F_\varepsilon(x^*) - \mathbf{E}F_\varepsilon(\bar{x}^k) \leq (-\alpha)^{-1}F_\varepsilon^{1-\alpha}(x^*)(\mathbf{E}F_\varepsilon^\alpha(\bar{x}^k) - F_\varepsilon^\alpha(x^*)),$$

from where the required estimate follows.

Let $0 < \alpha < 1$. The function $y^\alpha, y > 0$, is concave, so due to Jensen's inequality

$$F_\varepsilon^\alpha(x^*) - \mathbf{E}F_\varepsilon^\alpha(\bar{x}^k) \geq F_\varepsilon^\alpha(x^*) - (\mathbf{E}F_\varepsilon(\bar{x}^k))^\alpha.$$

Due to concavity of $y^\alpha, y > 0$, we have

$$(\mathbf{E}F_\varepsilon(\bar{x}^k))^\alpha \leq F_\varepsilon^\alpha(x^*) + \alpha F_\varepsilon^{\alpha-1}(x^*)(\mathbf{E}F_\varepsilon(\bar{x}^k) - F_\varepsilon(x^*)),$$

hence

$$F_\varepsilon(x^*) - \mathbf{E}F_\varepsilon(\bar{x}^k) \le \alpha^{-1} F_\varepsilon^{1-\alpha}(x^*)(F_\varepsilon^\alpha(x^*) - \mathbf{E}F_\varepsilon^\alpha(\bar{x}^k)).$$

Let $\alpha = 0$. Function $\ln y, y > 0$, is concave, so by Jensen's inequality

$$\ln F_\varepsilon(x^*) - \mathbf{E}\ln F_\varepsilon(\bar{x}^k) \ge \ln F_\varepsilon(x^*) - \ln \mathbf{E}F_\varepsilon(\bar{x}^k) = \ln(F_\varepsilon(x^*)/F_\varepsilon(\bar{x}^k)),$$

hence

$$F_\varepsilon(x^*) - \mathbf{E}F_\varepsilon(\bar{x}^k) \le F_\varepsilon(x^*)(1 - \exp(-(\ln F_\varepsilon(x^*) - \mathbf{E}\ln F_\varepsilon(\bar{x}^k)) \le$$
$$\le F_\varepsilon(x^*)(\ln F_\varepsilon(x^*) - \mathbf{E}\ln F_\varepsilon(\bar{x}^k)).$$

The required estimates for the difference $(F_\varepsilon(x^*) - \mathbf{E}F_\varepsilon(\bar{x}^k))$ now follow from the appropriate estimates of Lemma 4.1

Let $\alpha \ge 1$. Then both $F_\varepsilon^\alpha(x)$ and $F_\varepsilon(x)$ are concave, so the required estimate for the difference $(F_\varepsilon(x^*) - \mathbf{E}F_\varepsilon(\bar{x}^k))$ is obtained from the last estimate of Lemma 4.1 with $\alpha = 1$.

Corollary 4.1 *If*

$$\lim_{i\to+\infty} \rho_i = 0, \qquad \sum_{i=0}^{\infty} \rho_i = +\infty,$$

then stochastic quasi-gradient method with trajectory averaging converges in mean, i.e. $\lim_{k\to\infty} \mathbf{E}(F_\varepsilon(x^*) - F_\varepsilon(\bar{x}^k)) = 0$.

Corollary 4.2 *If* $\rho_k = 1/\sqrt{k}$ *then*

$$\mathbf{E}(F_\varepsilon(x^*) - F_\varepsilon(\bar{x}^k)) \le$$
$$\le \left(\frac{F_\varepsilon(x^*)}{\delta}\right)^{\max(1-\alpha,0)} \left(\frac{\mathbf{E} \parallel x^0 - x^* \parallel^2 + C(\varepsilon)\cdot \ln k}{2\sqrt{k}} + 0(1/k)\right).$$

5 Conclusions

We considered an approach to probability function optimization. In general this function can be nonsmooth, nonconvex and even discontinuous. But under certain conditions it is continuous and quasi-concave (α-concave). In such a case, we could apply a subgradient algorithm for its maximization. But calculation of subgradients for probability functions still remains a challenge. So we uniformly approximate the original function by a sequence of quasi-concave (α-concave) functions for which the calculation of subgradients is easy. For solution of approximate problems we apply efficient Ermoliev-Nemirowski-Yudin's stochastic subgradient algorithm with trajectory averaging. Convergence and rate of convergence results are obtained. The algorithms accuracy estimates are similar to ones in convex stochastic programming case and differ only by a multiplier. This multiplier is $(F_{\max}/F_{\min})^{\max(0,1-\alpha)}$, where $F_{\max}$ and $F_{\min}$ are maximum and minimum values of the probability functions F over optimization range, α is a concavity parameter. If $F_{\min} = 0$, then the obtained estimates become only asymptotic because $F(\bar{x}^k) \to F_{\max}$.

References

[1] C. Borrell. Convex set functions in d-spaces, *Period. Math. Hung.*, V.6,N.2, (1975) 111-136.

[2] H.J. Brascamp and E.H. Lieb. On extensions of the Brunn-Minkowski and Prekopa-Liendler theorems, including inequalities for log-concave functions, and with application to the diffusion equations, *J. of Func, Anal.*, 22 (1976) 366-389.

[3] H. Brunn. *Uber Ovale und Eiflachen. Inaugural Dissertation* (München, 1987).

[4] F.H. Clarke. *Optimization and Nonsmooth Analysis* (Wiley, N.Y., 1983).

[5] S. Das Gupta. S-unimodal functions: related inequalities and statistical applications, *Sankhya. Ser.B*, 38 (1976) 301-314.

[6] S. Das Gupta. Brun-Minkowski inequality and its aftermath, *J. of Multivariate Anal.*, 10 (1980) 296-318.

[7] Yr.M. Ermoliev. *Methods of Stochastic Programming* (Nauka, Moscow, 1976) (In Russian).

[8] Y. Ermoliev, V. Norkin and R. Wets. The Minimization of Discontinuous Functions: Mollifier Subgradients, Working Paper WP-92-73, International Institute for Applied Systems Analysis, Laxenburg, Austria (1992).

[9] Yu.M. Ermoliev and E.A. Nurminski. Limit extremal problems, *Kibernetika*, 4 (1973) 130-132.

[10] Yu.M. Ermoliev and R.J.-B Wets (Eds.) *Numerical Techniques for Stochastic Optimization* (Springer-Verlag, 1988).

[11] A. Gaivoronski. On nonstationary stochastic optimization problems, *Kibernetika*, 4 (1978) 89-92, (In Russian, English translation in: *Cybernetics*, V. 14, N.4).

[12] Ch.M. Harvey. Special conditions on risk attitudes, Working Paper WP-87-3, International Institute for Applied Systems Analysis, Laxenburg, Austria, 1987.

[13] V. Kankova. On the convergence rate of empirical estimates in chance constrained stochastic programming, *Kybernetika*, V. 26, N. 6 (1990) 449-461.

[14] R.L. Keeney and H. Raiffa. *Decisions with Multiple Objectives: Preferences and Value Tradeoffs* (Wiley, 1976).

[15] A.I. Kibzun and V.Yu. Kurbakovskiy. Numerical algorithms of quantile optimization and their application to chance-constrained programming problems, *Izvestia Akademii Nauk SSSR. Technicheskaya Kibernetika*, 1 (1991) 75-81 (In Russian).

[16] A.I. Kibzun and V.V. Malyshev. Probabilistic optimization problems, Izvestia Akademii Nauk SSSR. *Technicheskaya Kibernetika*, 6 (1989) 46-55. (In Russian).

[17] V.S. Kiriluk. On two-stage stochastic programming problems and marginal functions properties, in: *Models and Methods for Operations Research, Risk Analysis and Reliability Theory,* ed. Yu.M. Ermoliev (Glushkov Institute of Cybernetics, Kiev, 1992) (In Russian).

[18] I.N. Kovalenko and A.N. Nakonechniy. Approximate Calculation and Optimization of Reliability (Naukova Dumka, Kiev, 1989) (In Russian).

[19] R. Lepp. The maximization of a probability function over simple sets, *Izvestia Akademii Nauk Estonskoy SSR. Physics and Mathematics*, V. 28, N. 4 (1979) 303-309 (In Russian).

[20] R. Lepp. Stochastic approximation type algorithm for the maximization of the probability function, *Izvestia Akademii Nauk Estonskoi SSR. Physics and Mathematics*, V. 32, N. 2 (1983) 150-156.

[21] L. Lusternik. Brunn-Minkowski inequality for Lebesgue measurable sets, *Dokladi AN SSSR*, V. 3, N. 2 (1935) 55-58.

[22] H. Minkowski. *Geometrie der Zahlen* (B.G. Teubner, Leipzig, Berlin, 1910).

[23] A.S. Nemirowski and D.B. Yudin. *Problems complexity and efficiency of optimization methods* (Moscow, Nauka, 1979) (In Russian).

[24] V.I. Norkin. Optimization of probabilities, Preprint 89-6 (Institute of Cybernetics, Kiev, 1989) (In Russian).

[25] V.I. Norkin and N.V. Roenko. α-concave functions and measures and their application, Manuskripte, Institüt für Operations Reseach der Universität Zürich, Zürich, Switzerland (1990).

[26] V.I. Norkin and N.V. Roenko. α-concave functions and measures and their application, *Kibernetika i Sistemny Analiz*, 6 (1991) 77-88 (In Russian, English translation in *Cybernetics and Systems Analysis*, V.27, N.6).

[27] E.A. Nurminski. *Numerical methods for solving deterministic and stochastic minimax problems* (Kiev, Naukova Dumka, 1979) (In Russian).

[28] E. Parzen. On estimation of a probability density function and the mode, *Annals of Mathematical Statistics*, V. 33 (1962) 1065-1076.

[29] B.T. Polyak. *Introduction to Optimization* (Nauka, Moscow, 1983) (In Russian).

[30] A. Prekopa. On probabilistic constrained programming, in: *Proceedings of the Princeton Symposium on Mathematical Programming,* Princeton University Press, New York (1970) 113-138.

[31] A. Prekopa. Logarithmic concave measures with applications to stochastic programming, *Acta Sci. Math.*, 32 (1971) 301-318.

[32] A. Prekopa. Logarithmic concave measures and related topics, in: *Stochastic Programming,* Ed. M.A.H. Dempster, Academic Press, London (1980) 63-82.

[33] A. Prekopa. Numerical solution of probabilistic constrained programming problems. in: *Numerical techniques for Stochastic Optimization,* Eds. Yu. Ermoliev and R.J.-B Wets, Springer, Berlin (1988) 123-139.

[34] E. Raik. Qualitative Studies in nonlinear stochastic programming problems, *Izvestia Akademii Nauk Estonskoy SSR. Phys and Math.*, V.20, N. 8 (1971) (In Russian).

[35] E. Raik. On stochastic programming problems with probability functions and quantile functions, *Izvestia Akademii Nauk Estonskoy SSR. Phys and Math.*, V.21, N. 2 (1972) 142-148 (In Russian).

[36] E. Raik. The differentiability in the parameter of the probability function and optimization of the probability function via the stochastic pseudogradient method, *Izvestia Akademii Nauk Estonskoy SSR. Phys and Math.*, V.21, N. 1 (1975) 3-9.

[37] Y. Rinott. On convexity of measures, *Ann. Prob.* 4 (1976) 1020-1026.

[38] N.V. Roenko. Stochastic programming problems with integral functionals from multivalued mappings, Abstract of Ph.D.Thesis (V.M. Glushkov Institute of Cybernetics, Kiev, 1983) (In Russian).

[39] R.T. Rockafellar. *Convex Analysis* (Princeton University Press, 1970).

[40] R.T. Rockafellar. *The Theory of Subgradients and Its Application to Problems of Optimization: Convex and Nonconvex Functions* (Helderman Verlag, Berlin, 1981).

[41] M. Rosenblatt. Remarks on some nonparametric estimates of a density function, *Annals of Mathematical Statistics*, V. 27 (1956) 832-835.

[42] W. Rõemisch and R. Schultz. On distribution sensitivity in chance constrained programming, in: *Advances in Mathematical Optimization,* Ed. J.Guddat (Akademie-Verlag, Berlin, 1988) 161-168.

[43] G. Salinetti. Approximations for chance-constrained programming problems, *Stochastics*, 10 (1983) 157-179.

[44] J. Simon. Second Variation in Domain Optimization Problems, in: *International Series of Numerical Mathematics,* 91, eds. F. Kappel, K. Kunish and W. Schappacher (Birkhauser Verlag, 1989) 361-378.

[45] S.L. Sobolev. *Some Applications of Functional analysis in Mathematical Physics* (1-st ed., Leningrad State University Press, 1950; 3-d ed., Nauka, Moscow, 1988) (In Russian).

[46] V.A. Steklov. *Main Problems of Mathematical Physics* (1-st ed.m Part 1,2, Petrograd, 1922, 1923; 2-nd ed., Nauka, Moscow, 1983) (In Russian).

[47] T. Szantai. Calculation of the multivariate probability distribution function values and their gradients, Working Paper WP-87-82, International Institute for Applied Systems Analysis, Laxenburg, Austria (1987).

[48] E. Tamm. On g-concave functions and probability measures, *Izvestia Akademii Nauk Estonskoy SSR. Phys and Math.*, V.26, N. 4 (1977) 376-379.

[49] E. Tamm. On a probability function optimization, *Izvestia Akademii Nauk Estonskoy SSR. Phys and Math.*, V.28, N. 1 (1979) 17-24 (In Russian).

[50] S.P. Uryas'ev. Differentiability of an Integral Over a Set Defined by Inclusion, *Kibernetika* (Kiev), 5 (1988) 83-86 (In Russian, English translation in: *Cybernetics*, V. 24, N. 5 (1988) 638-642).

[51] S.P. Uryas'ev. A differentiation formula for integrals over sets given by inclusion, *Numerical Functional Analysis and Optimization*, 10 (1989) 827-841.

[52] R. Wets. Stochastic programming: solution techniques and approximation schemes, in : *Mathematical Programming: the State of the Art*, ed.A. Bachen, M. Groetschel and B. Korte (Springer Verlag, Berlin, 1983) 566-603.

[53] E. Yubi. Minimization of a probability function by means of Monte-Carlo method, *Proceedings of Tallin Polytechnical Institute,* 411 (1976) 57-76 (In Russian).

[54] E. Yubi. On probability function minimization problem, *Proceedings of Tallin Polytechnical Institute,* 482 (1980) 121-130 (In Russian).