Comparison of approximations in stochastic and robust optimization programs

Michal Houda

Abstract: The paper deals with two wide areas of optimization theory: stochastic and robust programming. We specialize to different approaches when solving an optimization problem where some uncertainties in constraints occur. To overcome uncertainty, we can request the solution to be feasible to all but a small part of constraints. Both approaches gives us different methods to deal with this requirement. We try to find fundamental differencies between them and illustrate the differencies on a simple numerical example.

MSC 2000: 90C15, 90C25

Key words: Chance-constrained programming, robust programming, approximations, sampling method

1 Introduction

Many engineering and economic problems are mathematically viewed as optimization problems subject to convex constraints; but usually, input parameters of these problems are not known precisely. Sometimes, randomness of some of the parameters could be disregarded – e. g., replaced by some deterministic version, often by their average value – but there are examples where this approach does not satisfy our needs (see e. g. Kall's linear programming example, [9]). Different methods are to be used to deal with such a class of problems.

Generally, there are two main approaches to deal with constrained optimization with uncertainty: robust programming approach and stochastic programming approach. In *robust programming* one seeks for a solution which simultaneously satisfy all possible realizations of the constraints. The *stochastic programming* approach works with the probabilistic distribution of uncertainty and the constraints are required to be satisfied up to prescribed level of probability (the last is known as the chance-constrained optimization). However, both approaches lead to computationally intractable problems and we have to consider some kind of approximation for them.

Acknowledgements. This research was supported by the Czech Science Foundation under the projects No. 402/03/H057, 402/04/1294, and No. 402/05/0115.

1.1 Uncertain convex program

An uncertain convex program is an optimization problem in which the constraints are not precisely known. Formally, consider $\xi \in \Xi \subset \mathbb{R}^s$ (a random, uncertainty, or instance parameter), $X \subset \mathbb{R}^n$ convex and closed set, and a function $f(x;\xi) : X \times \Xi \to \mathbb{R}$ convex in x for all $\xi \in \Xi$; an *uncertain convex program* (UCP) is then a problem

$$\min_{x \in X} c'x \quad \text{subject to} \quad f(x;\xi) \leqslant 0. \tag{UCP}$$

The function f is assumed to be scalar here; in fact, multiple scalar constraints $f_i(x;\xi) \leqslant 0$ can be converted to a single scalar constraint of the form $f(x;\xi) = \max_{i=1,\dots,m} f_i(x;\xi) \leqslant 0$. Without loss of generality, we also assume the objective of (UCP) to be linear. If a realization of ξ is known and fixed, deterministic optimization could be easily used to solve (UCP). Unfortunately, such solutions are often very sensitive to a perturbation of ξ. Next, we introduce two main concepts that appear in the literature dealing with the uncertainty of ξ.

1.2 Chance-constrained approach

The stochastic programming approach, or precisely the chance-constrained approach assumes that ξ is a random variable on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with known distribution and checks the constraints in (UCP) to be fulfilled with a certain degree of probability. If P is a probability distribution of ξ and $\varepsilon \in [0;1]$ is an acceptable level of constraint violation, then the *chance (probability) constrained* version of the uncertain program is

$$\min_{x \in X} c'x \quad \text{subject to} \quad x \in X_\varepsilon := \big\{x \in X;\ P\{f(x;\xi) > 0\} \leqslant \varepsilon\big\}. \tag{PCP}$$

(PCP) problem is not necessarily a convex optimization problem even if function f is convex in x for all ξ. Another difficulties arise when we evaluate the probability measure in X_ε because it often involves a multidimensional integral. We deal with approximation to this problem in Section 2.

The chance-constrained optimization dates a long history, starting at least by the work of Charnes and Cooper [5]. Above all, an extensive presentation of the topic (in particular we mention conditions implying convexity of X_ε) is given in Prékopa's book [12].

1.3 Robust programming approach

The robust programming approach is an alternative way to deal with uncertainty parameters in (UCP). It is also known as 'min-max' or 'worst-case' approach due to the nature of the problem. In robust optimization we look for a solution which is feasible for *all* possible instances of ξ; this approach leads to the problem

$$\min_{x \in X} c'x \quad \text{subject to} \quad f(x;\xi) \leqslant 0 \text{ for all } \xi \in \Xi. \tag{RCP}$$

Throughout, we assume that there exists a feasible solution to (RCP). The robust convex programming problem is convex but it is numerically hard to solve because of infinite number of constraints. The robust optimization methods propose some relaxation techniques to deal with such a problem. In the next section, we consider a solution method based on 'randomization' of the parameter ξ and the sampling techniques. Another disadvantage is the fact that robust programming aproach gives the same weight to all of the values of the uncertain parameter.

The framework of robust optimization problem was introduced by Ben-Tal and Nemirovski [1] and developed by other authors in various direction, see e. g. [2], [7] and references therein.

2 Sampled convex programs

The probability distribution P of ξ is rarely known completely. Instead, various approximations and estimates are used. One of the very important techniques is using a (random) sample of the parameter ξ. There are again two ways of using this technique which we may call again the chance-constrained programming approach and the robust programming approach.

2.1 Chance-constrained sampled program

Consider a set of independent samples $\xi_1, \dots, \xi_N$ distributed according to P, the original distribution of the parameter ξ. We define the *empirical distribution function* as a discrete random variable of the form $P_N := \frac{1}{N} \sum \delta_{\xi_i}$ where δ_ξ denotes the Dirac measure placing the unit mass at ξ. The (PCP) problem is now approximated, for the given sample, by replacing the original probability distribution P by P_N and the problem then reads

$$\min_{x \in X} c'x \quad \text{subject to} \quad x \in X[\varepsilon, N] := \big\{ x \in X; \frac{1}{N} \operatorname{card}\{i; f(x;\xi_i) > 0\} \leqslant \varepsilon \big\}. \tag{PCP$_N$}$$

The essential idea of (PCP$_N$) is that the relative frequency of constraint violations correspond to the desired upper level of infeasibility in (PCP). (PCP$_N$) is the program with a single constraint and in some simple cases it is computationally tractable.

There exists many results in the theory of stability of stochastic optimization problems dealing with a question how far is the optimal solution of (PCP$_N$) from the original optimal solution of (PCP); among all we refer to works [8], [10], [11],

[13], [14], references therein, and many of other authors. Here, general stability theorem (Theorem 1 in [8]) can be considered as a base for further special results. As an example we recall the following proposition, formulated as in the original paper for a linear (vector) function $f(x;\xi)$.

Proposition 2.1 (Corollary 2 in [8] or Theorem 47 in [14]). *Assume that*

1. $f(x;\xi) = \xi - Tx$, $T \in \mathbb{R}^n \times \mathbb{R}^s$ *is a constant matrix of parameters;*

2. P *is a logarithmic concave measure, i. e.,* $P\{\lambda B_1 + (1-\lambda)B_2\} \geqslant P(B_1)^\lambda P(B_2)^{1-\lambda}$ *is valid for all* $\lambda \in [0;1]$ *and all convex Borel* $B_1, B_2 \subset \mathbb{R}^s$ *such that* $\lambda B_1 + (1-\lambda)B_2$ *is also Borel set;*

3. *the optimal solution set* $\Psi(P)$ *of* (PCP) *is nonempty and bounded and* $\Psi(P) \cap \operatorname{argmin}_{x\in X} c'x = \emptyset$*;*

4. *there exists an* $\bar{x} \in X$ *such that* $F_P(T\bar{x}) > 1-\varepsilon$, F_P *is the distribution function corresponding to* P*;*

5. $\log F_P$ *is strongly concave on some convex neighbourhood* $T\Psi(P)$*, i. e., there is a constant* $c > 0$ *such that* $\log F_P(\lambda t_1 + (1-\lambda)t_2) \geqslant \lambda \log F_P(t_1) + (1-\lambda)\log F_P(t_2) + \frac{1}{2}c\lambda(1-\lambda)||t_1 - t_2||^2$ *is valid for all* $\lambda \in [0;1]$ *and all* t_1, t_2 *from the convex neighbourhood of* $T\Psi(P)$*.*

Then there are constants $L > 0, \delta > 0$ *such that*

$$d_H(\Psi(P), \Psi(P_N)) \leqslant L\sqrt{d_K(P, P_N)}$$

whenever $d_K(P, P_N) < \delta$*.*

Here, d_H denotes the Hausdorff distance on subsets on $\mathbb{R}^n$, $d_K(P, P_N)$ is the Kolmogorov metric ($\sup_t |F_P(t) - F_{P_N}(t)|$). Kolmogorov metric $d_K(P, P_N)$ converges almost surely to zero under rather general conditions, hence the distance between optimal solutions of (PCP) and (PCP$_N$) are expected to converge to zero. This will be illustrated in Section 3.

2.2 Robust sampled program

Recently, Calafiore and Campi [4] and de Farias and Van Roy [6] independently proposed the following approximations to (RCP). Consider again a set of independent samples $\xi_1, \ldots, \xi_N$ distributed according to P. The (RCP) is then approximated by asking the constraints to be satisfied for all ξ_i:

$$\min_{x\in X} c'x \quad \text{subject to} \quad X[N] := \left\{x \in X; f(x;\xi_i) \leqslant 0 \text{ for } i = 1, \ldots, N\right\} \qquad (\mathrm{SCP}_N)$$

(SCP$_N$) is again a convex program with a finite number of convex constraints and therefore it is computationally tractable. It is an approximation to (RCP) in the following framework: we do not require the constraints to be satisfied for all realizations of ξ, but only for a high number of samples, which are moreover the most probable to happen. Calafiore and Campi [3] found a rule to set up N in order to have the optimal solution of (SCP$_N$) feasible in (PCP):

Proposition 2.2 (Theorem 2 in [3]). *For a fixed $\varepsilon, \beta > 0$, the optimal solution of* (SCP$_N$) *is feasible in* (PCP) *with a probability at least $1 - \beta$ if*

$$N \geqslant \frac{2n}{\varepsilon} \ln \frac{1}{\varepsilon} + \frac{2}{\varepsilon} \ln \frac{1}{\beta} + 2n.$$

3 Numerical study

In the current section, we illustrate both propositions from Section 2 on a simple numerical example. Thus, consider the following uncertain convex program:

$$\min_{x \in \mathbb{R}} x \quad \text{subject to} \quad x \geqslant \xi \tag{UCP}$$

where $\xi \subset \mathbb{R}$ is distributed according to standard normal distribution N(0; 1), and exponential distribution Exp(1) respectively, with the distribution function denoted by F for both cases. According to Sections 1 and 2, we define the following deterministic programs

$$\min_{x \in \mathbb{R}} x \quad \text{subject to} \quad x \geqslant F^{-1}(1 - \varepsilon) \tag{PCP}$$

$$\min_{x \in \mathbb{R}} x \quad \text{subject to} \quad \frac{1}{N} \operatorname{card}\{i; x < \xi_i\} \leqslant \varepsilon \tag{PCP$_N$}$$

$$\min_{x \in \mathbb{R}} x \quad \text{subject to} \quad x \geqslant \max_{i=1,\dots,N} \xi_i \tag{SCP$_N$}$$

Normal and exponential distribution functions are defined on an unbouded set, hence the robust program is not well defined – there is not a solution feasible to all the instances of ξ. However, practical interest of this fact is small; we can use some suitable transformation of the distribution in order to obtain a bounded support and define the new problem. We do not pursue this direction in the following.

In our simple case, the lower boundary of the feasibility set of each of the problems coincides with the optimal solution of the problem. In the sequel, we compute the optimal solution of (PCP) (that is $1 - \varepsilon$-quantile of F) and the approximated solutions of (PCP$_N$) and (SCP$_N$) for different values of N. To create the array of graphics in Figure 1, we set up $\varepsilon = 0.05$, $N = 30, 300,$ and 3000 respectively, and F to the distribution functions of two probability distributions (the left-hand column of the array represents normal distribution, the right-hand column represents exponential distribution). The sampling procedure is repeated 200 times for each sample size in order to estimate densities (histograms) for the optimal solutions.

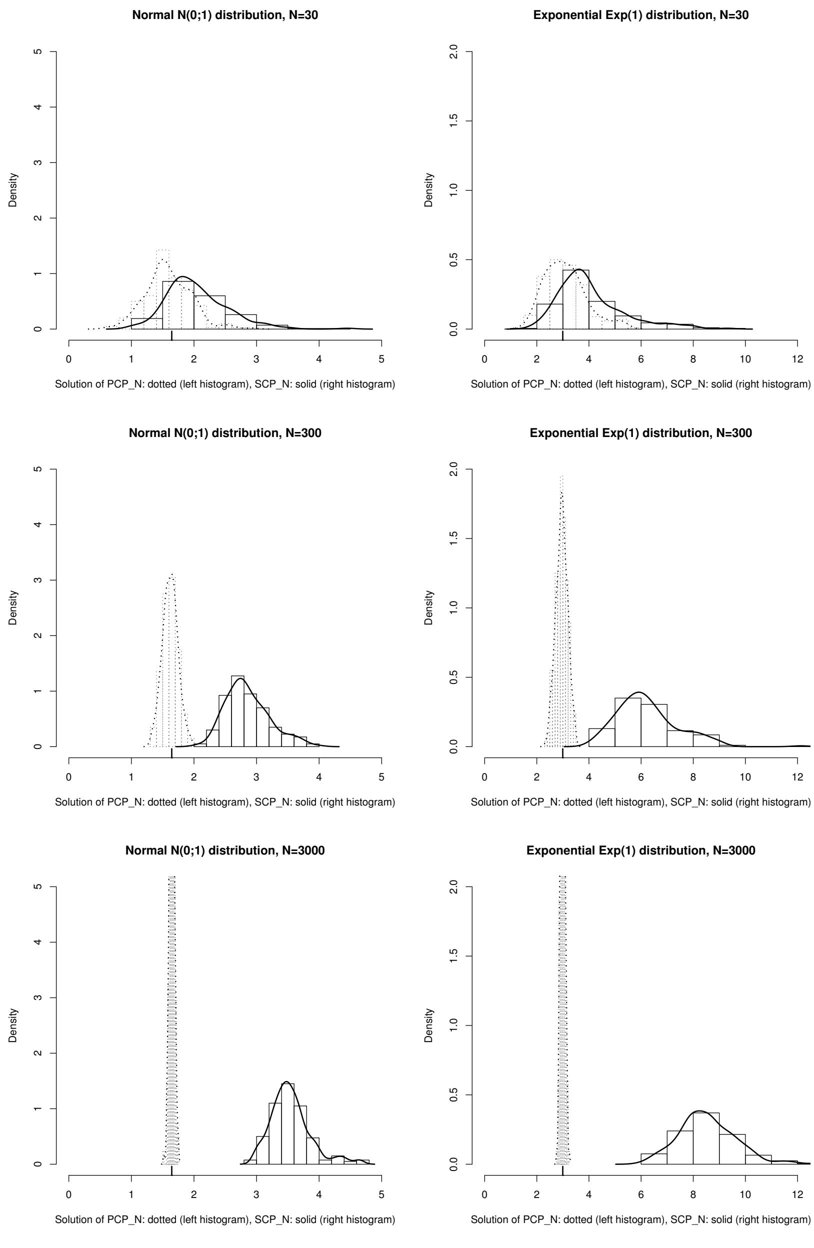

Figure 1: Convergence of optimal values for (SCP_N) and (PCP_N)

In the first rank, dotted histograms represent the fact that the optimal solution of the chance-constrained sampled problem (PCP_N) converges, as N goes to infinity, to the solution of (PCP), marked by a short tickmark on x-axis. This sampling method is useful espetially if the number N of samples is high, as the possible error in estimating optimal solution decreases. The optimal solution of the second mentioned approach, robust sampled problem (SCP_N), goes to the upper boundary of the support of F (i. e. to infinity in our cases), but with rapidly decreasing rate. The tickmark now represents the lower boundary of ε-feasibility set X_ε, i. e. the limiting point for which a solution is feasible for (PCP) with a high probability. You could observe the fact mentioned in Proposition 2.2 – if N is greater than 241, then the optimal value of the (SCP_N) program is feasible in (PCP) with probability of 0.95.

4 Conclusion

Choosing the approximation method to an uncertain convex program is ambiguous. There is no reason to measure difference between the two optimal solutions as they originate in different context. But the selected method has to fill up the needs of practical dimension of the problem:

- how much the probability of violation of the constraints is crucial,
- how many samples one has at disposition or can generate.

Getting an answer to the first question stronger, one's preferences have to be directed towards the robust sampled problems assuring the high probability of fulfilling the constraints. Chance to fulfill the constraints by the optimal solution of chance-constrained sampled problem is only *approximately* the desired value $1 - \varepsilon$, especially if the number of samples is low. But this solution could be useful in cases where the $1 - \varepsilon$ level is not crucial and our preferences are pointed more likely towards costs savings solutions.

We have illustrated how these general theses apply in a simple optimization problem and two rather 'representative' distributions. The generalization is possible – for other distributions it is straightforward, for the multidimensional case the problem is likely to be a problem of getting data and obtaining a clear representation.

References

[1] A. Ben-Tal and A. Nemirovski. Robust convex optimization. *Mathematics of Operations Research*, 23(4):769–805, 1998.

[2] D. Bertsimas and M. Sim. Tractable approximations to robust conic optimization problems. *Mathematical Programming*, published online, Dec. 2005.

[3] G. Calafiore and M. C. Campi. Decision making in an uncertain environment: the scenario based optimization approach. In J. Andrýsek, M. Kárný, and J. Kracík, editors, *Multiple Participant Decision Making*, pages 99–111. Advanced Knowledge International, 2004.

[4] G. Calafiore and M. C. Campi. Uncertain convex programs: randomized solutions and confidence levels. *Mathematical Programming*, 102 (1, Ser. A):25–46, 2005.

[5] A. Charnes and W. W. Cooper. Chance constrained programming. *Management Science*, 6:73–79, 1959.

[6] D. P. de Farias and B. Van Roy. On constraint sampling in the linear programming approach to approximate dynamic programming. *Mathematics of Operations Research*, 29(3):462–478, 2004.

[7] L. El Ghaoui, F. Oustry, and H. Lebret. Robust solutions to uncertain semidefinite programs. *SIAM Journal of Optimization*, 9(1):33–52, 1998.

[8] R. Henrion and W. Römisch. Metric regularity and quantitative stability in stochastic programs with probabilistic constraints. *Mathematical Programming*, 84(1):792–818, 1999.

[9] P. Kall. *Stochastic Linear Programming.* Springer-Verlag, Berlin, 1976.

[10] V. Kaňková. On the convergence rate of empirical estimates in chance constrained stochastic programming. *Kybernetika*, 26:449–461, 1990.

[11] V. Kaňková. A note on relationship between distribution function estimation and estimation in stochastic programming. In *Transactions of the 12th Prague Conference on Information Theory, Statistical Decision Functions and Random Processes*, pages 122–125. Academy of Sciences of the Czech Republic, Prague, 1994.

[12] A. Prékopa. *Stochastic Programming.* Akadémiai Kiadó, Budapest, 1995.

[13] S. T. Rachev and W. Römisch. Quantitative stabilisy in stochastic programming: The method of probability metrics. *Mathematics of Operations Research*, 27(4):792–818, 2002.

[14] W. Römisch. Stability of stochastic programming problems. In A. Ruszczynski and A. Shapiro, editors, *Stochastic Programming. Handbooks in Operations Research and Management Science*, volume 10, pages 483–554. Elsevier, Amsterdam, 2003.

Michal Houda: Academy of Sciences of the Czech Republic, Institute of Information Theory and Automation, Pod Vodárenskou věží 4, Praha 8, 182 08, Czech Republic, *houda@karlin.mff.cuni.cz*

Empirical Processes in Stochastic Programming

Vlasta Kaňková, Michal Houda

Abstract: Usually, it is very complicated to investigate and to solve optimization problems depending on a probability measure. To this end a stability of them, considered with respect to a probability measure space, has been discussed in the stochastic programming literature many times. We focus on the investigation of the stability with respect to the Wasserstein and to the Kolmogorov metrics with "underlying" $\mathcal{L}_1$ space. Moreover, we apply achieved stability results to empirical estimates.

MSC 2000: 90C15, 62G20
Key words: Stochastic programming, stability, empirical estimates, Wasserstein metric, Kolmogorov metric, independent and dependent random samples

1 Introduction

Let $(\Omega, \mathcal{S}, P)$ be a probability space, $\xi (:= \xi(\omega) = [\xi_1(\omega), \ldots, \xi_s(\omega)])$ be an s–dimensional random vector defined on $(\Omega, \mathcal{S}, P)$, $F (:= F(z), z \in R^s)$ be the distribution function of ξ, P_F denote the probability measure corresponding to F, $Z_F \subset R^s$ denote the support of P_F. Let, furthermore, $g_i (:= g_i(x, z))$, $i = 0, 1, \ldots, m$ be real–valued (say continuous) functions defined on $R^n \times R^s$, $X_F \subset R^n$ be a nonempty set depending (generally) on F, $X \subset R^n$ be a nonempty set, $\alpha, \alpha_i \in \langle 0, 1 \rangle$, $i = 1, \ldots, m$. (The symbol R^n, $n \geqslant 1$ is reserved for the n–dimensional Euclidean space.)

A rather general stochastic programming problem can be introduced in the form.

Find

$$\varphi(F) = \inf\{\mathsf{E}_F g_0(x, \xi) | x \in X_F\}. \tag{1}$$

E_F denotes the operator of mathematical expectation corresponding to F.

It is known from the stochastic programming literature that the case $X_F = X$ not depending of F includes the well–known stochastic programming problems with penalty and recourse, whereas

$$X_F := X_F(\alpha) = \{x \in X : P_F\{g_i(x, \xi) \leqslant 0, i = 1, 2, \ldots, m\} \geqslant \alpha\} \tag{2}$$

Acknowledgements. This research was supported by the Grant Agency of the Czech Republic under Grants 402/04/1294, 402/05/0115 and 402/03/H057.

corresponds to stochastic programming problem with joint probability constraints,

$$X_F := \bar{X}_F(\bar{\alpha}) = \bigcap_{i=1}^{m} \{x \in X : P_F\{g_i(x, \xi) \leqslant 0\} \geqslant \alpha_i\}, \quad \bar{\alpha} = (\alpha_1, \ldots, \alpha_m) \tag{3}$$

corresponds to stochastic programming problem with individual probability constraints.

To solve the problem (1), a complete information on F is assumed. However in applications very often at least one of the following cases happen: F must be replaced by its statistical estimate; F must be (for numerical difficulties) replaced by some simpler one; the actual distribution function is a little modified F. Consequently, it is suitable (or even necessary) to investigate the stability w.r.t. the probability measure space and statistical estimates. Great attention has been paid to these both problems; see e.g. [2], [4], [5], [12], [13], [16], [20].

It follows from the stability results that it is practically impossible to evaluate numerically (in the case $s > 1$) an "approximation" error obtained by stability results. According to this fact, we try to employ the $\mathcal{L}_1$ norm to obtain some stability results. An effort to reduce the stability problems (to $\mathcal{L}_1$ space) has appeared already in the literature, see e.g. [7], [10] or [18]. The achieved results can be applied to empirical estimates including some weak dependent random samples. For information about dependent samples we refer to [1], [4], [9], [21].

2 Problem Analysis

Let F, G be two s–dimensional distribution functions. Knowing the stochastic programming literature, we can suppose that (under rather general assumptions) there exist constants $m_{W_1}, m_K \geqslant 0$ such that

$$|\varphi(F) - \varphi(G)| \leqslant m_{W_1}(d_{W_1}(P_F, P_G)) + m_K[d_K(P_F, P_G)]^{\frac{1}{s}}, \tag{4}$$

where d_{W_1}, d_K denote the Wasserstein and the Kolmogorov metrics. Of course, other metrics can replace the two ones mentioned above (for more details see e.g. [11], [13]). The Kolmogorov metric is given by the relation

$$d_K(F, G) := d_K(P_F, P_G) = \sup_{z \in R^s} |F(z) - G(z)|. \tag{5}$$

To define the Wasserstein metric $d_{W_1}(F, G)$, let $\mathcal{P}(R^s)$ denote the set of all (Borel) probability measures on R^s. If $\mathcal{M}_1(R^s) = \{\nu \in \mathcal{P}(R^s) : \int_{R^s} \|z\| \nu(dz) < \infty\}$ and $\mathcal{D}(P_F, P_G)$ denotes the set of those measures on $\mathcal{P}(R^s \times R^s)$ whose marginal measures are P_F and P_G, $\|\cdot\|$ denotes a suitable norm in R^s, then

$$\begin{aligned} & d_{W_1}(F, G) := d_{W_1}(P_F, P_G) = \\ & \inf\{ \int_{R^s \times R^s} \|z - \bar{z}\| \kappa(dz \times d\bar{z}) : \kappa \in \mathcal{D}(P_F, P_G)\}, \quad P_F, P_G \in \mathcal{M}_1(R^s). \end{aligned} \tag{6}$$

Moreover, it has been proven in [19] that in the case of the Euclidean norm

$$d_{W_1}(P_F,\, P_G) = \int\limits_{-\infty}^{+\infty} |F(z) - G(z)| dz \quad \text{for} \quad s = 1. \tag{7}$$

It is easy to see that right hand side of (4) is, generally, suitable for theoretical investigation, however it is not suitable for a bounds determination. The only exception is the case of $s = 1$. Moreover, if (in the case $s = 1$) F^N denotes an empirical distribution function determined by an independent random sample $\xi^1, \ldots, \xi^N$, corresponding to the measure P_F, then assumptions under which $d_{W_1}(P_F,\, P_{F^N}) \longrightarrow_{N \to \infty} 0$ can be found in [17]. Furthermore, for $s = 1$ the process

$$\sqrt{N} d_{W_1}(P_F,\, P_{F^N}) = \int\limits_{-\infty}^{+\infty} \sqrt{N} |F^N(t) - F(t)| dt \tag{8}$$

is called the integrated empirical process. It has been, in the case of uniform distribution on (0, 1) investigated in [17], where a limit distribution can be found. A numerical investigation for some other (one–dimensional) distribution functions are presented in [9]. There were also recalled that if $\int_{-\infty}^{\infty} \sqrt{F(z)(1 - F(z))} dz < +\infty$, then

$$\int\limits_{-\infty}^{\infty} \sqrt{N} |F^N(z) - F(z)| dz \longrightarrow_d \int\limits_{-\infty}^{\infty} |U(F(z))| dz, \quad U \quad \text{is the Brownian bridge.} \tag{9}$$

In the case of the Kolmogorov metric, if P_F is absolutely continous w.r.t. the Lebesgue measure on R^1, then

$$P\{\sqrt{N} \sup_{-\infty < t < +\infty} |F^N(t) - F(t)| < z\} \quad \longrightarrow \quad \begin{array}{ll} \sum\limits_{k=-\infty}^{\infty} (-1)^k e^{-2k^2 z^2} & \text{for} \quad z > 0, \\ 0 & \text{for} \quad z \leqslant 0. \end{array}$$

3 Some Auxiliary Assertions

Let $\eta(:= \eta(\omega) = [\eta_1(\omega), \ldots, \eta_s(\omega)])$ be an s–dimensional random vector defined on $(\Omega, \mathcal{S}, P)$. We denote the distribution function of η by G.

Lemma 3.1. *Let $X \subset R^n$ be a nonempty set, F_i, G_i, $i = 1, 2, \ldots, s$ denote one–dimensional marginal distribution functions corresponding to F, G; P_F, $P_G \in$*

$\mathcal{M}_1(R^s)$. *If for every* $x \in X$, g_0 *is a Lipschitz (with respect to* $\mathcal{L}_1$ *norm) function of* $z \in R^s$ *with the Lipschitz constant* L *not depending on* $x \in X$, *then for* $x \in X$

$$|\mathsf{E}_F g_0(x, \xi) - \mathsf{E}_G g_0(x, \eta)| \leqslant L \sum_{i=1}^{s} \int\limits_{-\infty}^{+\infty} |F_i(z_i) - G_i(z_i)| dz_i. \tag{10}$$

Proof. First, let $x \in X$ be arbitrary. We prove the assertion for $s = 2$.

First, evidently, $F_1(\xi_1)$, $F_2(\xi_2)$, $G_1(\eta_1)$, $G_2(\eta_2)$ and the conditional distribution functions $F^{\xi_2|\xi_1}(\xi_2|\xi_1 = z_1)$, $G^{\eta_2|\eta_1}(\eta_2|\eta_1 = z_1)$ (ξ_2 conditioned by ξ_1, η_2 conditioned by η_1) have uniform distribution function on $\langle 0, 1\rangle$. To construct $2s$–dimensional measure ν with s–dimensional marginals we employ the idea of the proof of Theorem in [19] (see also Remark 1 there). To this end let ζ_1, ζ_2 be random variables with uniform probability measure on $\langle 0, 1\rangle$. We set

$$\begin{aligned} \xi_1 &= F_1^{-1}(\zeta_1), & \eta_1 &= G_1^{-1}(\zeta_1), \\ \xi_2 &= F_2^{-1}(\zeta_2), & \eta_2 &= G_2^{-1}(\zeta_2), \end{aligned}$$

Evidently we can obtain by this a relationship between ξ, η and $\zeta = (\zeta_1, \zeta_2)$. Moreover, we obtain $2s$–dimensional measure ν with s–dimensional marginals P_F, P_G and 2–dimensional ν_1, ν_2 corresponding to P_{F_i}, P_{G_i}, $i = 1, 2$. According to the results of [19] we obtain

$$\begin{aligned} &|\mathsf{E}_F g_0(x, \xi) - \mathsf{E}_G g_0(x, \eta)| = |\mathsf{E}_\nu[g_0(x, \xi) - g_0(x, \eta)]| \leqslant \\ &\mathsf{E}_\nu|g_0(x, \xi) - g_0(x, \eta)| \leqslant L\mathsf{E}_\nu\{|\xi_1 - \eta_1| + |\xi_2 - \eta_2|\} = \\ &L\{\mathsf{E}_{\nu_1}\{|\xi_1 - \eta_1|\} + \mathsf{E}_{\nu_2}|\xi_2 - \eta_2|\} = \\ &L\{\int\limits_{-\infty}^{\infty} |F_1^\xi(z_1) - G_1^\eta(z_1)| dz_1 + \int\limits_{-\infty}^{\infty} |F_2^\xi(z_2) - G_2^\eta(z_2)| dz_2\}. \end{aligned}$$

We have finished the proof of Lemma 3.1 for $s = 2$. Evidently, we can employ this idea to prove the assertion for an arbitrary natural number s. □

Lemma 3.2. *Let* $s = 1$ *and let the probability measure* P_F *be absolutely continuous with respect to the Lebesgue measure on* R^1, Z_F *be a compact set with a radius* $R = \sup\limits_{z \in Z_F} z - \inf\limits_{z \in Z_F} z$. *If* F^N *is determined by independent random sample, then*

1. $P\{d_{W_1}(F, F^N) > 2Rt\} \leqslant (\frac{1}{t} + 1)\exp\{-2Nt^2\}$, $\quad N = 1, 2, \ldots,\quad t \in (0, 1)$,
2. $P\{N^\beta d_{W_1}(F, F^N) > 2Rt\} \longrightarrow_{\longrightarrow\infty} 0$ for $\beta \in (0, \frac{1}{2})$, $t \in (0, 1)$.

Proof. Let $t \in (0, 1)$ be arbitrary. Then there exists a natural number k, $\frac{1}{t} \leqslant k \leqslant (\frac{1}{t} + 1)$ and $z_1, \ldots, z_k \in R^1$, $\inf\limits_{z \in Z_F} z = z_1 \leqslant z_2 \leqslant \ldots \leqslant z_k = \sup\limits_{z \in Z_F} z$ such that $F(z_{i+1}) - F(z_i) \leqslant t$, $i = 1, \ldots, k-1$. According to the properties of the distribution function we can successively obtain

$$
\begin{aligned}
&P\{d_{W_1}(F, F^N) > 2Rt\} = P\{\int\limits_{Z_F} |F(z) - F^N(z)| dz > 2Rt\} \leqslant \\
&P\{|F(z) - F^N(z)| > 2t \quad \text{for at least one} \quad z \in R^1\} \leqslant \\
&P\{|F(z) - F^N(z)| > t \quad \text{for at least one} \quad z \in \{z_1, \ldots, z_k\}\} \leqslant \\
&\sum_{i=1}^{k} P\{|F(z_i) - F^N(z_i)| > t\} \leqslant k \exp\{-2Nt^2\}.
\end{aligned}
$$

To obtain the first assertion we have employed the inequality for large deviation (see e.g. [3] or [17]). The assertion 2 follows immediately from assertion 1. $\square$

Exponentional rate of the convergence for the Kolmogorov metric has been proven already e.g. in [3]. However, the exact forms of the corresponding inequalities for $s > 1$ and finite values N depend on unknown constants. It follows from the recalled results that the properties of the Wasserstein and the Kolmogorov metrics are very suitable (especially) in the case of $s = 1$.

Lemma 3.3. *Let $\bar{X}_F(\bar{\alpha})$, $X_F(\alpha)$ be defined by the relations (2), (3), h be an arbitrary function defined on R^n. Let, moreover, $\alpha \in (0, 1)$, $\hat{\alpha} = (\alpha, \ldots \alpha)$, $\alpha^* = (\frac{\alpha+m-1}{m}, \ldots, \frac{\alpha+m-1}{m})$. If*

1. *$\bar{X}_F(\hat{\alpha}) = \bar{X}_F(\bar{\alpha})$, where $\alpha_i = \alpha$, $i = 1, \ldots, m$,*
2. *$\bar{X}_F(\alpha^*) = \bar{X}_F(\bar{\alpha})$, where $\alpha_i = \frac{\alpha+m-1}{m}$, $i = 1, \ldots, m$,*

then

$$
\inf\{h(x) | x \in \bar{X}_F(\hat{\alpha})\} \leqslant \inf\{h(x) | x \in X_F(\alpha)\} \leqslant \inf\{h(x) | x \in \bar{X}_F(\alpha^*)\}.
$$

Proof. The assertion of Lemma 3.3 has been proven in [8]. $\square$

Example 3.4. If $m = 2$, $\alpha = 0.98$, $h(x)$ fulfils the assumption of Lemma 3.3, then

$$
\inf\{h(x) | \bar{X}_F(0.98, 0.98)\} \leqslant \inf\{h(x) | X_F(0.98)\} \leqslant \inf\{h(x) | \bar{X}_F(0.99, 0.99).
$$

The assertion of Lemma 3.3 is reasonable for approximation of the set $X_F(\alpha)$ by the set $\bar{X}_F(\bar{\alpha})$ (with suitable value $\bar{\alpha}$) only for a small value of m and of course for a suitable value α.

Lemma 3.5. *Let $s = m$ and, simultaneously, there exist functions $f_i (:= f_i(x),\ i = 1, \ldots, s)$ defined on R^n such that*

$$g_i(x, z) = f_i(x) - z_i,\ i = 1, \ldots, s, \quad z = (z_1, \ldots, z_s). \tag{11}$$

If $\alpha_i \in (0, 1)$, $\mathcal{K}_i(z_i)$ $\ i = 1, \ldots, s$ are defined by the relations

$$\begin{aligned} \mathcal{K}_i(z_i) &= \{x \in X |\, f_i(x) \leqslant z_i\}, \quad i = 1, \ldots, s, \\ k_{F_i}(\alpha_i) &= \sup\{z_i | P_{F_i}\{\omega |\, z_i \leqslant \xi_i(\omega)\} \geqslant \alpha_i\},\ i = 1, \ldots, s, \end{aligned}$$

then $\quad \bar{X}_F(\bar{\alpha}) = \bigcap\limits_{i=1}^{s} \mathcal{K}_i(k_{F_i}(\alpha_i)).$

Proof. The proof of Lemma 3.5 follows from the corresponding definitions. □

4 Main Results

To present the main results, we define the set $\mathcal{K}(z)$, $z = (z_1, \ldots, z_s)$ by the relation

$$\mathcal{K}(z) = \bigcap_{i=1}^{s} \mathcal{K}_i(z_i). \tag{12}$$

Theorem 4.1. *Let $m = s$, $\alpha \in (0, 1)$, $X \subset R^n$ be a nonempty compact set, g_0 be a uniformly continuous function on $R^n \times R^s$. Let, moreover, $X_F(\alpha)$, $\bar{X}_F(\bar{\alpha})$ be defined by the relation (2), (3). If $\hat{\alpha} = (\alpha, \ldots, \alpha)$, $\alpha^* = (\frac{\alpha+m-1}{m}, \ldots, \frac{\alpha+m-1}{m})$ and if for every $x \in X$ there exists a finite $\mathsf{E}_F g_0(x, \xi)$, then*

$$\inf\{\mathsf{E}_F g_0(x, \xi) | \bar{X}_F(\hat{\alpha})\} \leqslant \inf\{\mathsf{E}_F g_0(x, \xi) | X_F(\alpha)\} \leqslant \inf\{\mathsf{E}_F g_0(x, \xi) | \bar{X}_F(\alpha^*)\}.$$

($\bar{X}_F(\hat{\alpha})$, $\bar{X}(\alpha^*)$ are defined by Lemma 3.3.)

Proof. The assertion of Theorem 4.1 follows from the assertion of Lemma 3.3 setting $h(x) = \mathsf{E}_F g_0(x, \xi)$. □

Corollary 4.2. *Let the assumptions of Theorem 4.1 be fulfilled. If G is an s–dimensional distribution function and, moreover, for every $x \in X$ there exists a finite $\mathsf{E}_G g_0(x, \eta)$, $X_F(\alpha)$, $X_G(\alpha)$, $\bar{X}_F(\hat{\alpha})$, $\bar{X}_G(\hat{\alpha})$, $\bar{X}_F(\alpha^*)$, $\bar{X}_G(\alpha^*)$ are nonempty compact sets, then*

$$\begin{aligned} &|\inf\{\mathsf{E}_F g_0(x, \xi) | X_F(\alpha)\} - \inf\{\mathsf{E}_G g_0(x, \eta) | X_G(\alpha)\}| \leqslant \\ &|\inf\{\mathsf{E}_F g_0(x, \xi) | \bar{X}_F(\alpha^*)\} - \inf\{\mathsf{E}_F g_0(x, \xi) | \bar{X}_F(\hat{\alpha})\}| + \\ &|\inf\{\mathsf{E}_F g_0(x, \xi) | \bar{X}_F(\hat{\alpha})\} - \inf\{\mathsf{E}_G g_0(x, \eta) | \bar{X}_G(\hat{\alpha})\}| + \\ &|\inf\{\mathsf{E}_G g_0(x, \eta) | \bar{X}_G(\alpha^*)\} - \inf\{\mathsf{E}_G g_0(x, \eta) | \bar{X}_G(\hat{\alpha})\}|. \end{aligned}$$

Proof. Employing the triangular inequality we can see that the assertion of Corollary 4.2 follows immediately from the assertion of Theorem 4.1. □

4.1 Stability Results

Theorem 4.3. *Let X be a nonempty compact set, $X_F = X$ independent of F, g_0 be a uniformly continuous function on $X \times R^s$. If the assumptions of Lemma 3.1 are fulfilled, then*

$$|\inf\{\mathsf{E}_F g_0(x, \xi)|x \in X\} - \inf\{\mathsf{E}_G g_0(x, \eta)|x \in X\}| \leqslant L \sum_{i=1}^{s} \int_{-\infty}^{\infty} |F_i(z_i) - G_i(z_i)| dz_i. \tag{13}$$

Proof. The assertion follows from the assertion of Lemma 3.1. □

The assumptions under which there exist $\bar{C}_1, C > 0$ such that

$$\Delta[\bar{X}_F(\bar{\alpha}), \bar{X}_G(\bar{\alpha})] \leqslant \bar{C}_1 \sum_{i=1}^{s} |\mathcal{K}_i(k_{F_i}(\alpha_i)) - \mathcal{K}_i(k_{G_i}(\alpha_i))| \leqslant C \sum_{i=1}^{s} \sup_{z_i} |F_i(z_i) - G_i(z_i)| \tag{14}$$

were introduced in [6] and [7]. ($\Delta[\cdot, \cdot]$ denotes the Hausdorff distance in the space of closed subsets of R^n. For the definition of the Hausdorff distance see e.g. [14].)

Theorem 4.4. *Let $\alpha_i \in (0, 1)$, $i = 1, \ldots, s$, X, $\bar{X}_F(\bar{\alpha})$, $\bar{X}_G(\bar{\alpha}) \subset R^n$ be nonempty compact sets. If*

1. *there exist functions $f_i (:= f_i(x), i = 1, \ldots, s)$ defined on R^n, fulfilling the relation (11),*
2. *the assumptions of Lemma 3.1 and the relation (14) are fulfilled.*
3. *P_{F_i}, $i = 1, \ldots s$ are absolutely continuous w.r.t. the Lebesgue measure on R^1 and, moreover, there exists $\vartheta_i > 0$ such that $\bar{f}_i(z_i) > \vartheta_i > 0$, $z_i \in Z_{F_i}$, $i = 1, \ldots, s$ (the symbol $\bar{f}_i$ denotes the probability density corresponding to F_i),*
4. *g_0 is a Lipschitz function on $X \times R^s$,*

then there exist constants $L > 0$, $C > 0$ such that

$$\begin{aligned} &|\inf\{\mathsf{E}_F g_0(x, \xi)|x \in \bar{X}_F(\bar{\alpha})\} - \inf\{\mathsf{E}_G g_0(x, \eta)|x \in \bar{X}_G(\bar{\alpha})\}| \leqslant \\ &L \sum_{i=1}^{s} \int_{-\infty}^{\infty} |F_i(z_i) - G_i(z_i)| dz_i + C \sum_{i=1}^{s} \sup_{z_i \in R^1} |F_i(z_i) - G_i(z_i)|. \end{aligned} \tag{15}$$

Proof. First, it follows from the assumptions that g_0 is a Lipschitz (with respect to $\mathcal{L}_1$) norm function of $z \in R^s$ with the Lipschitz constant not depending on $x \in X$. Furthermore, following the idea of the proofs presented in [7] we can see that

$$|\inf\{\mathsf{E}_F g_0(x, \xi)|x \in \bar{X}_F(\bar{\alpha})\} - \inf\{\mathsf{E}_G g_0(x, \eta)|x \in \bar{X}_G(\bar{\alpha})\}| \leqslant$$

$$|\inf_{x \in \bar{X}_F(\bar{\alpha})} \mathsf{E}_F g_0(x, \xi) - \inf_{x \in \bar{X}_F(\bar{\alpha})} \mathsf{E}_G g_0(x, \eta)| +$$

$$|\inf_{x \in \bar{X}_F(\bar{\alpha})} \mathsf{E}_G g_0(x, \eta) - \inf_{x \in \bar{X}_G(\bar{\alpha})} \mathsf{E}_G g_0(x, \eta)|.$$

Employing the assumptions, it follows from Lemma 3.1 that

$$|\inf_{x \in \bar{X}_F(\bar{\alpha})} \mathsf{E}_F g_0(x, \xi) - \inf_{x \in \bar{X}_F(\bar{\alpha})} \mathsf{E}_G g_0(x, \eta)| \leqslant L \sum_{i=1}^{s} \int_{-\infty}^{\infty} |F_i(z_i) - G_i(z_i)| dz_i.$$

Now according to the relation (14), it is easy to see that the assertion of Theorem 4.4 is valid. □

4.2 Empirical Estimates

Evidently, the stability results can be employed for empirical estimates investigation. To this end, we replace distribution functions G, G_i, $i = 1, \ldots,$ by F^N, F_i^N, $i = 1, \ldots, s$, where F^N is an empirical distribution function determined by independent random sample $\xi^1, \ldots, \xi^N$, $N = 1, \ldots,$ corresponding to F.

Theorem 4.5. *Let X be a compact set, $X_F = X$ independent of F, g_0 be a uniformly continuous function on $X \times Z_F$. Let, moreover, P_F and g_0 fulfil the assumptions of Lemma 3.1, $\int_{-\infty}^{\infty} \sqrt{F_i(z_i)(1 - F_i(z_i))} dz_i < +\infty$, $i = 1, \ldots s$, then almost surely*

$$\begin{aligned} &\sqrt{N}|\inf\{\mathsf{E}_F g_0(x, \xi)|\, x \in X\} - \inf\{\mathsf{E}_{F^N} g_0(x, \xi)|\, x \in X\}| \leqslant \\ &\sqrt{N} \sum_{i=1}^{s} \int_{-\infty}^{\infty} |F_i^N(z_i) - F_i(z_i)| dz_i, \qquad \text{where} \\ &\sqrt{N} \sum_{i=1}^{s} \int_{-\infty}^{\infty} |F_i^N(z_i) - F_i(z_i)| dz_i \longrightarrow_d \sum_{i=1}^{s} \int_{-\infty}^{\infty} |U(F_i(z_i))| dz_i. \end{aligned} \tag{16}$$

If, moreover, Z_F is a compact set, and F_i, $i = 1, 2, \ldots, s$ are absolutely continuous with respect to Lebesgue measure in R^1, then for $t \in (0, 1)$ and $\beta \in (0, \frac{1}{2})$

$$P\{N^{\beta}|\inf\{\mathsf{E}_F g_0(x, \xi)|x \in X\} - \inf\{\mathsf{E}_{F^N} g_0(x, \xi)|x \in X\}| \geqslant t\} \longrightarrow_{N \longrightarrow \infty} 0. \tag{17}$$

Proof. The proof of the first assertion follows from the assertion of Lemma 3.1 and the relation (9). The proof of the second assertion follows from Lemma 3.2 and from the well–known properties of the exponentional function. □

Theorem 4.6. *Let* $\alpha_i \in (0, 1)$, $i = 1, \ldots, s$, $P_F \in \mathcal{M}_1(R^s)$, g_0 *be uniformly continuous on* $X \times Z_F$. *Let, moreover,* g_0 *fulfil the assumptions of Lemma 3.1, the relation (14) be fulfilled,* X, Z_F, $\bar{X}_F(\bar{\alpha})$, $\bar{X}_{F^N}(\bar{\alpha})$ *be nonempty compact sets. If*

1. *the relation (11) is fulfilled,*
2. *the assumptions 2, 3 and 4 of Theorem 4.4 are fulfilled and moreover* F_i *are increasing in some neighbourhood of* $k_{F_i}(\alpha_i)$, $i = 1, \ldots, s$,

then for every $t \in (0, 1)$ *there exist a constants* $C(t)$, $k > 0$ *such that*

$$P\{|\inf\{\mathsf{E}_F g_0(x, \xi)|x \in \bar{X}_F(\bar{\alpha})\} - \inf\{\mathsf{E}_{F^N} g_0(x, \eta)|x \in \bar{X}_{F^N}(\bar{\alpha})\}| \geqslant t\} \leqslant$$

$$C(t)\exp\{-kNt^2\} \qquad \text{and, moreover,}$$

$$P\{N^\beta|\inf\{\mathsf{E}_F g_0(x, \xi)|x \in \bar{X}_F(\bar{\alpha})\} - \inf\{\mathsf{E}_{F^N} g_0(x, \eta)|x \in \bar{X}_{F^N}(\bar{\alpha})\}| \geqslant t\}$$

$$\longrightarrow_{N \longrightarrow \infty} 0, \quad t \in (0, 1) \quad \beta \in (0, \tfrac{1}{2}).$$

Proof. The proof of the first assertion follows from the assertion of Theorem 4.4 and from the well–known results (mentioned already above) published e.g. in [3]. The proof of the second assertion follows from the first assertion, from the results on quantile empirical estimates (see e.g. [15]), from the well–known properties of exponentional function and the form of the constant $C(t)$ given by Lemma 3.2. □

References

[1] L. Dai, C. H. Chen, and J. R. Birge. Convergence properties of two–stage stochastic programming. *J. Optim. Theory Appl.*, 106: 489–509, 2000.

[2] J. Dupačová and R. J.-B. Wets. Asymptotic behaviour of statistical estimates and optimal solutions of stochastic optimization problems. *Ann. Statist.*, 16: 1517–1549, 1984.

[3] A. Dvoretzky, J. Kiefer, and J. Wolfowitz. Asymptotic minimax character of the sample distribution function and the classical multinomial estimate. *Ann. Math. Statist.*, 27: 642–669, 1956.

[4] V. Kaňková. A note on estimates in stochastic programming. *J. Comput. Math.*, 56: 97–112, 1994.

[5] V. Kaňková. A note on the relationship between distribution function estimation and estimation in stochastic programming. In: *Transactions of the 12th Prague Conference 1994*, Academia, 122–125.

[6] V. Kaňková. A note on objective functions in multistage stochastic programming problems. In: *Proceedings of the 17th IFIP TC7 Conference on System Modelling and Optimization, Prague 1995* (J. Doležal and J. Fidler, eds.), Chapman Hall, London – Glasgow – New York – Tokyo 1995, 582–589.

[7] V. Kaňková. On the stability in stochastic programming: the case of individual probability constraints. *Kybernetika*, 33: 525–546, 1997.

[8] V. Kaňková. Unemployment problem, restruktualization and stochastic programming. In: *Mathematical Methods in Economocs'99* (J. Plešinger, ed.), 151–158.

[9] V. Kaňková and M. Houda. Dependent samples in empirical estimation of stochastic programming problem. *Austrian Journal of Statistics* (to appaear).

[10] G. Ch. Pflug. Scenario tree generation for multiperiod financial optimization by optimal discretization. *Math. Program. Ser. B*, 89: 251–271, 2001.

[11] S. T. Rachev. *Probability Metrics and the Stability of Stochastic Models.* Wiley, New York, 1991.

[12] W. Römisch and R. Schulz. Stability of solutions for stochastic programs with complete recourse. *Math. of Oper. Res.*, 18: 590–609, 1993.

[13] W. Römisch. Stability of stochastic programming problems. In: *Stochastic Programming* (A. Ruszczynski and A. A. Shapiro, eds.) Handbooks in Operations Research and Managemennt Science, Vol 10, Elsevier, Amsterodam 2003, 483–554.

[14] G. Salinetti and R. J.-B. Wets. On the convergence of closed–valued measurable multifunction. *Trans. of the Americal Society*, 266: 278–289, 1981.

[15] J. R. Serfling. *Approximation Theorems of Mathematical Statistics* Wiley, New York, 1980.

[16] A. Shapiro. Quantitative stability in stochastic programming. *Math. Program.*, 67: 99–108, 1994.

[17] G. R. Shorack and J. A. Wellner. *Empirical Processes with Applications to Statistics.* Wiley, New York, 1986.

[18] M. Šmíd: On approximation of stochastic programming problems. Doctoral Thesis, Charles University Prague, Faculty of Mathematics and Physics.

[19] S. S. Valander. Calculation of the Wasserstein distance between probability distributions on the line. *Theor. Probab. Appl.*, 18: 784–786, 1973.

[20] S. Vogel. On stability in multiobjective programming – a stochastic approach. *Math. Progr.*, 56: 91–119, 1992.

[21] L. Wang and J. Wang. Limit distribution of statistical estimators for stochastic programs with dependent samples. *Z. Angew. Math. Mech.*, 79: 257–266, 1999.

Vlasta Kaňková: Academy of Sciences of the Czech Republic, Institute of Information Theory and Automation, Pod Vodárenskou věží 4, Praha 8, CZ 182 08, Czech Republic, *kankova@utia.cas.cz*

Michal Houda: Academy of Sciences of the Czech Republic, Institute of Information Theory and Automation, Pod Vodárenskou věží 4, Praha 8, CZ 182 08, Czech Republic, *houda@csmat.karlin.mff.cuni.cz*