KYBERNETIKA — VOLUME *51* (2015), NUMBER 3, PAGES 433–456

THIN AND HEAVY TAILS IN STOCHASTIC PROGRAMMING

Vlasta Kaňková and Michal Houda

Optimization problems depending on a probability measure correspond to many applications. These problems can be static (single-stage), dynamic with finite (multi-stage) or infinite horizon, single- or multi-objective. It is necessary to have complete knowledge of the "underlying" probability measure if we are to solve the above-mentioned problems with precision. However this assumption is very rarely fulfilled (in applications) and consequently, problems have to be solved mostly on the basis of data. Stochastic estimates of an optimal value and an optimal solution can only be obtained using this approach. Properties of these estimates have been investigated many times.

In this paper we intend to study one–stage problems under unusual (corresponding to reality, however) assumptions. In particular, we try to compare the achieved results under the assumptions of thin and heavy tails in the case of problems with linear and nonlinear dependence on the probability measure, problems with probability and risk measure constraints, and the case of stochastic dominance constraints. Heavy-tailed distributions quite often appear in financial problems [26] while nonlinear dependence frequently appears in problems with risk measures [22, 30]. The results we introduce follow mostly from the stability results based on the Wasserstein metric with the "underlying" $\mathcal{L}_1$ norm. Theoretical results are completed by a simulation investigation.

Keywords: stochastic programming problems, stability, Wasserstein metric, $\mathcal{L}_1$ norm, Lipschitz property, empirical estimates, convergence rate, linear and nonlinear dependence, probability and risk constraints, stochastic dominance

Classification: 90C15

1. INTRODUCTION

Let $(\Omega, \mathcal{S}, \mathbb{P})$ be a probability space, $\xi := \xi(\omega) = (\xi_1(\omega), \ldots, \xi_s(\omega))$ an s–dimensional random vector defined on $(\Omega, \mathcal{S}, \mathbb{P})$, $F := F_\xi(\cdot)$ the distribution function of ξ, and $\mathbb{P}_F$ and Z_F the probability measure and the support corresponding to F, respectively. Moreover, let $g_0 : \mathbb{R}^s \times \mathbb{R}^n \to \mathbb{R}$ be a real-valued function, $X \subset R^n$ a nonempty deterministic set, and $X_F \subset X$ its nonempty subset, in general depending on F. The operator of mathematical expectation corresponding to F will be denoted by $\mathbb{E}_F$. To further simplify the notation, we will also use the abbreviation $\mathbb{P}\{A\}$ standing for $\mathbb{P}\{\omega \in \Omega \mid \omega \in A\}$ for a random event A from $(\Omega, \mathcal{S}, \mathbb{P})$.

DOI: 10.14736/kyb-2015-3-0433

Assuming that for an $x \in X$ there exists $\mathbb{E}_F g_0(x, \xi)$, we introduce a general one-stage (static) "classical" stochastic optimization problem in the form

$$\text{to find} \quad \varphi(F, X_F) := \inf\big\{\mathbb{E}_F g_0(x, \xi) \mid x \in X_F\big\}. \tag{1}$$

The objective function in (1) linearly depends on the probability measure $\mathbb{P}_F$. We also consider the following relaxed version of the problem (with nonlinear dependence):

$$\text{to find} \quad \bar{\varphi}(F, X_F) := \inf\big\{\mathbb{E}_F \bar{g}_0(x, \xi, \mathbb{E}_F h(x, \xi)) \mid x \in X_F\big\}, \tag{2}$$

where $h : \mathbb{R}^n \times \mathbb{R}^s \to \mathbb{R}^{m_1}$ is an m_1-dimensional vector function (with components $h_1, \ldots, h_{m_1}$), and $\bar{g}_0 : \mathbb{R}^n \times \mathbb{R}^s \times \mathbb{R}^{m_1} \to \mathbb{R}$.

The type of problems represented by (2) have recently appeared in the literature (see, e. g., Ermoliev and Norkin [8]). The classical problem (1) is also covered by the problem (2) if $\bar{g}_0(x, z, y) := g_0(x, z)$ for every $y \in \mathbb{R}^{m_1}$. "Mean-Risk" problems can also be covered by the form (2), see [21, 30, 40].

In this paper we restrict our focus only to several special cases of X_F:

1. *fixed constraint set*:
$$X_F := X; \tag{3}$$

2. *individual probability constraints*:
$$X_F := X_F(p) := \bigcap_{i=1}^{s} \big\{x \in X \mid \mathbb{P}[g_i(x) \le \xi_i] \ge p_i\big\}, \tag{4}$$
where $p = (p_1, \ldots, p_s)$, $p_i \in (0, 1)$, and $g_i : \mathbb{R}^n \to \mathbb{R}$ for $i = 1, \ldots, s$;

3. *constraints with loss functions*:
$$X_F := X_F(u_0, p) = \bigcap_{i=1}^{s} \Big\{x \in X \mid \min\big\{u^i \mid \mathbb{P}[L_i(x, \xi) \le u^i] \ge p_i\big\} \le u_0^i\Big\}, \tag{5}$$
where $u_0 = (u_0^1, \ldots, u_0^s)$ and $p = (p_1, \ldots, p_s)$ with $u_0^i > 0$, $p_i \in (0, 1)$ for $i = 1, \ldots, s$. In our setting, the loss functions are defined for $i = 1, \ldots, s$ by
$$L_i(x, z) = g_i(x) - z_i$$
where $g_i : \mathbb{R}^n \to \mathbb{R}$ and $z = (z_1, \ldots, z_s) \in \mathbb{R}^s$. Such type of loss functions can appear for example in connection with an inner problem in two-stage stochastic (generally nonlinear) programming problems (for a definition of two-stage problems see, e. g., Birge and Louveaux [3]).

4. *stochastic (first and second order) dominance constraints*: given $g : \mathbb{R}^n \times \mathbb{R}^s \to \mathbb{R}$ (such that $g(x, \xi)$ is a random variable for every $x \in X$) and a random variable $Y := Y(\xi)$ with the distribution function F_Y, we define a first-order stochastic dominance constraint by
$$X_F := \big\{x \in X \mid F_{g(x,\xi)}(u) \le F_Y(u) \text{ for every } u \in \mathbb{R}\big\}, \tag{6}$$

and a second order stochastic dominance constraint by

$$X_F := \{x \in X \mid F^{(2)}_{g(x,\xi)}(u) \le F^{(2)}_Y(u) \text{ for every } u \in \mathbb{R}\}, \tag{7}$$

where, for $u \in \mathbb{R}$,

$$F^{(2)}_{g(x,\xi)}(u) := \int_{-\infty}^{u} F_{g(x,\xi)}(y)\mathrm{d}y, \qquad F^{(2)}_Y(u) := \int_{-\infty}^{u} F_Y(y)\mathrm{d}y.$$

For more information about stochastic dominance see, e. g., [40].

In applications we often have to replace the measure $\mathbb{P}_F$ by an empirical measure $\mathbb{P}_{F^N}$ determined from a random sample (not necessarily independent) corresponding to the measure $\mathbb{P}_F$. Consequently, instead of solving the problems (1) and (2), the following problems are solved:

$$\text{to find } \varphi(F^N, X_{F^N}) = \inf\{\mathbb{E}_{F^N} g_0(x,\xi) \mid x \in X_{F^N}\}, \tag{8}$$

$$\text{to find } \bar{\varphi}(F^N, X_{F^N}) = \inf\{\mathbb{E}_{F^N} \bar{g}_0(x,\xi,\mathbb{E}_{F^N} h(x,\xi)) \mid x \in X_{F^N}\}. \tag{9}$$

Solving (8) and (9) we obtain (empirical) estimates of the optimal values and optimal solutions of the problems (1) and (2) in question. The aim of this paper is to compare the properties of the above-mentioned estimates when the underlying distributions possess thin or heavy tails. We shall see that the properties of empirical estimates are more favorable in the case of thin tails, although the fundamental limit properties are mostly valid also in the case of heavy-tailed distributions.

The paper is organized as follows: first, we recall a short historical survey on empirical estimates for the classical stochastic programming problem (1) with $X_F = X$ in Section 2. In the text, we emphasize a significance of heavy-tailed distributions. Auxiliary definitions and assertions can be found in Section 3. The main results will be presented in Section 4. The paper is completed by a simulation study in Section 5, and concluded by a discussion in Section 6.

2. BRIEF HISTORICAL SURVEY

Investigation of empirical estimates in stochastic programming started in [44] for the problem of type (1) with $X_F = X$, see also [13]. It was followed by many papers, see, e. g., [4, 6, 12, 14, 17, 29, 34, 37, 38]. Consistency, convergence rate, and asymptotic distributions have been studied there under the assumptions of thin-tailed distributions and $X_F = X$. Exceptional examples introducing heavy-tailed distributions are the papers [11, 21, 31]. Nonlinear dependence on probability measure appeared, e. g., in [8, 23, 31]. However, we will demonstrate in this section that many of the original results already intrinsically cover the case of heavy-tailed distributions as well. To recall these assertions we start with the following consistency theorem.

Theorem 2.1. (Kaňková [13]) If

1. X is a compact set, $g_0(x, z)$ is a uniformly continuous bounded function on $X \times \mathbb{R}^s$,
2. $\{\xi^i\}_{i=-\infty}^{+\infty}$ is an ergodic sequence, and F^N is determined by $\{\xi^i\}_{i=1}^{N}$,

then

$$\mathbb{P}\Big\{|\varphi(F^N,\, X) - \varphi(F,\, X)| \xrightarrow[N\to\infty]{} 0\Big\} = 1.$$

Theorem 2.1 was proven under the assumption that $\{\xi^i\}_{i=-\infty}^{+\infty}$ is an ergodic sequence and g_0 is a bounded function (for the definition of the ergodic random sequence see [2]). Of course, the ergodic property covers the independent case. It is very important to note that the theorem is valid independently of the distribution tails. Observe that it also covers stable distributions.

Later we shall see that the consistency is guaranteed (under some additional assumptions) by the existence of finite first moments. Let us recall one of the essential results concerning this question in which g_0 is not assumed to be bounded.

Theorem 2.2. Let X be a nonempty compact set. If

1. $g_0(\cdot, z)$ is a continuous function for almost every $z \in Z_F$ (with respect to $\mathbb{P}_F$),
2. $g_0(x, \cdot)$ is dominated by a function integrable with respect to F,
3. $\{\xi^i\}_{i=1}^{N}, N = 1, 2, \ldots$ is an independent random sample,

then

$$P\Big\{|\varphi(F^N, X) - \varphi(F, X)| \xrightarrow[N\to\infty]{} 0\Big\} = 1.$$

P r o o f. The assertion of Theorem 2.2 follows immediately from Proposition 5.2 and Theorem 7.48 proven in [40]. □

Investigation of the convergence rate dates back to 1978 [14] when it was started by the following assertion.

Theorem 2.3. (Kaňková [14]) Let $t > 0$, X be a nonempty compact, convex set. If

1. $g_0(x, z)$ is a uniformly continuous function on $X \times Z_F$, bounded by $M' > 0$ (i. e., $|g_0(x, z)| \le M'$),
2. $g_0(x, z)$ is a Lipschitz function on X with the Lipschitz constant L' not depending on z,
3. $\{\xi^i\}_{i=1}^{N},\ N = 1, 2, \ldots$ is an independent random sample,

then there exist constants $K(t, X, L')$, $k_1(M') > 0$ such that

$$\mathbb{P}\Big\{\varphi(F, X) - \varphi(F^N, X)| > t\} \le K(t, X, L') \exp\{-Nk_1(M')t^2\Big\}.$$

The assertion of Theorem 2.3 is valid independently of the distribution tails. However, the integrand function $g_0(x, \xi)$ in the problem (1) is supposed to be bounded.

Remark 2.4. The estimation of the constants $K(t, X, L')$, $k_1(M')$ [14] gives a possibility to determine a sample size that guarantees reasonable error estimates, and to prove that

$$\mathbb{P}\{N^{\beta}|\varphi(F, X) - \varphi(F^N, X)| > t\} \xrightarrow[N \longrightarrow \infty]{} 0 \quad \text{for } \beta \in (0, \tfrac{1}{2})$$

(see [15]).

Later it was the theory of large deviations which was employed to investigate the convergence rate (see, e. g., [4, 12]). To present this, we define the moment generating function $M_{\bar{g}}$ corresponding to a function $\bar{g} : Z_F \to \mathbb{R}$ by the relation

$$M_{\bar{g}}(u) := \mathbb{E}_F e^{u[\bar{g}(\xi) - \mathbb{E}_F \bar{g}(\xi)]}, \ u \in \mathbb{R}.$$

We also denote $\|\cdot\|_2$ to be the Euclidean norm in $\mathbb{R}^n$. The following assertion was proven in [39].

Theorem 2.5. (Shapiro and Hu [39]) Let X be a nonempty compact set. If

1. for every $x \in X$ the moment generating function $M_{g_0}(u)$ is finite-valued for all u in a neighborhood of zero,
2. there exists a measurable function $\kappa \colon Z_F \to R_+$ and a constant $\gamma' > 0$ such that
$$|g_0(x', z) - g_0(x, z)| \le \kappa(z)\|x' - x\|_2^{\gamma'} \quad \text{for } z \in Z_F \text{ and } x, x' \in X,$$
3. the moment generating function $M_\kappa(u)$ of κ is finite-valued for all u in a neighborhood of zero,
4. $\{\xi^i\}_{i=1}^N$, $N = 1, 2, \ldots$ is an independent random sample,

then for any $t > 0$ there exist positive constants $\bar{C} := \bar{C}(t)$ and $\bar{\beta} := \bar{\beta}(t)$, independent of N, such that

$$\mathbb{P}\Big\{\sup_{x \in X} |\mathbb{E}_{F^N} g_0(x, \xi) - \mathbb{E}_F g_0(x, \xi)| \ge t\Big\} \le \bar{C}(t) e^{-N\bar{\beta}(t)}.$$

In Theorem 2.5, it is assumed that the function g_0 is Hölder continuous with respect to the decision vector x; the Hölder constant can depend on the random element and its moment generating function is supposed to exist. The upper bound introduced in Theorem 2.5 is, under the remaining assumptions, exponential; however, the form of functions $\bar{C}(\cdot)$, $\bar{\beta}(\cdot)$ is not explicitly specified. Consequently, the result of Theorem 2.5 cannot be employed to determine N that would guarantee an acceptable error of this estimate. An assumption similar to assumption 2 of Theorem 2.5 also appeared in [4].

Summarizing this part: while the first results formulated by Theorem 2.1, Theorem 2.3, and Remark 2.4, are valid independently of distribution tails, later results represented by Theorem 2.5 are valid only if the "underlying" distributions exhibit thin tails. This is probably due to the boundedness of the function under the operator of mathematical expectation in the former papers.

3. SOME DEFINITIONS AND AUXILIARY ASSERTIONS

In this section we recall basic notions, definitions and auxiliary assertions useful in our investigation of empirical estimates in the stochastic programming problems. First, we recall that the distribution function F_η of a random variable η (defined on $(\Omega, \mathcal{S}, \mathbb{P})$) has thin tails if there exists its finite moment generating function in a neighborhood of zero. Similarly, η is said to have a distribution function with heavy tails if its finite moment generating function does not exist (for the definition of the moment generating function see, e. g., [9]). Consequently, all moments of a thin-tailed distribution are finite; however, the existence of all finite moments does not guarantee the finite moment generating function and consequently, thin tails.

We consider several examples of distributions. The first of these is the Weibull distribution, defined by the probability density function

$$f(z) = \begin{cases} \frac{\beta}{\eta}\left(\frac{z}{\eta}\right)^{\beta-1} \exp\left\{-\left(\frac{z}{\eta}\right)^{\beta}\right\} & \text{for } z > 0, \\ 0 & \text{for } z \le 0, \end{cases}$$

where $\beta > 0$ is the shape parameter, and $\eta > 0$ the scale parameter of the distribution. According to the definition mentioned above, the Weibull distribution lies on a boundary between thin- and heavy-tailed distributions: its finite moment generating function exists only for some parameter values, but the rate of the convergence of empirical estimates remains the same as in the exponential case. A log-normal distribution presents a similar situation where the finite moment generating function no longer exists.

The situation is rather different in the case of a Pareto distribution. The type I Pareto probability density function is given by

$$f(z) = \begin{cases} \alpha C^{\alpha} z^{-(\alpha+1)} & \text{for } z \ge C, \\ 0 & \text{for } z < C, \end{cases}$$

where $\alpha > 0$ is the shape parameter, and $C > 0$ the scale parameter. For this distribution, the finite moment generating function does not exist, and neither do most of the moments (only the moments up to α exist). The convergence rate is worse for this distribution – see Section 4.

First, we recall some stability results.

3.1. Stability

Consider F, G to be two s-dimensional distribution functions with finite $\mathbb{E}_F g_0(x, \xi)$, $\mathbb{E}_G g_0(x, \xi)$ for $x \in X$, and denote $\mathcal{X}(F, X_F)$ a solution set of the problem (1). If we employ triangular inequality we can obtain

$$|\varphi(F, X_F) - \varphi(G, X_G)| \le |\varphi(F, X_F) - \varphi(G, X_F)| + |\varphi(G, X_F) - \varphi(G, X_G)|. \quad (10)$$

If, moreover, $\mathcal{X}(F, X_F)$, $\mathcal{X}(G, X_F)$, $\mathcal{X}(G, X_G)$ are singletons, then also

$$\|\mathcal{X}(F, X_F) - \mathcal{X}(G, X_G)\|_2 \le \|\mathcal{X}(F, X_F) - \mathcal{X}(G, X_F)\|_2 + \|\mathcal{X}(G, X_F) - \mathcal{X}(G, X_G)\|_2. \quad (11)$$

Evidently, relations similar to (10) and (11) are also valid in the case of problem (2).

To recall the first auxiliary assertion, let $\mathcal{P}(\mathbb{R}^s)$ denote the set of all (Borel) probability measures on $\mathbb{R}^s$ and let the system $\mathcal{M}_1^1(\mathbb{R}^s)$ be defined by the relation

$$\mathcal{M}_1^1(\mathbb{R}^s) := \left\{ \nu \in \mathcal{P}(\mathbb{R}^s) \mid \int_{\mathbb{R}^s} \|z\|_1 \mathrm{d}\nu(z) < \infty \right\}$$

where $\|\cdot\|_1$ denotes the $\mathcal{L}_1$-norm in $\mathbb{R}^s$. We introduce the following systems of assumptions:

A.1
- $g_0(x, z)$ is either a uniformly continuous function on $X \times \mathbb{R}^s$, or there exists $\varepsilon > 0$ such that $g_0(x, z)$ is a function convex on $X(\varepsilon)$ and bounded on $X(\varepsilon) \times Z_F$, where $X(\varepsilon)$ denotes the ε–neighborhood of the set X;
- $g_0(x, z)$ is, for $x \in X$, a Lipschitz function of $z \in \mathbb{R}^s$ with the Lipschitz constant L (corresponding to the $\mathcal{L}_1$ norm) not depending on x.

B.1 $\mathbb{P}_F, \mathbb{P}_G \in \mathcal{M}_1^1(\mathbb{R}^s)$ and there exists $\varepsilon > 0$ such that

- for every $x \in X(\varepsilon)$ and $z \in \mathbb{R}^s$, the function $\bar{g}_0(x, z, y)$ is a Lipschitz function of $y \in Y(\varepsilon)$ with a Lipschitz constant L_y where

  $$Y(\varepsilon) = \left\{ y \in \mathbb{R}^{m_1} \mid y = h(x, z) \text{ for some } x \in X(\varepsilon),\ z \in \mathbb{R}^s \right\},$$

  and $\mathbb{E}_F h(x, \xi)$, $\mathbb{E}_G h(x, \xi) \in Y(\varepsilon)$;
- for every $x \in X(\varepsilon)$ and $y \in Y(\varepsilon)$, there exist finite mathematical expectations $\mathbb{E}_F \bar{g}_0(x, \xi, \mathbb{E}_F h(x, \xi))$, $\mathbb{E}_F g_0^1(x, \xi, \mathbb{E}_G h(x, \xi))$, $\mathbb{E}_G g_0^1(x, \xi, \mathbb{E}_F h(x, \xi))$, and $\mathbb{E}_G g_0^1(x, \xi, \mathbb{E}_G h(x, \xi))$;
- for every $x \in X(\varepsilon)$, the functions $h_i(x, z)$, $i = 1, \ldots, m_1$ are Lipschitz functions of z with the Lipschitz constants L_h^i, corresponding to $\mathcal{L}_1$ norm,
- for every $x \in X(\varepsilon)$ and $y \in Y(\varepsilon)$, the function $\bar{g}_0(x, z, y)$ is a Lipschitz function of z with the Lipschitz constant L_z, corresponding to $\mathcal{L}_1$ norm.

B.2 $\mathbb{E}_F \bar{g}_0(x, \xi, \mathbb{E}_F h(x, \xi))$, $\mathbb{E}_G \bar{g}_0(x, \xi, \mathbb{E}_G h(x, \xi))$ are continuous functions on X.

Proposition 3.1. (Kaňková [19, 23]) Let $\mathbb{P}_F, \mathbb{P}_G \in \mathcal{M}_1^1(\mathbb{R}^s)$ and let X be a compact set. If

1. Assumption A.1 is fulfilled, then

$$|\varphi(F, X) - \varphi(G, X)| \le L \sum_{i=1}^{s} \int_{-\infty}^{+\infty} |F_i(z_i) - G_i(z_i)| \, \mathrm{d}z_i. \tag{12}$$

2. Assumptions B.1, B.2 are fulfilled, then there exist $\hat{C} > 0$ such that

$$|\bar{\varphi}(F, X) - \bar{\varphi}(G, X)| \le \hat{C} \sum_{i=1}^{s} \int_{-\infty}^{\infty} |F_i(z_i) - G_i(z_i)| \, \mathrm{d}z_i. \tag{13}$$

Proposition 3.1 reduces (from the mathematical point of view) an s-dimensional case to a one-dimensional one. Of course, a stochastic dependence between the components of the random vector is neglected there. The idea to reduce an s-dimensional case, $s > 1$ to a one-dimensional case is credited to Pflug [28], see also Šmíd [43].

Furthermore, we introduce some auxiliary assertions dealing with optimal solution sets. To this end, we recall the definition of strongly convex functions.

Definition 3.2. Let $\bar{h}(x)$ be a real-valued function defined on a nonempty convex set $\mathcal{K} \subset \mathbb{R}^n$. $\bar{h}(x)$ is said to be a strongly convex function with parameter $\rho > 0$ if for every $x^1, x^2 \in \mathcal{K}$ and $\lambda \in [0, 1]$

$$\bar{h}(\lambda x^1 + (1-\lambda)x^2) \le \lambda \bar{h}(x^1) + (1-\lambda)\bar{h}(x^2) - \lambda(1-\lambda)\rho \|x^1 - x^2\|_2^2.$$

Proposition 3.3. (Kaňková [16]) Let $\mathcal{K} \subset \mathbb{R}^n$ be a nonempty compact convex set. Let, moreover, $\bar{h}(x)$ be a continuous real-valued function defined on $\mathcal{K}$, strongly convex with a parameter $\rho > 0$. Let $x^0 = \arg\min_{x \in \mathcal{K}} \bar{h}(x)$. Then

$$\|x - x^0\|_2^2 \le \frac{2}{\rho} |\bar{h}(x) - \bar{h}(x^0)| \text{ for every } x \in \mathcal{K}.$$

To recall equivalent forms of the constraint sets X_F, we first introduce the following notation: let $F_i(z_i),\ i = 1, \ldots, s$, denote one-dimensional distributions corresponding to F and let us define

$$\begin{aligned} k_F(p) &= (k_{F_1}(p_1), \ldots, k_{F_s}(p_s)), \\ k_{F_i}(p_i) &= \sup\bigl\{z_i \mid \mathbb{P}_{F_i}[z_i \le \xi_i] \ge p_i\bigr\} \end{aligned} \tag{14}$$

for $p = (p_1, \ldots, p_s),\ p_i \in (0, 1),\ i = 1, \ldots, s$. We also denote by $\Delta[\cdot,\cdot] := \Delta_n[\cdot,\cdot]$ the Hausdorff distance on $\mathbb{R}^n$ (for its definition see [32]) .

Lemma 3.4. Let $g_i(x),\ i = 1, \ldots, s$ be real-valued continuous functions defined on $\mathbb{R}^n$. Let, moreover, $\mathbb{P}_{F_i},\ i = 1, \ldots, s$ be absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}$. Then

$$X_F = \bar{X}(k_F(p)),$$

where, for $v = (v_1, \ldots, v_s) \in \mathbb{R}^s$,

$$\bar{X}(v) := \bigcap_{i=1}^{s} \{x \in X \mid g_i(x) \le v_i\} \tag{15}$$

if X_F is defined by (4), or

$$\bar{X}(v) := \bigcap_{i=1}^{s} \{x \in X \mid g_i(x) - u_0^i| \le v_i\} \tag{16}$$

if X_F is defined by (5).

P r o o f . The first assertion has been proven in [18], the second in [23]. □

Proposition 3.5. Let X be a nonempty set. If

1. $\hat{g}_0(x)$ is a real-valued function, Lipschitz on X with a Lipschitz constant L,
2. $\bar{X}(v)$ are nonempty sets for every $v \in Z_F$,
3. there exists a constant $\hat{C} > 0$ such that for all $v^{(1)}, v^{(2)} \in Z_F$,
$$\Delta \left[\bar{X}(v^{(1)}), \bar{X}(v^{(2)})\right] \le \hat{C} \|v^{(1)} - v^{(2)}\|_2,$$

then
$$\left| \inf_{x \in \bar{X}(v^{(1)})} \hat{g}_0(x) - \inf_{x \in \bar{X}(v^{(2)})} \hat{g}_0(x) \right| \le L\hat{C} \|v^{(1)} - v^{(2)}\|_2.$$

Proposition 3.5 is a slightly modified version of Proposition 1 in [18]. The conditions under which $g_i, \ i = 1, \ldots, s$ fulfill assumption 3 of Proposition 3.5 can be found in the same paper.

To recall an equivalent form of the constraint set X_F defined by the second-order stochastic dominance (7), we introduce the following lemma.

Lemma 3.6. Let $g(x, z)$, $Y(z)$ be, for every $x \in X$, Lipschitz functions of $z \in \mathbb{R}^s$ with the Lipschitz constant L_g not depending on x. Let, moreover, $\mathbb{P}_F, \mathbb{P}_G \in \mathcal{M}_1^1(\mathbb{R}^s)$. If X_F is defined by relation (7), then

1. $[u - g(x, z)]^+$, $[u - Y(z)]^+$ are Lipschitz functions of $z \in \mathbb{R}^s$ with the Lipschitz constant L_g not depending on $u \in \mathbb{R}$, $x \in \mathbb{R}^n$;
2. $X_F = \{x \in X \mid E_F[u - g(x, \xi)]^+ \le \mathbb{E}_F[u - Y(\xi)]^+, u \in \mathbb{R}\}$;
3. for $u \in R$, $x \in X$ it holds that

$$\left|\mathbb{E}_F[u - g(x, \xi)]^+ - \mathbb{E}_G[u - g(x, \xi)]^+\right| \le L_g \sum_{i=1}^{s} \int_{-\infty}^{+\infty} |F_i(z_i) - G_i(z_i)| \mathrm{d}z_i,$$

$$\left|\mathbb{E}_F[u - Y(\xi)]^+ - \mathbb{E}_G[u - Y(\xi)]^+\right| \le L_g \sum_{i=1}^{s} \int_{-\infty}^{+\infty} |F_i(z_i) - G_i(z_i)| \mathrm{d}z_i,$$

$$x \in X_F \Longrightarrow \mathbb{E}_G[u - g(x, \xi)]^+ - \mathbb{E}_G[u - Y(\xi)]^+ \le 2L_g \sum_{i=1}^{s} \int_{-\infty}^{+\infty} |F_i(z_i) - G_i(z_i)| \mathrm{d}z_i,$$

$$x \in X_G \Longrightarrow \mathbb{E}_F[u - g(x, \xi)]^+ - \mathbb{E}_F[u - Y(\xi)]^+ \le 2L_g \sum_{i=1}^{s} \int_{-\infty}^{+\infty} |F_i(z_i) - G_i(z_i)| \mathrm{d}z_i.$$

P r o o f . The first assertion of Lemma 3.6 follows from the properties of the Lipschitz functions, the second assertion follows from relation (4.7) in [40]. The third assertion follows from the first and second assertions and from the assertion of Proposition 3.1. □

3.2. Empirical estimates

If we replace the distribution G by the empirical distribution function F^N in the conclusions of the previous section, then we can formulate auxiliary assertions concerning empirical estimates useful for the investigation of problems (1) and (2). To this end we first introduce the following system of assumptions:

A.2
- $\{\xi^i\}_{i=1}^{\infty}$ is an independent random sequence corresponding to F, and
- F^N is an empirical distribution function determined by $\{\xi^i\}_{i=1}^{N}$, $N = 1, 2, \ldots$.

A.3 $\mathbb{P}_{F_i}, i = 1, \ldots, s$ are absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}$.

A.4 For every $i \in \{1, \ldots, s\}$ there exist $\delta > 0$ and $\vartheta > 0$ such that $f_i(z_i) > \vartheta$ for every $z_i \in Z_{F_i}$ with $|z_i - k_{F_i}(p_i)| < 2\delta$.

According to Proposition 3.1, it is reasonable to first focus on the case $s = 1$ and to recall the following auxiliary assertions.

Lemma 3.7. (Shor [42]) Let $s = 1$, $\mathbb{P}_F \in \mathcal{M}_1^1(\mathbb{R}^1)$ and Assumption A.2 be fulfilled. Then

$$\mathbb{P}\Big\{ \int_{-\infty}^{\infty} |F(z) - F^N(z)| \mathrm{d}z \xrightarrow[N\to\infty]{} 0 \Big\} = 1.$$

Proposition 3.8. (Houda and Kaňková [11]) Let $s = 1$, $t > 0$ and Assumptions A.2, A.3 be fulfilled. If there exist $\beta > 0$, and $R := R(N) > 0$ defined on $\mathbb{N}$ (the set of natural numbers) such that $R(N) \xrightarrow[N\to+\infty]{} \infty$ and, moreover,

$$\begin{aligned} &N^{\beta} \int_{-\infty}^{-R(N)} F(z)\mathrm{d}z \xrightarrow[N\to\infty]{} 0, \qquad N^{\beta} \int_{R(N)}^{\infty} [1 - F(z)]\mathrm{d}z \xrightarrow[N\to\infty]{} 0, \\ &2NF(-R(N)) \xrightarrow[N\to\infty]{} 0, \qquad 2N[1 - F(R(N))] \xrightarrow[N\to\infty]{} 0, \\ &\left(\frac{12N^{\beta}R(N)}{t} + 1\right) \exp\left\{-2N\left(\frac{t}{12N^{\beta}R(N)}\right)^2\right\} \xrightarrow[N\longrightarrow\infty]{} 0, \end{aligned} \tag{17}$$

then

$$\mathbb{P}\left\{ N^{\beta} \int_{-\infty}^{\infty} \left|F(z) - F^N(z)\right| \mathrm{d}z > t \right\} \xrightarrow[N\to\infty]{} 0. \tag{18}$$

According to the assertion of Proposition 3.8 and to old results of Dvoretzky–Kiefer–Wolfowitz (see [7]) we can recognize that fulfillment of relation (18) depends on the distribution tails. To introduce the corresponding assertions, we first recall one well-known result concerning distribution tails.

Lemma 3.9. (Gut [9]) Let $r > 0$, $s = 1$. Suppose that ξ is a non-negative random variable with $\mathbb{E}_F \xi^r < +\infty$. Then

$$z^r \mathbb{P}\{\xi > z\} \to 0 \quad \text{as } z \to +\infty.$$

The above recalled Lemma 3.9 is a basis for the next proposition.

Proposition 3.10. (Houda and Kaňková [11]) Let $s = 1$, $t > 0$, $r > 0$, and the assumptions A.2, A.3 be fulfilled. Let, moreover, ξ be a random variable such that $\mathbb{E}_F |\xi|^r < \infty$. If a constant $\beta > 0$ fulfills the inequalities

$$0 < \beta < \frac{1}{2} - \frac{1}{r},$$

then

$$\mathbb{P}\left\{ N^\beta \int\limits_{-\infty}^{\infty} |F(z) - F^N(z)| \mathrm{d}z > t \right\} \xrightarrow[N \to \infty]{} 0.$$

P r o o f . The assertion of Proposition 3.10 follows immediately from Theorem 19 [22]. □

The value of the convergence rate $\beta := \beta(r)$ introduced in Proposition 3.10 depends on the existence of finite absolute moments. Evidently, it is seen that the convergence rate $\beta \in (0, 1/2)$ is the same for the distributions with exponential tails as for the Weibull or log-normal distribution. In fact, for nearly every distribution for which all absolute moments exist and are finite, the convergence rate exhibits this behavior − that is, it is also true for some distributions with heavy tails. However, the convergence rate β is much smaller if only some of absolute moments are finite (as in the case of a Pareto distribution, where only moments up to a certain $r > 0$ exist).

It follows from the results published in [41] that Pareto tails can approximate the heavy tails of the continuous stable distributions (it means of all stable distributions with exception of the normal distribution). The class of univariate stable distributions can be considered a "generalization" of the class of normal distributions. There are four known equivalent definitions of the stable distributions. We recall one of them (see [36]); a direct relationship to the normal distribution can be seen from this definition. A random variable ξ is said to have a (univariate) stable distribution if, for all positive numbers A and B, there are a positive number C and a real number D such that

$$A\xi_1 + B\xi_2 =_d C\xi + D,$$

where ξ_1 and ξ_2 are independent copies of ξ, and " $=_d$ " denotes equality in distribution.

The stable distributions are characterized by four parameters: index of stability $\nu \in (0, 2]$, which says how heavy the tails of the distributions are; the scale parameter $\sigma \geq 0$; the skewness parameter $b \in [-1, 1]$ (usually denoted β, but we use b here to avoid the notation conflict); and the shift parameter $\mu \in \mathbb{R}$. The stable distribution is Gaussian when $\nu = 2$, and in this case σ is proportional to the standard deviation, b can be taken as zero and μ is the mean. It is known that the probability densities of stable random variables exist and are continuous but, with a few exceptions, they are not known in a closed form. Moreover, it is known that the finite first moment exists if $\nu > 1$.

Unfortunately, we cannot obtain (by this approach) any results similar to Proposition 3.10 if there exist $\mathbb{E}_F|\xi|^r$ only for $r < 2$. It is known (see, e. g., [24]) that this is an example of heavy-tailed stable distributions. However, to obtain at least a weaker result for this case we recall the result of [1].

Proposition 3.11. (Bario et al. [1]) Let $s = 1$, $\{\xi^i\}_{i=1}^N$, $N = 1, 2, \ldots$ be a sequence of independent random values corresponding to a stable distribution F with the index of stability $\nu \in (1, 2)$. Then the sequence

$$\frac{N}{N^{1/\nu}} \int_{-\infty}^{\infty} |F^N(z) - F(z)| \mathrm{d}z, \; N = 1, 2, \ldots \tag{19}$$

is stochastically bounded if and only if

$$\sup_{t>0} t^{\nu} \, \mathbb{P}\{|\xi| > t\} < \infty. \tag{20}$$

The assertion of Proposition 3.11 follows from Theorem 2.2 [1]. According to the definition of the stochastically bounded random sequences it holds (under the validity of relation (20)) that

$$\lim_{M \to \infty} \sup_N \mathbb{P} \left\{ \frac{N}{N^{1/\nu}} \int_{-\infty}^{\infty} |F(z) - F^N(z)| \mathrm{d}z > M \right\} = 0. \tag{21}$$

To complete this Section, we recall the assertion corresponding to the case $\beta = \frac{1}{2}$.

Proposition 3.12. (Houda [10]) Let the assumptions A.2 and A.3 be fulfilled, $s = 1$, and $\mathbb{P}_F \in \mathcal{M}_1^1(\mathbb{R})$. If

$$\int_{-\infty}^{+\infty} \sqrt{F(z)(1 - F(z)} \mathrm{d}z < +\infty,$$

then

$$\int_{-\infty}^{+\infty} \sqrt{N} |F^N(z) - F(z)| \mathrm{d}z \to_d \int_{-\infty}^{+\infty} |\mathcal{U}(F(z))| \mathrm{d}z,$$

where $\mathcal{U}$ denotes the Brownian bridge.

Furthermore, we recall some auxiliary assertions useful for "empirical" estimates of the constraint set X_F fulfilling (4) and (5).

Lemma 3.13. (Kaňková [22]) Let $s = 1$, $p \in (0, 1)$. If the assumptions A.2, A.3 and A.4 are fulfilled, and $0 < t' < \delta$, then

$$\mathbb{P}\{|k_{F^N}(p) - k_F(p)| > t'\} \leq 2 \exp\{-2N(\vartheta t')^2\}, \quad N \in \mathbb{N},$$

($k_F(p), k_{F^N}(p)$ are defined by the relation (14)).

4. EMPIRICAL ESTIMATES IN OPTIMIZATION PROBLEMS

4.1. The case $X_F = X$

First, according to Proposition 3.1 and Lemma 3.7, we can obtain the following consistency result for optimal values of stochastic optimization problems.

Theorem 4.1. Let $\mathbb{P}_F \in \mathcal{M}_1^1(\mathbb{R}^s)$, X be a compact set.

1. If the assumptions A.1, A.2 and A.3 are fulfilled, then
$$\mathbb{P}\{|\varphi(F^N, X) - \varphi(F, X)| \xrightarrow[N\to\infty]{} 0\} = 1;$$
2. if the assumptions B.1, B.2, A.2 and A.3 are fulfilled, then
$$\mathbb{P}\{|\bar{\varphi}(F^N, X) - \bar{\varphi}(F, X)| \xrightarrow[N\to\infty]{} 0\} = 1.$$

According to Theorem 4.1 we can see that $\varphi(F^N, X)$, $\bar{\varphi}(F^N, X)$ are consistent estimates of $\varphi(F, X)$ and $\bar{\varphi}(F, X)$ if the underlying distributions have finite first moments (and under the remaining assumptions mentioned above). It means that these estimates are also consistent for the heavy-tailed distributions (if there exist first absolute moments); for example, for the stable distributions with the index of stability $\nu \in (1, 2]$.

The next result deals with the convergence rate.

Theorem 4.2. (Houda, Kaňková [11, 23]) Consider an $r > 2$ for which $\mathbb{E}_{F_i}|\xi_i|^r < +\infty$, $i = 1, \ldots, s$. Let, moreover, $\hat{\beta} = 1/2 - 1/r$.

1. If the assumptions A.1, A.2, A.3 are fulfilled, $\mathbb{P}_F \in \mathcal{M}_1^1(\mathbb{R}^s)$, $t > 0$, X is a compact set, then
$$\mathbb{P}\big\{N^\beta|\varphi(F,\, X) - \varphi(F^N, X)| > t\big\} \xrightarrow[N\to\infty]{} 0 \quad \text{for} \quad 0 < \beta < \hat{\beta}; \tag{22}$$
2. if the assumptions A.2, A.3. B.1 and B.2 are fulfilled, $\mathbb{P}_F \in \mathcal{M}_1^1(\mathbb{R}^s)$, $t > 0$, X is a compact set, then
$$\mathbb{P}\big\{N^\beta|\bar{\varphi}(F, X) - \bar{\varphi}(F^N, X)| > t\big\} \xrightarrow[N\to\infty]{} 0 \quad \text{for} \quad 0 < \beta < \hat{\beta}. \tag{23}$$

Generally, the convergence rate $\hat{\beta} := \hat{\beta}(r)$ introduced in Theorem 4.2 depends on the existence of finite absolute moments. In particular, the following relation is valid:
$$\hat{\beta}(r) \xrightarrow[r\to+\infty]{} \frac{1}{2}, \qquad \hat{\beta}(r) \xrightarrow[r\to 2^+]{} 0.$$
Consequently, relations (22) and (23) hold with $\beta \in (0, 1/2)$ not only for the distributions with exponential tails, but also for every distribution for which all moments exist and are finite (that is, Weibull or log-normal distributions). The convergence rates become worse for distributions like the Pareto one (see, e. g., [11]) if only some moments (for a certain $r > 2$) are finite. For a more detailed analysis see the commentary following Proposition 3.10.

The following weaker assertion covering the stable distribution with $\nu \in (1, 2)$ can be formulated.

Theorem 4.3. (Kaňková [22]) Let the assumptions A.2, A.3 be fulfilled, $\mathbb{P}_F \in \mathcal{M}_1^1(\mathbb{R}^s)$, $M > 0$, and X be a compact set. Let the one-dimensional components ξ_i, $i = 1, \ldots, s$ of the random vector ξ have the stable distribution functions F_i with the indices of stability $\nu_i \in (1, 2)$ and fulfill the relations

$$\sup_{t>0} t^{\nu_i}\, \mathbb{P}\{|\xi_i| > t\} < \infty, \quad i = 1, \ldots, s.$$

Then

1. if the assumption A.1 is fulfilled, it holds that

$$\lim_{M\to\infty} \sup_N \mathbb{P}\left\{ \frac{N}{N^{1/\nu}} |\varphi(F^N, X) - \varphi(F, X)| > M \right\} = 0 \tag{24}$$

with $\nu = \min(\nu_1, \ldots, \nu_s)$;

2. if the assumptions B.1 and B.2 are fulfilled, it holds that

$$\lim_{M\to\infty} \sup_N \mathbb{P}\left\{ \frac{N}{N^{1/\nu}} |\bar{\varphi}(F^N, X) - \bar{\varphi}(F, X)| > M \right\} = 0 \tag{25}$$

with $\nu := \min\{\nu_1, \ldots, \nu_s\}$.

Let us assume that the assumptions of Theorem 4.3 are fulfilled and define $\beta := \beta(\nu) = 1 - 1/\nu$ (that corresponds to "convergence rate"). Then $\beta(\nu)$ is an increasing function of ν and

$$\lim_{\nu\to 1^+} \beta(\nu) = 0, \quad \lim_{\nu\to 2^-} \beta(\nu) = \frac{1}{2}.$$

It will be seen in Section 5 that the convergence rate is much worse for a stable distribution with an index of stability $\nu \in (1, 2)$.

4.2. The case of probability and risk constraints

In this subsection we admit a constraint set depending on the probability measure. In particular, we assume that one of the relations (4) or (5) is fulfilled. We prove the following theorem.

Theorem 4.4. Let the assumptions A.2, A.3, and A.4 be fulfilled, $p = (p_1, \ldots, p_s)$, $u_0 = (u_0^1, \ldots, u_0^s)$, $p_i \in (0, 1)$, $u_0^i > 0$, $i = 1, \ldots, s$, and $t > 0$. Let, moreover, $g_0(x, z)$ be a real-valued Lipschitz function on X with the Lipschitz constant L' not depending on $z \in Z_F$. If

1. X_F is defined by relation (4) as $X_F = X_F(p)$, or by relation (5) as $X_F = X_F(u_0, p)$,

2. for every $v \in Z_F$, $\bar{X}(v)$ are nonempty sets and, moreover, there exists a constant $\hat{C} > 0$ such that

$$\Delta\big[\bar{X}(v^{(1)}), \bar{X}(v^{(2)})\big] \le \hat{C} \|v^{(1)} - v^{(2)}\|_2 \quad \text{for } v^{(1)}, v^{(2)} \in Z_F,$$

3. there exists $\beta \in (0, 1/2)$ such that

$$\mathbb{P}\{N^{\beta}|\varphi(F, X_F) - \varphi(F^N, X_F)| > t\} \xrightarrow[N\to\infty]{} 0, \tag{26}$$

then

$$\mathbb{P}\Big\{N^{\beta}| \inf_{\bar{X}(k_F(p))} \mathbb{E}_F g_0(x, \xi) - \inf_{\bar{X}(k_{F^N}(p))} \mathbb{E}_{F^N} g_0(x, \xi)| > t\Big\} \xrightarrow[N\to\infty]{} 0.$$

If, moreover, $g_0(x, z)$ is, for every $z \in Z_F$, a strongly convex function of $x \in X$ with a parameter $\rho > 0$, then also

$$\mathbb{P}\{N^{\beta}\|\mathcal{X}(F, X_F) - \mathcal{X}(F^N, X_{F^N})\|_2^2 > t\} \xrightarrow[N\to\infty]{} 0$$

(where $\bar{X}(v)$ is defined by relation (15) or (16) according to the definition of X_F; and $\Delta[\cdot, \cdot]$ denotes the Hausdorff distance, see p. 440).

Proof. Let F^N be determined by $\{\xi^i\}_{i=1}^N$. First, under the assumptions of the theorem, $\mathbb{E}_{F^N} g_0(x, \xi)$ is a Lipschitz function of $x \in X$ with the Lipschitz constant L' not depending on $\xi^1, \dots, \xi^N$; consequently also not depending on $\omega \in \Omega$. According to Proposition 3.5 and Lemma 3.4 there exists C' such that, successively,

$$\left| \inf_{\bar{X}(k_F(p))} \mathbb{E}_{F^N} g_0(x, \xi) - \inf_{\bar{X}(k_{F^N}(p))} \mathbb{E}_{F^N} g_0(x, \xi) \right| \le L'C'\|k_F(p) - k_{F^N}(p)\|_2,$$

$$\mathbb{P}\Big\{ \inf_{\bar{X}(k_F(p))} \mathbb{E}_{F^N} g_0(x, \xi) - \inf_{\bar{X}(k_{F^N}(p))} \mathbb{E}_{F^N} g_0(x, \xi)| > t\Big\} \le \mathbb{P}\big\{L'C'\|k_{F^N}(p) - k_F(p)\|_2 \ge t\big\}.$$

Employing now this last inequality, the inequality (26), Lemma 3.13 and properties of the exponential function we obtain the first assertion of Theorem 4.4. If, moreover, $g_0(\cdot, z)$ is strongly convex on X with a parameter $\rho > 0$, then employing relation (11), Proposition 3.3 we obtain the second assertion. □

The following result follows immediately from Theorem 4.4.

Corollary 4.5. Let the assumptions of Theorem 4.4 be valid and let a function $\hat{g}_0(x)$ defined on X be such that $g_0(x, z) = \hat{g}_0(x)$ for $x \in X$, $z \in Z_F$. Then the assertions of Theorem 4.4 are valid with $\beta \in (0, 1/2)$.

It follows from Corollary 4.5 that, if the objective function does not depend on the probability measure, the convergence rate $\beta \in (0, 1/2)$ (under very general assumptions) is independent of the underlying distribution function. Consequently, this assertion is also valid for all heavy-tailed distributions, including the stable case.

4.3. The case of second order stochastic dominance constraints

To consider the constraint set X_F fulfilling the relation (7), or if $g(x, \xi) \in \mathcal{M}_1^1(\mathbb{R})$ for $x \in X$, equivalently the relation

$$X_F := X_F^0 := \{x \in X \mid \mathbb{E}_F[u - g(x, \xi)]^+ - \mathbb{E}_F[u - Y(\xi)]^+ \le 0,\ u \in \mathbb{R}\},$$

let us first define, for $\varepsilon > 0$, a modified constraint set X_F^ε by the relation

$$X_F^\varepsilon = \{x \in X \mid \mathbb{E}_F[u - g(x, \xi)]^+ - \mathbb{E}_F[u - Y(\xi)]^+ \le \varepsilon,\ u \in \mathbb{R}\}. \tag{27}$$

We prove the following assertion.

Proposition 4.6. Let $\mathbb{P}_F \in \mathcal{M}_1^1(\mathbb{R}^s)$, $g(x, z)$ be a Lipschitz function of $z \in Z_F$ with the Lipschitz constant not depending on $x \in X$, and $\delta > 0$, then

$$X_{F^N}^{\delta - \varepsilon(N)} \subset X_F^\delta \subset X_{F^N}^{\delta + \varepsilon(N)} \quad \text{with} \quad \varepsilon(N) = 2L_g \sum_{i=1}^{s} \int_{-\infty}^{+\infty} |F_i(z_i) - F_i^N(z_i)| \mathrm{d}z_i.$$

If, moreover, the assumptions A.1, A.2, and A.3 are fulfilled, then also

$$\varphi(F, X_{F^N}^{\varepsilon(N)}) \le \varphi(F, X_F^0) \le \varphi(F, X_{F^N}^{-\varepsilon(N)}),$$
$$\varphi(F^N, X_{F^N}^{\varepsilon(N)}) \le \varphi(F^N, X_F^0) \le \varphi(F^N, X_{F^N}^{-\varepsilon(N)}).$$

Proof. Employing the assertion of Lemma 3.6 and setting $G := F^N$ we obtain

$$\left|\mathbb{E}_{F^N}[u - g(x, \xi)]^+ - \mathbb{E}_{F^N}[u - Y(\xi)]^+ - \mathbb{E}_F[u - g(x, \xi)]^+ + \mathbb{E}_F[u - Y(\xi)]^+\right|$$
$$\le 2L_g \sum_{i=1}^{s} \int_{-\infty}^{+\infty} |F_i(z_i) - F_i^N(z_i)| \mathrm{d}z_i,\ u \in \mathbb{R};$$

hence also

$$x \in X_F^\delta \Longrightarrow x \in X_{F^N}^{\delta + \varepsilon(N)}, \qquad x \in X_{F^N}^{\delta - \varepsilon(N)} \Longrightarrow x \in X_F^\delta.$$

Now we can already see that the first assertion of Proposition 4.6 follows from relation (27). The second assertion follows from the first one and from the properties of infimum. □

We can even prove a stronger assertion.

Theorem 4.7. Let $\mathbb{P}_F \in \mathcal{M}_1^1(\mathbb{R}^s)$, $t > 0$, and X be a compact set, with Assumptions A.1, A.2 and A.3 fulfilled. If

1. • $g(x, z)$ is a Lipschitz function of $z \in Z_F$ with the Lipschitz constant not depending on $x \in X$,

- $g_0(x, z)$ is a Lipschitz function of $x \in X$ with the Lipschitz constant L' not depending on $z \in Z_F$,

2. there exists $\varepsilon_0 > 0$ such that X_F^{ε} are nonempty compact sets for every $\varepsilon \in [-\varepsilon_0, \varepsilon_0]$, and, moreover, there exists a constant $\hat{C} > 0$ such that

$$\Delta[X_F^{\varepsilon},\ X_F^{\varepsilon'}] \le \hat{C}|\varepsilon - \varepsilon'| \quad \text{for } \varepsilon, \varepsilon' \in [-\varepsilon_0, \varepsilon_0],$$

3. for some $r > 2$ it holds that $\mathbb{E}_{F_i}|\xi_i|^r < +\infty$, $i = 1, \ldots, s$ and a constant β fulfills the inequality

$$0 < \beta < \frac{1}{2} - \frac{1}{r},$$

then

$$\mathbb{P}\big\{N^{\beta}|\varphi(F, X_F^0) - \varphi(F^N, X_{F^N}^0)| > t\big\} \xrightarrow[N \to +\infty]{} 0. \tag{28}$$

P r o o f. Setting $G := F^N$ in relation (10) we obtain

$$|\varphi(F, X_F^0) - \varphi(F^N, X_{F^N}^0)| \le |\varphi(F, X_F^0) - \varphi(F^N, X_F^0)| + |\varphi(F^N, X_F^0) - \varphi(F^N, X_{F^N}^0)|. \tag{29}$$

According to assumption 2, X_F^0 is a nonempty compact set. Consequently it follows from Theorem 4.2 that

$$\mathbb{P}\{N^{\beta}|\varphi(F, X_F^0) - \varphi(F^N, X_F^0)| > t\} \xrightarrow[N \to +\infty]{} 0. \tag{30}$$

It follows from Proposition 4.6 that for every $\delta \ge 0$

$$X_{F^N}^{\delta - \varepsilon(N)} \subset X_F^{\delta} \subset X_{F^N}^{\delta + \varepsilon(N)} \quad \text{with} \quad \varepsilon(N) = 2L_g \sum_{i=1}^{s} \int_{-\infty}^{+\infty} |F_i(z_i) - F_i^N(z_i)| \mathrm{d}z_i. \tag{31}$$

Since the Hausdorff distance is a metric in the space of compact subsets of $\mathbb{R}^n$ (see, e. g., [35]) we can obtain that

$$\Delta[X_F^0, X_{F^N}^0] \le \Delta[X_F^0, X_F^{-\varepsilon(N)}] + \Delta[X_F^{-\varepsilon(N)}, X_{F^N}^0]. \tag{32}$$

According to (31) we can obtain that also

$$X_F^{-\varepsilon(N)} \subset X_{F^N}^0 \subset X_F^{\varepsilon(N)} \quad \text{with} \quad \varepsilon(N) = 2L_g \sum_{i=1}^{s} \int_{-\infty}^{+\infty} |F_i(z_i) - F_i^N(z_i)| \mathrm{d}z_i \tag{33}$$

and, consequently, according to the definition of the Hausdorff distance, assumption 2, relation (32) and definition (27) we can successively obtain that

$$\begin{aligned} \Delta\big[X_F^{-\varepsilon(N)}, X_{F^N}^0\big] \le \Delta\big[X_F^{-\varepsilon(N)}, X_F^{\varepsilon(N)}\big] &\le C''\varepsilon(N) \quad \text{for some } C'' > 0, \quad \varepsilon(N) < \varepsilon_0, \\ \Delta\big[X_F^0, X_{F^N}^0\big] &\le D'\varepsilon(N) \quad \text{for some } D' > 0, \quad \varepsilon(N) < \varepsilon_0. \end{aligned}$$

Furthermore, since it follows from assumption 1 that $\mathbb{E}_{F^N} g_0(x,\, \xi)$ is a Lipschitz function of $x \in X$ with the Lipschitz constant not depending on $\xi^1, \dots, \xi^N$ (and consequently not depending on $\omega \in \Omega$), employing a slightly modified assertion of Proposition 1 of [18] (see also Proposition 3.5) and the last inequalities we obtain that

$$|\varphi(F^N, X^0_F) - \varphi(F^N, X^0_{F^N})| \le \bar{D}\varepsilon(N) \quad \text{if} \quad \varepsilon(N) < \varepsilon_0$$

for some $\bar{D} > 0$.

Consequently, it follows (successively) from the properties of probability measure for $t > 0$

$$\begin{aligned} &\mathbb{P}\{N^\beta|\varphi(F^N, X^0_F) - \varphi(F^N, X^0_{F^N})| > t\} \\ &\qquad = \mathbb{P}\{N^\beta|\varphi(F^N, X^0_F) - \varphi(F^N, X^0_{F^N})| > t \ \wedge \ \varepsilon(N) \ge \varepsilon_0\} \\ &\qquad + \mathbb{P}\{N^\beta|\varphi(F^N, X^0_F) - \varphi(F^N, X^0_{F^N})| > t \ \wedge \ \varepsilon(N) \le \varepsilon_0\} \\ &\qquad\qquad \le \mathbb{P}\{\varepsilon(N) \ge \varepsilon_0\} + \mathbb{P}\{N^\beta \bar{D}\varepsilon(N) > t\}. \end{aligned}$$

Now already it follows from Lemma 3.7, Proposition 3.10 and definition of $\varepsilon(N)$ that

$$P\{N^\beta|\varphi(F, X^0_F) - \varphi(F^N, X^0_{F^N}) > t\} \xrightarrow[N\to\infty]{} 0. \tag{34}$$

The assertion of Theorem 4.7 follows now from the relations (30), (31) and (34). □

Theorem 4.7 deals with the second-order stochastic dominance constraints. In particular, it is shown that the convergence rate (of the empirical estimates of the optimal value) is determined by the distribution's tails – in a way similar to deterministic constraint set X.

5. SIMULATION STUDY

To illustrate convergence properties treated in previous sections, we provide a simple simulation study. In particular, we compute values for the integrated empirical process (IEP)

$$N^\beta \int_{-\infty}^{\infty} |F(z) - F^N(z_i)| \mathrm{d}z \tag{35}$$

for different distributions, namely for

- the standard normal distribution $N(0; 1)$,
- the Pareto I distribution with scale parameter $C = 1$ and different values of shape parameter α (from 3 to 1.8),
- a symmetric stable distribution with scale parameter $\gamma = 1$, location parameter $\delta = 0$, and a selected value $\alpha = 1.8$ of tail parameter,

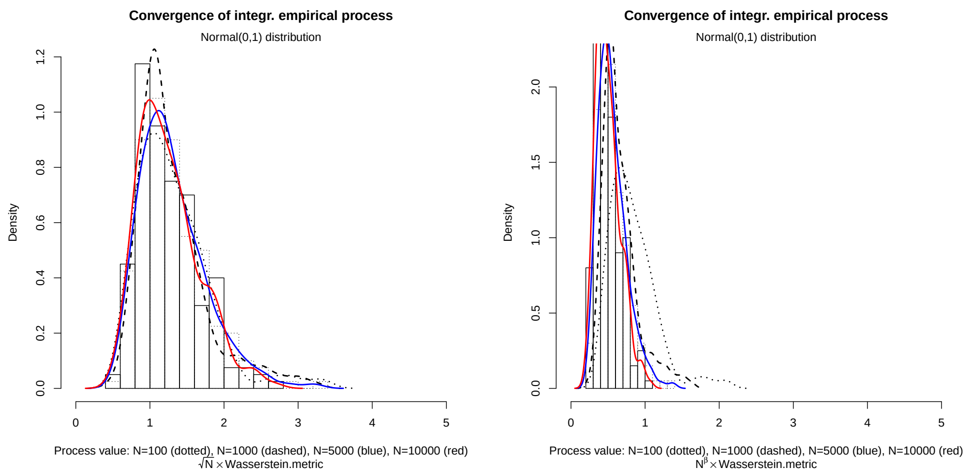

Fig. 1. Normal distribution, IEP with $\beta = 1/2$ and $2/5$.

and for two different values of β, namely $1/2$ and $2/5$. We select $N = 100, 1000, 5000,$ and 10000, respectively, as sample sizes. Each simulation is repeated 200 times to get a kernel estimator of the resulting density. The convergence properties of the IEP are particularly interesting for such problems of stochastic programming where the stability of optimal values is driven by the Wasserstein distance, see, e. g., [10, 20].

1. Standard Normal Distribution. According to Propositions 3.10 and 3.12, the IEP converge to zero if $\beta < \frac{1}{2}$, and to an integrated Brownian bridge of the standard normal distribution function, if $\beta = \frac{1}{2}$. We provide normal distribution results as a benchmark for the other computational results. The convergence is illustrated in Figure 1.

2. Pareto distribution. The shape parameter α is the largest available moment of the considered distribution. If $\alpha \to +\infty$ then the IEP converges to zero for some $\beta(\alpha) \in (0; \frac{1}{2})$, approaching to $\frac{1}{2}$. If $\alpha \to 2^-$ then the tails of IEP converge to zero for a certain $\beta(\alpha) \in (0; \frac{1}{2})$, approaching $\frac{1}{2}$. The convergence is illustrated in Figures 2 (for $\beta = 1/2$) and 3 (for $\beta = 2/5$). One can notice poor convergence properties in the case of small values of α (represented here by $\alpha = 1.8$) – the limiting distribution is not clearly identified even after 10,000 samples in that case.

3. Symmetric stable distribution. The tail parameter α is again the largest available moment of this distribution. If $\alpha = 2$ then it is the standard normal distribution. The general convergence properties follow the same rules as for the Pareto distribution. Fourth of the images in Figures 2 and 3 represents the convergence of the IEP with stable distribution with $\alpha = 1.8$.

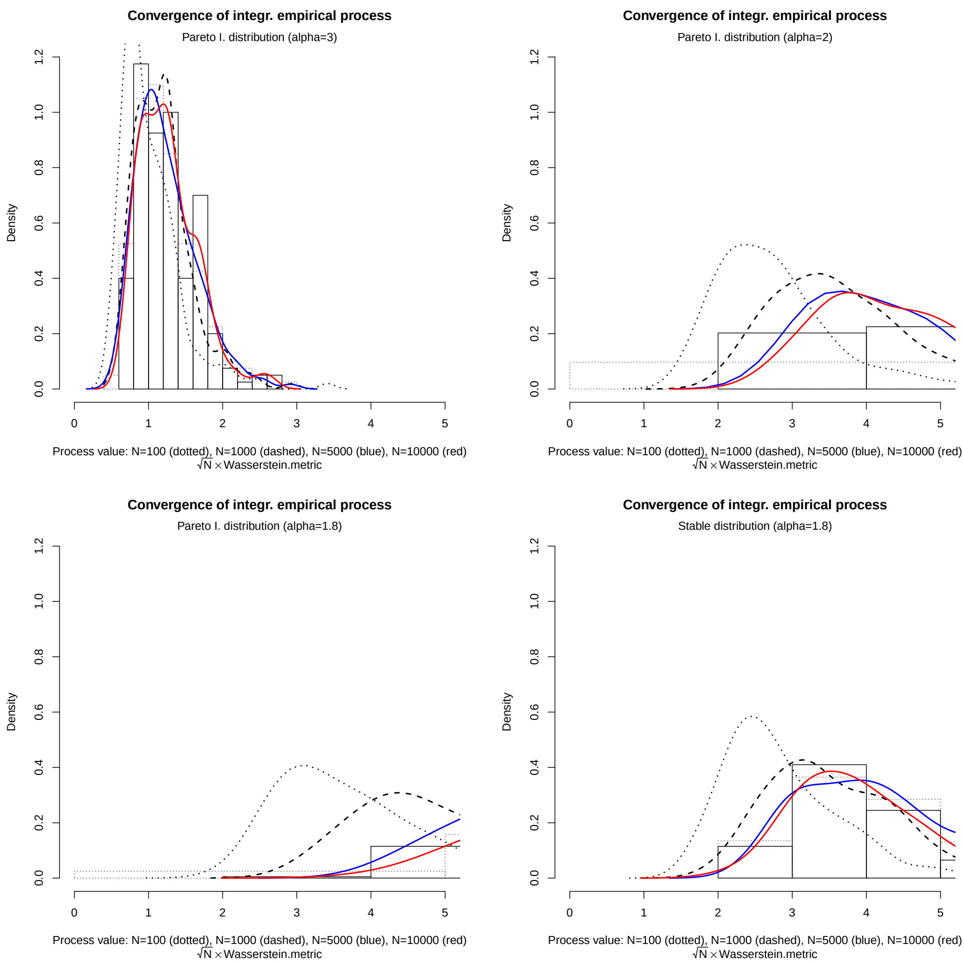

Fig. 2. Pareto ($\alpha = 3.0, 2.0, 1.8$) and stable ($\alpha = 1.8$) distributions, IEP with $\beta = 1/2$.

6. CONCLUSION – DISCUSSION

The aim of the paper is to summarize and to compare the rates of convergence for empirical estimates in stochastic optimization problems from the point of view of thin- and heavy-tailed distributions. The paper covers the case of a deterministic constraint set, a constraint set depending on a probability measure, and a constraint set defined through stochastic second-order dominance. In spite of the fact that the introduced convergence properties also cover the stable distributions in the case of probability and risk constraints, the situation is more complicated in the case of stochastic dominance constraints. First, it is possible to follow the approach of Dentcheva and

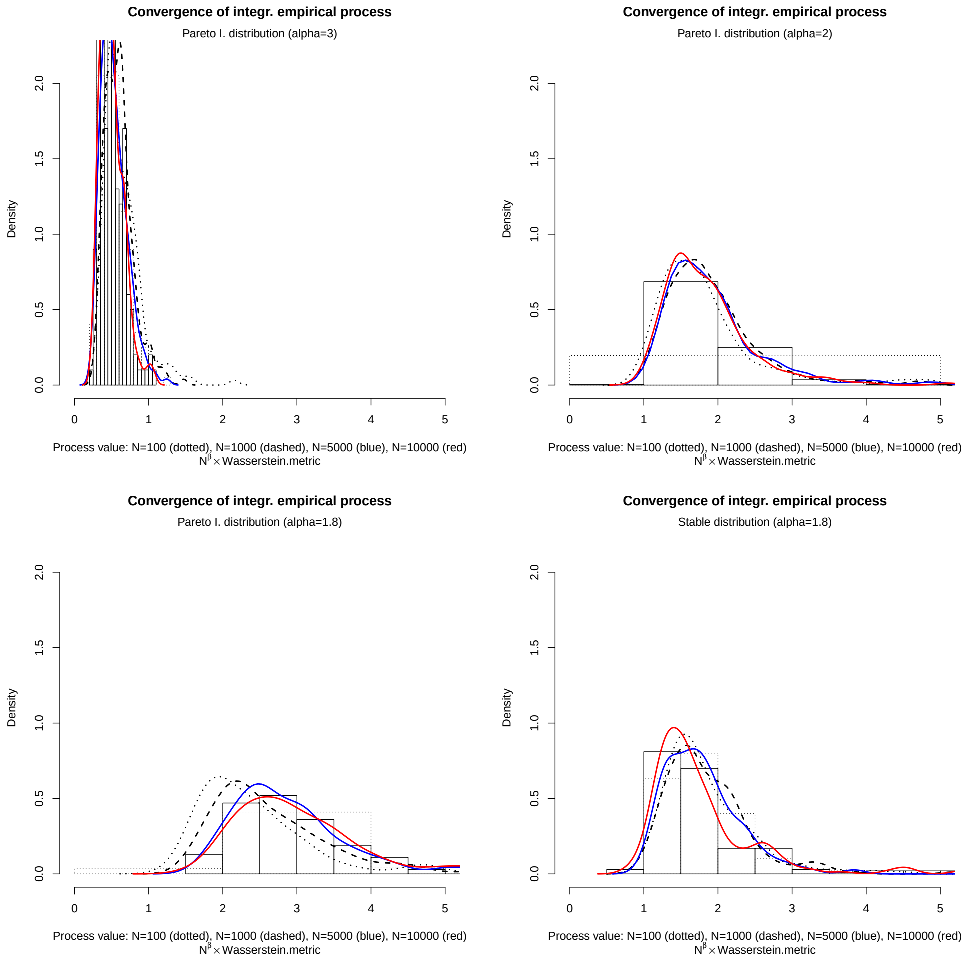

Fig. 3. Pareto ($\alpha = 3.0, 2.0, 1.8$) and stable ($\alpha = 1.8$) distributions, IEP with $\beta = 2/5$.

Ruszczyński (see [5]) with relaxing the constraints set; in particular, they propose replacing the constraint set (7) by

$$X_F = \left\{x \in X \mid \mathbb{E}_F[u - g(x,\xi)]^+ \leq \mathbb{E}_F[u - Y(\xi)]^+,\ u \in [a,b]\right\}$$

for some $a, b \in \mathbb{R}$. A question remains how to select the constants a, b in the case of stable distributions. However, the main question arises if the constraint set is nonempty. A detailed investigation in this direction still remains open.

ACKNOWLEDGMENT

This research was supported by the Czech Science Foundation under Grant No. 13-14445S.

(Received June 24, 2014)

REFERENCES

[1] E. Barrio, E. Giné, and E. Matrán: Central limit theorems for a Wasserstein distance between empirical and the true distributions. Ann. Probab. *27* (1999), 2, 1009–1071. DOI:10.1214/aop/1022677394

[2] P. Billingsley: Ergodic Theory and Information. John Wiley and Sons, New York 1965.

[3] J. R. Birge and F. Louveaux: Introduction in Stochastic Programming. Springer, Berlin 1992.

[4] L. Dai, C.-H. Chen, and J. R. Birge: Convergence properties of two–stage stochastic programming. J. Optim. Theory Appl. *106* (2000), 489–509. DOI:10.1023/a:1004649211111

[5] D. Dentcheva and A. Ruszczynski: Porfolio optimization with stochastic dominance constraints. J. Banking and Finance *30* (2006), 433–451. DOI:10.1016/j.jbankfin.2005.04.024

[6] J. Dupačová and R. J. B.Wets: Asymptotic behaviour of statistical estimates and optimal solutions of stochastic optimization problems. Ann. Statist. *16* (1984), 1517–1549. DOI:10.1214/aos/1176351052

[7] A. Dvoretzky, J. Kiefer, and J. Wolfowitz: Asymptotic minimax character of the sample distribution function and the classical multinomial estimate. Ann. Math. Statist. *56* (1956), 642–669. DOI:10.1214/aoms/1177728174

[8] Y. M. Ermoliev and V. Norkin: Sample everage approximation method for compound stochastic optimization problems. SIAM J. Optim. *23* (2013), 4, 2231–2263. DOI:10.1137/120863277

[9] A. Gut: Probability: A Graduate Course. Springer, New York 2005.

[10] M. Houda: Stability and Approximations for Stochastic Programs. Doctoral Thesis, Faculty of Mathematics and Physics, Charles University Prague, Prague 2009.

[11] M. Houda and V. Kaňková: Empirical estimates in economic and financial optimization problems. Bull. Czech Econometr. Soc. *19* (2012), 29, 50–69.

[12] Y. M. Kaniovski, A. J. King, and R. J.-B. Wets: Probabilistic bounds (via large deviations) for the solutions of stochastic programming problems. Ann. Oper. Res. *56* (1995), 189–208. DOI:10.1007/bf02031707

[13] V. Kaňková: Optimum solution of a stochastic optimization problem. In: Trans. 7th Prague Conf. 1974, Academia, Prague 1977, pp. 239–244.

[14] V. Kaňková: An approximative solution of stochastic optimization problem. In: Trans. 8th Prague Conf., Academia, Prague 1978, pp. 349–353.

[15] V. Kaňková and P. Lachout: Convergence rate of empirical estimates in stochastic programming. Informatica *3* (1992), 4, 497–523.

[16] V. Kaňková: Stability in stochastic programming – the case of unknown location parameter. Kybernetika *29* (1993), 1, 97–112.

[17] V. Kaňková: A note on estimates in stochastic programming. J. Comput. Appl. Math. *56* (1994), 97–112. DOI:10.1016/0377-0427(94)90381-6

[18] V. Kaňková: On the stability in stochastic programming: the case of individual probability constraints. Kybernetika *33* (1997), 5, 525–544.

[19] V. Kaňková and M. Houda: Empirical estimates in stochastic programming. In: Proc. Prague Stochastics 2006 (M. Hušková and M. Janžura, eds.), MATFYZPRESS, Prague 2006, pp. 426–436.

[20] V. Kaňková and M. Houda: Dependent samples in empirical estimation of stochastic programming problems. Austrian J. Statist. *35* (2006), 2 – 3, 271–279.

[21] V. Kaňková: Empirical estimates in stochastic programming via distribution tails. Kybernetika *46* (2010), 3, 459–471.

[22] V. Kaňková: Empirical estimates in optimization problems: survey with special regard to heavy tails and dependent samples. Bull. Czech Econometr. Soc. *19* (2012), 30, 92–111.

[23] V. Kaňková: Risk measures in optimization problems via empirical estimates. Czech Econom. Rev. *VII* (2013), 3, 162–177.

[24] L. B. Klebanov: Heavy Tailed Distributions. MATFYZPRESS, Prague 2003.

[25] M. M. Meerschaert and H.-P.Scheffler: Limit Distributions for Sums of Independent Random Vectors (Heavy Tails in Theory and Practice). John Wiley and Sons, New York 2001.

[26] M. M. Meerschaert and H.-P.Scheffler: Portfolio Modelling with Heavy Tailed Random Vectors. In: Handbook of Heavy Tailed Distributions in Finance (S. T. Rachev, ed.), Elsevier, Amsterdam 2003, pp. 595–640.

[27] M. M. Meerschaert and H.-P.Scheffler: Portfolio Modeling with Heavy Tailed Random Vectors. In: Handbook of Heavy Tailed Distributions in Finance (S. T. Rachev, ed.), Elsevier, Amsterdam 2003, pp. 595–640.

[28] G. Ch. Pflug: Scenario tree generation for multiperiod financial optimization by optimal discretization. Math. Program. Ser. B *89* (2001), 251–271. DOI:10.1007/pl00011398

[29] G. Ch. Pflug: Stochastic Optimization and Statistical Inference. In: Handbooks in Operations Research and Managemennt *10*, Stochastic Programming (A. Ruszczynski and A. A. Shapiro, eds.) Elsevier, Amsterdam 2003, pp. 427–480.

[30] G. Ch. Pflug and W. Römisch: Modeling Measuring and Managing Risk. World Scientific Publishing Co. Pte. Ltd, New Jersey 2007.

[31] S. T. Rachev and W. Römisch: Quantitative stability and stochastic programming: the method of probabilistic metrics. Math. Oper. Res. *27* (2002), 792–818. DOI:10.1287/moor.27.4.792.304

[32] R. Rockafellar, and R. J. B. Wets: Variational Analysis. Springer, Berlin 1983.

[33] W. Römisch and A. Wakolbinger: Obtaining Convergence Rate for Approximation in Stochastic Programming. In: Parametric Optimization and Related Topics (J. Guddat, H. Th. Jongen, B. Kummer and F. Nožička, eds.), Akademie–Verlag, Berlin 1987, pp. 327–343.

[34] W. Römisch: Stability of Stochastic Programming Problems. In: Handbooks in Operations Research and Managemennt Science *10*, Stochastic Programming (A. Ruszczynski and A. A. Shapiro, eds.) Elsevier, Amsterdam 2003, pp. 483–554.

[35] G. Salinetti and R. J.-B. Wets: On the convergence of sequence of convex sets in finite dimensions. SIAM Rev. *21* (1979), 16–33. DOI:10.1137/1021002

[36] G. Samarodnitsky and M. Taqqu: Stable Non-Gaussian Random Processes. Chapman and Hall, New York 1994.

[37] R. Schulz: Rates of convergence in stochastic programs with complete integer recourse. SIAM J. Optim. *6* (1996), 4, 1138–1152. DOI:10.1137/s1052623494271655

[38] A. Shapiro: Quantitative stability in stochastic programming. Math. Program. *67* (1994), 99–108. DOI:10.1007/bf01582215

[39] A. Shapiro and H. Xu: Stochastic mathematical programs with equilibrium constraints, modelling and sample average approximation. Optimization *57* (2008), 395–418. DOI:10.1080/02331930801954177

[40] A. Shapiro, D. Dentcheva, and A. Ruszczynski: Lectures on Stochastic Programming (Modeling and Theory). Published by Society for Industrial and Applied Mathematics and Mathematical Programming Society, Philadelphia 2009.

[41] A. N. Shiryaev: Essential of Stochastic Finance (Facts, Models, Theory). World Scientific, New Jersey 2008.

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

[43] M. Šmíd: The expected loss in the discretezation of multistage stochastic programming problems – estimation and convergence rate. Ann. Oper. Res. *165* (2009), 1, 29–45. DOI:10.1007/s10479-008-0355-9

[44] R. J.-B. Wets: A Statistical Approach to the Solution of Stochastic Programs with (Convex) Simple Recourse. Research Report, University of Kentucky 1974.

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

Michal Houda, Institute of Information Theory and Automation — Academy of Sciences of the Czech Republic, Pod Vodárenskou věží 4, 182 08 Praha 8. Czech Republic.
e-mail: houda@ef.jcu.cz