AUSTRIAN JOURNAL OF STATISTICS
Volume 35 (2006), Number 2&3, 271–279

Dependent Samples in Empirical Estimation of Stochastic Programming Problems

Vlasta Kaňková and Michal Houda
Academy of Sciences of the Czech Republic, Prague, Czech Republic

Abstract: Stochastic optimization models are built with the assumption that the underlying probability measure is entirely known. This is not true in practice, however: empirical approximation or estimates are used instead. The question then arises if such inaccuracy does not perturb resulting solutions and optimal values. We measure the ”distance” between the probability distributions by suitable metrics on the space of probability measures.

It is known that, under certain assumptions, the stability of the stochastic optimization model is assured with respect to the selected metric, and, moreover, the empirical estimate of the unknown distribution has suitable convergence properties, including a sufficient rate of convergence. In the case of Kolmogorov metric, the convergence rate is known if the random sample is independent and the probability measure is ”continuous”. In the case of Wasserstein (Mallows) metric, uniform distribution and independent random sample, the rate of convergence is the same in the case of the traditional uniform process and its limiting distribution is known; for other distributions, metrics, and the multidimensional case, the convergence properties and the rate of convergence have to be estimated e.g. by simulations. In our contribution, we show some numerical results for independent and dependent samples and make some backward interpretation of the results applied to stochastic programming.

Keywords: Stability, Probability Metrics, Wasserstein Metric, Kolmogorov Metric, Simulations.

1 Introduction

A problem of stochastic programming with a fixed constraint set can be mathematically formulated as

$$\inf_{x \in X} \int_{\Xi} g(x;\xi)\ \mu(\mathrm{d}\xi) \tag{1}$$

where $\xi \in \Xi$ is a random vector defined on a given probability space $(\Omega, \mathcal{A}, \mathbb{P})$, $\Xi \subset \mathbb{R}^s$ is a closed set (support of μ), $\mathcal{P}(\Xi)$ is the set of all the probability measures on $\mathcal{B}(\Xi)$ (Borel σ-field), $\mu \in \mathcal{P}(\Xi)$ is the distribution of ξ, $X \subset \mathbb{R}^m$ is a closed constraint set not depending on μ, and $g : \mathbb{R}^m \times \Xi \to \overline{\mathbb{R}}$ is a function lower semicontinuous in x and measurable in ξ. Denote $\varphi(\mu)$ the optimal value of the problem (1).

In order to solve the problem (1), most of the stochastic programming models require a *full knowledge* of the distribution μ. Applying the stochastic programming theory, there are two important issues in this regard:

- *the distribution is not known;* we need to find an *estimate*, from some historical data, for example;
- the distribution is known but *too complicated* to solve the problem efficiently; for example, solving the problem involves very often multidimensional integrals; in this case the distribution needs to be *approximated* by some simpler version of it, usually by a discrete distribution (sometimes called "scenarios" in stochastic programming).

Let us, therefore, replace the original distribution μ in (1) by another distribution denoted as ν. An important question then arises: how the optimal value $\varphi(\nu)$ and the optimal solution set of (1) change with respect to the "difference" between μ and ν. In order to quantify the changes in the probability distribution, we have to introduce a distance on some subspace of $\mathcal{P}(\Xi)$ of the probability measures on Ξ. This is the purpose of Section 2. We use a known stability result and show how the notion of probability metrics applies there. Section 3 is devoted to the empirical distribution and convergence properties of probability metrics when the empirical distribution is considered as the approximation ν to μ in (1). Numerical study is given in Section 4, comparing the role of independent and dependent samples in simulations.

2 Probability Metrics

A right selection of the distance on the space of probability measures is crucial in the study of stability in stochastic programming. Some examples of improperly selected metrics are given in Kaňková and Houda (2003); another illustration may also be found at the end of this paper (Cauchy distribution). In general, we could claim that the process of selection of the metric is closely related to the (mathematical) nature of the original problem.

2.1 Wasserstein Metric

For a large and common class of problems with fixed constraints having the form (1), *Wasserstein metric* is used. It is considered an "ideal" metric if the functional g is Lipschitz continuous in ξ – see Römisch (2003) for details about this notion of "ideality" in stochastic programming. One-dimensional **1-Wasserstein metric** is defined for $\mu, \nu \in \mathcal{P}_1(\Xi)$ by

$$W(\mu,\nu) = \int_{-\infty}^{+\infty} |F_\mu(t) - F_\nu(t)| \mathrm{d}t$$

where $\mathcal{P}_1(\Xi)$ is the class of probability measures on $\Xi \subset \mathbb{R}$ having finite first moments and F_μ and F_ν are (right continuous) distribution functions corresponding to μ and ν.

Wasserstein metric can be extended to the classes of probability distributions with finite higher moments or multidimensional distributions in several ways; we will not proceed in this direction and we refer the reader to the book of Rachev (1991) for further information.

Proposition 1. *If, in* (1), $\mu \in \mathcal{P}_1(\Xi)$, $\nu \in \mathcal{P}_1(\Xi)$, X *is compact,* g *is uniformly continuous on* $\mathbb{R}^m \times \mathbb{R}^s$, *and* g *is Lipschitz continuous in* ξ *for all* $x \in X$ *with a constant* L *not* *depending on x, then*

$$|\varphi(\mu) - \varphi(\nu)| \le L\, W(\mu, \nu).$$

See Houda (2002) for the proof based on the results of Römisch and Schultz (1993). Proposition 1 shows that the stability of the problem (1) depends on its structure (represented by the constant L) and on the probability (Wasserstein) distance. If $s > 1$ then we consider the multidimensional version of the Wasserstein metric.

2.2 Kolmogorov Metric

The **Kolmogorov metric** is defined for $\mu, \nu \in \mathcal{P}(\Xi)$ by

$$\mathcal{K}(\mu, \nu) = \sup_{t \in \Xi} |F_\mu(t) - F_\nu(t)|$$

It is considered "ideal" for more complicated structures than the problem (1); however, the main importance of this notion lies in its (relative) computational simplicity – it is available even in cases where other metrics fail to be evaluated.

Proposition 2. *Let the assumption of Proposition 1 be fulfilled, let μ be an absolutely continuous distribution with compact support and such that its density f_μ fulfils $f_\mu \ge \vartheta > 0$ for some constant ϑ, and let the support of ν be contained in a sufficiently small neighborhood of the support of μ. Then*

$$|\varphi(\mu) - \varphi(\nu)| \le 16 L \sqrt{s} \Big(\frac{2\mathcal{K}(\mu, \nu)}{\vartheta} \Big)^{1/s}.$$

See Kaňková (1994b) for the proof. The upper bounds given in Proposition 2 are usually sharper than those of Proposition 1, but often the only available, especially if the dimension s of the support Ξ is large, moreover, the assumption of the existence of the first moment is not required.

3 Empirical Distributions and Processes

If the probability measure μ, needed for a successful solution of the stochastic optimization problem, is not available, then we have to use the empirical data at hand and replace the original distribution by the empirical version. This is the subject of the present part of the paper. Subsequently, one can successfully apply the stability results of the previous section in order to get upper bounds of the optimal value; similar results concerning optimal solution sets also exist (see e.g. Römisch, 2003, Kaňková and Houda, 2003, and references therein).

3.1 Empirical Distribution

Let $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ be independent random variables with the same probability distribution μ. For notational simplicity, denote its distribution function by F instead of F_μ. The random function

$$F_n(t) = \frac{1}{n} \sum_{i=1}^{n} I_{(-\infty;t]}(\xi_i), \quad t \in \mathbb{R} \tag{2}$$

is called **empirical distribution function** based on the sample $\xi_1, \dots, \xi_n$ (I_A is an indicator function of the set A). For each realization of the sample, $F_n(t)$ is actually a distribution function; we denote the associated probability measure as μ_n and call it *empirical measure*.

It is well known that the sequence of empirical distribution functions F_n converges almost surely to the distribution function F under general conditions as n goes to infinity. Considering the definition of Wasserstein and Kolmogorov metrics, the values of these metrics for F and F_n should converge, as well (see e.g. Shorack and Wellner, 1986). We illustrate it in Section 4 of the paper.

3.2 Empirical Processes

We are also interested in the behavior of the so-called **integrated empirical process** defined by

$$\sqrt{n}\, W(\mu_n, \mu) = \int_{-\infty}^{+\infty} \sqrt{n}\, |F_n(t) - F(t)| \,\mathrm{d}t \tag{3}$$

In the present paper, the (integrated) empirical process is considered in a more general sense than usual: we assume only that $\mu \in \mathcal{P}_1(\Xi)$. In the case of uniform distribution on [0;1], (3) describes the (integrated) empirical process in the usual sense. That process is known as Mallows statistic and its (weak) limit is the integral of the *Brownian bridge* $\mathbb{U}$. The probability distribution of the limit is known explicitly, see Section 3.8 of Shorack and Wellner (1986) for the exact formula.

In the case of other distributions, we apply Theorem 2.1 from Barrio et al. (1999): $\sqrt{n}(F_n(t) - F(t)) \to_w \mathbb{U}(F(t))$ in $L_1(\mathbb{R})$ if and only if $\int_{-\infty}^{+\infty} \sqrt{F(t)(1-F(t))}\mathrm{d}t < +\infty$. From this theorem it easily follows that, under the last condition,

$$\int_{-\infty}^{+\infty} \sqrt{n} \left| \frac{1}{n} \sum_{i=1}^{n} I_{(-\infty;t]}(\xi_i) - F(t) \right| \mathrm{d}t \to_d \int_{-\infty}^{+\infty} |\mathbb{U}(F(t))| \mathrm{d}t. \tag{4}$$

See Barrio et al. (1999) for details about the weak convergence in $L_1(\mathbb{R})$ (we have used the fact that if some processes Y_n converge weakly in $L_1(\mathbb{R})$ to Y, then, among others, $||Y_n||_{L_1} \to_d ||Y||_{L_1}$ where $||g||_{L_1} = \int_{-\infty}^{\infty} g(t)\mathrm{d}t$ for each non-negative $g \in L_1(\mathbb{R})$). In Section 4, we illustrate this convergence by simulations for a variety of distributions.

The convergence rate of Kolmogorov metric is well known for iid samples: if ξ_1, ξ_2, ..., ξ_n correspond to a probability measure that is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}$, then

$$\mathbb{P}\big\{\sqrt{n} \sup_{-\infty<t<\infty} |F_n(t) - F(t)| < x\big\} \to \begin{cases} \displaystyle\sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2x^2} & \text{for } x > 0 \\ 0 & \text{for } x \le 0. \end{cases}$$

The rate of convergence is exponential and independent on the original distribution. This kind of result is not known in case of Wasserstein metric. We no longer proceed in this direction.

4 Dependent and Independent Data

In economic and engineering applications, weakly dependent samples are of a very practical interest. We have seen that Wasserstein metric and its convergence properties play an important role in the stability of stochastic programming. Now let us relax the assumption of independence in (2) and assume one of the types of the weak dependence instead. Empirical estimates have already been investigated (in the literature) for some types of weakly (e.g. mixing) dependent random sequences (for some details see e.g. Dai, Chen, and Birge, 2000; Kaňková, 1994a; or Wang and Wang, 1999).

In this paper, we restrict our consideration to the very special case of M-dependent sequences. Let $\{\xi_t\}_{-\infty}^{+\infty}$ be a random sequence defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Let, moreover, $\mathcal{B}(-\infty, a)$ be the σ-field generated by $\ldots, \xi_{a-1}, \xi_a$, and $\mathcal{B}(b, +\infty)$ be the σ-field generated by $\xi_b, \xi_{b+1}, \ldots$. The sequence is said to be M*-dependent* if $\mathcal{B}(-\infty, a)$ and $\mathcal{B}(b, +\infty)$ are independent for $b - a > M$.

We can prove (by the techniques employed in Kaňková, 1994a) that for every natural n there exists $k \in \{0, 1, \ldots\}$ and $r \in \{0, \ldots, M\}$ such that $n = Mk + r$, and

$$|F_n(t) - F(t)| \leq \sum_{j=1}^{M} \frac{n_j}{n} |F_{n_j}(t) - F(t)|, \quad t \in \mathbb{R} \tag{5}$$

where F_{n_j} are empirical distribution functions determined by sequences of n_j independent random variables; moreover, $n_j = k + 1$ for $Mk + 1 \leq n \leq Mk + r$ and $n_j = k$ for $Mk + r < n \leq M(k + 1)$.

Clearly, it follows from the relation (5) that the asymptotic properties corresponding to M-dependent sequences are very similar to those proved by Proposition 1 and Proposition 2 for the independent case (the convergence is slower, of course). Moreover, it follows from Yoshihara (1992) that every stationary ϕ-mixing normal distributed sequence is also M-dependent for some $M \in \mathbb{N}$. A stationary Gaussian random sequence $\{\xi_t\}_{-\infty}^{+\infty}$ is ϕ-mixing if and only if the σ-fields $\mathcal{B}(-\infty, k)$ and $\mathcal{B}(k + n, +\infty)$ are independent for any n sufficiently large. For definitions of ϕ-mixing and more details about M-dependent sequences see e.g. Yoshihara (1992).

4.1 Simulation Study Overview

According to our analysis above, we focus on the numerical illustration of Proposition 1 (Wasserstein metric) for several "representative" one-dimensional probability measures. Of course, the investigation of both the more dimensional case and the "Kolomogorov empirical process" can be very useful, however, this investigation goes beyond the possibilities of the paper.

We have first generated iid samples $\zeta_1, \zeta_2, \ldots$ from the given distribution and then made up a new series defined as $\xi_k := 0.5\zeta_k + 0.5\zeta_{k-1}$. The theoretical distribution of ξ_k is given by convolution. In particular, it is

- triangular (Simpson's) in case of uniform samples on $[0; 1]$;
- gamma with the shape and rate parameters both equal to 2 in case of the exponential distribution with parameter $\lambda = 1$;

- normal with zero mean and variance 0.5 in case of normal distribution $N(0;1)$;
- Cauchy with the original parameters.

The empirical distribution function is given as before, based now on the (dependent) series (ξ_i); Wasserstein distance is then calculated with respect to the theoretical distribution.

A four-graph set for each of the examined distributions is given. The left column of the set is devoted to the independent data, the right one to the dependent data; the dotted line displays the samples of length 100 and the solid line the samples of length 1000. The first row of the graph shows the densities of the Wasserstein metric values (they should converge to zero), the second row are the densities of the empirical process (3).

4.2 Normal, Uniform, and Exponential Distribution

As expected, the convergence properties are well satisfied in this simple case of MA(1) process. The weak dependence (it is 2-dependence in this case) does not make great difficulties for the convergence; the differences between the densities for independent and 2-dependent data are very small if any. Theoretical background for these numerical results, and definitions of more complicated types of dependence can be found in Yoshihara (1992).

4.3 Cutted Cauchy Distribution

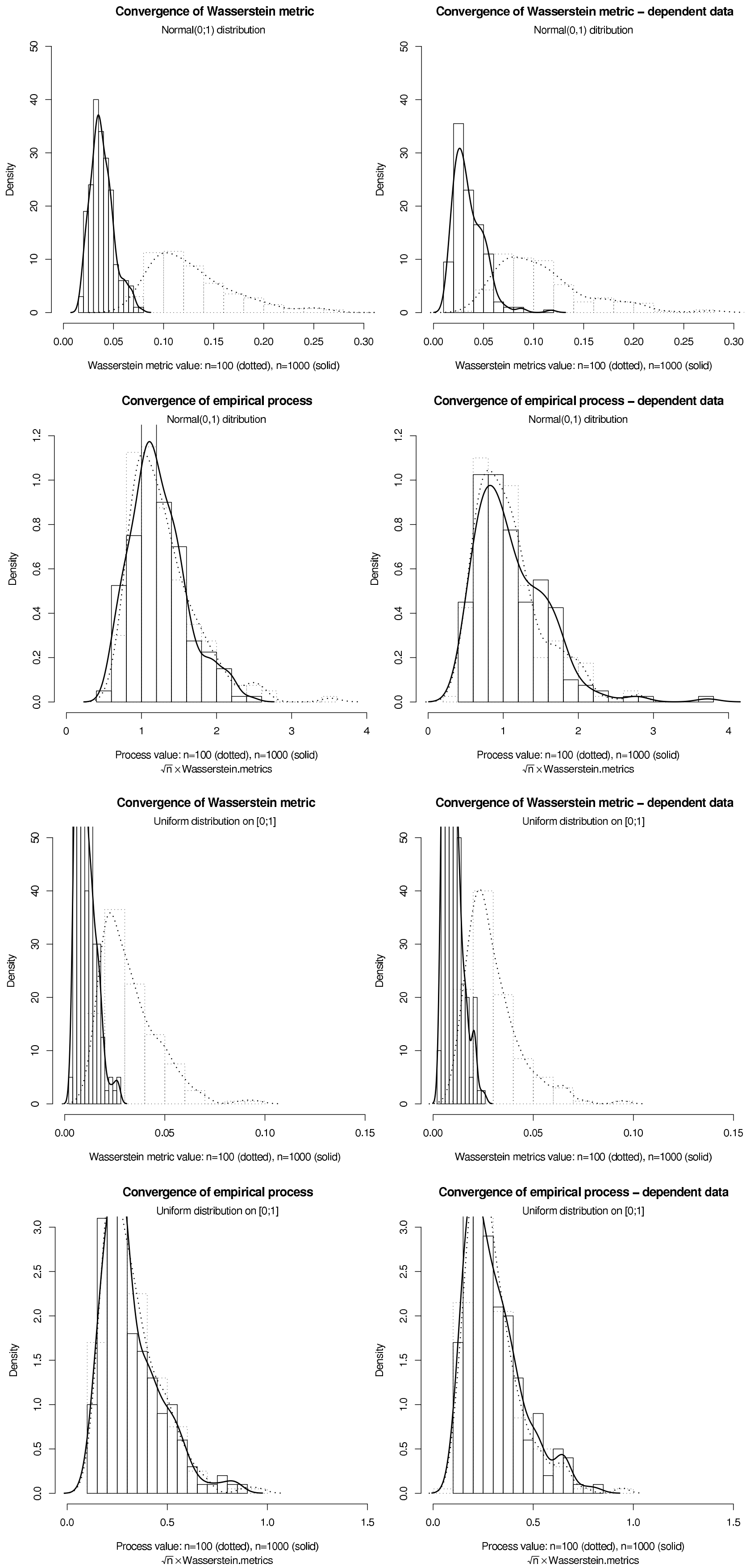
The cutted Cauchy distribution is an example of a distribution with heavy tail. It is evident that Wasserstein metric does not deal well with this type of distribution – the convergence is very slow and the distribution of the limiting process does *not* stabilize after a small number of samples. The dependence is still a problem in this case. Cutting is necessary because standard Cauchy distribution does not have the first moment and Wasserstein metric is not defined for it.

5 Conclusion

The present paper illustrates some results regarding the stability of stochastic programming problems with respect to changes in the underlying distribution. The numerical study is given for the simple case of weakly (2-)dependent data and we show that it does not represent essential difficulties. The theoretical part of the paper formulates some basic results which may be subject of our future research.

Acknowledgement

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


Convergence of Wasserstein metric
Normal(0;1) distribution
Density
Wasserstein metric value: n=100 (dotted), n=1000 (solid)
Convergence of Wasserstein metric – dependent data
Normal(0,1) ditribution
Density
Wasserstein metrics value: n=100 (dotted), n=1000 (solid)
Convergence of empirical process
Normal(0,1) ditribution
Density
Process value: n=100 (dotted), n=1000 (solid)
√n × Wasserstein.metrics
Convergence of empirical process – dependent data
Normal(0,1) ditribution
Density
Process value: n=100 (dotted), n=1000 (solid)
√n × Wasserstein.metrics
Convergence of Wasserstein metric
Uniform distribution on [0;1]
Density
Wasserstein metric value: n=100 (dotted), n=1000 (solid)
Convergence of Wasserstein metric – dependent data
Uniform distribution on [0;1]
Density
Wasserstein metrics value: n=100 (dotted), n=1000 (solid)
Convergence of empirical process
Uniform distribution on [0;1]
Density
Process value: n=100 (dotted), n=1000 (solid)
√n × Wasserstein.metrics
Convergence of empirical process – dependent data
Uniform distribution on [0;1]
Density
Process value: n=100 (dotted), n=1000 (solid)
√n × Wasserstein.metrics

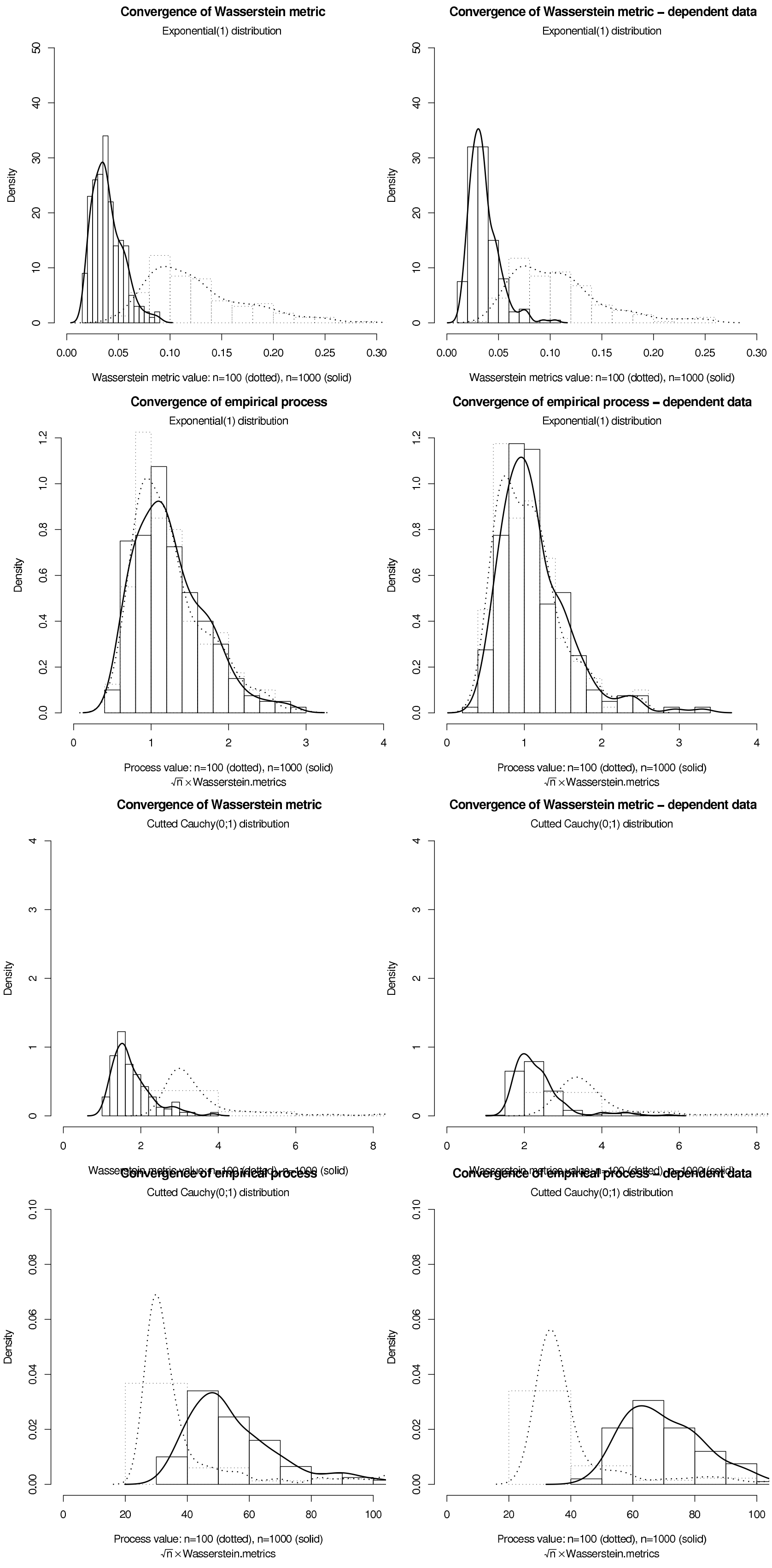


Figure 1: Convergence properties for dependent and independent data samples.

References

Barrio, E. del, Giné, E., and Matrán, C. (1999). Central limit theorems for the Wasserstein distance between the empirical and the true distributions. *The Annals of Probability*, *27*(2), 1009-1071.

Dai, L., Chen, C. H., and Birge, R. J. (2000). Convergence properties of two-stage stochastic programming. *Journal of Optimization Theory and Applications*, *106*(3), 489-509.

Houda, M. (2002). Probability metrics and the stability of stochastic programs with recourse. *Bulletin of the Czech Econometric Society*, *9*(17), 65-77.

Kaňková, V. (1994a). A note on estimates in stochastic programming. *Journal of Computational and Applied Mathematics*, *56*(1-2), 97-112.

Kaňková, V. (1994b). 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* (p. 122-125). Prague: Academy of Sciences of the Czech Republic.

Kaňková, V., and Houda, M. (2003). A note on quantitative stability and empirical estimates in stochastic programming. In U. Leopold-Wildburger, F. Rendl, and G. Wascher (Eds.), *Operations Research Proceedings 2002* (p. 413-418). Berlin: Springer.

Rachev, S. T. (1991). *Probability Metrics and the Stability of Stochastic Models*. Chichester: Wiley.

Römisch, W. (2003). Stability of stochastic programming problems. In A. Ruszczynski and A. Shapiro (Eds.), *Stochastic Programming. Handbooks in Operations Research and Management Science* (Vol. 10, p. 483-554). Amsterdam: Elsevier.

Römisch, W., and Schultz, R. (1993). Stability of solutions for stochastic programs with complete recourse. *Mathematics of Operations Research*, *18*(3), 590-609.

Shorack, G. R., and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics*. New York: Wiley.

Wang, L., and Wang, J. (1999). Limit distribution of statistical estimates for stochastic programs with dependent samples. *ZAMM Journal of Applied Mathematics and Mechanics*, *79*(4), 257-266.

Yoshihara, K. (1992). *Summation Theory for Weakly Dependent Sequences. Weakly Dependent Stochastic Sequences and their Applications* (Vol. 1). Tokyo: Sanseido.

Authors' address:

Vlasta Kaňková and Michal Houda
Institute of Automation and Information Theory
Academy of Sciences of the Czech Republic
Pod vodárenskou věží 4
CZ-182 08 Praha 8
Czech Republic

E-mail: kankova@utia.cas.cz, houda@karlin.mff.cuni.cz