

# Using Metrics in Stability of Stochastic Programming Problems<sup>#</sup>

Michal Houda<sup>\*</sup>

## Introduction

A variety of optimization techniques are widely used when we model economical situation. Input parameters of such models could take different nature; especially important are these that are random by origin. Many approaches are possible. First, the simplest way is to ignore the randomness. This is acceptable in many cases; post-optimisation sensitivity analysis has to confirm or reject such decision.

Examples show that – even in linear programming – there are examples where the approach described (ignoring the randomness) fails. Kall (1976) demonstrates this by the following classical example.

**Example 1** (Kall, 1976). Consider a linear optimisation problem: minimise  $x+y$  subject to  $ax+y \leq 7$ ,  $bx+y \leq 4$ ,  $x,y \geq 0$ , where  $a, b$  are random coefficients with uniform distribution on  $[1;4]$  and  $[1/3;1]$  respectively. If we replace  $a$  and  $b$  with their mean (expected) values, we obtain a solution that is not acceptable with probability of 75 %!

Example 1 shows that different approaches are necessary. A scientific discipline dealing with optimization models where the randomness is part of the model is known as *stochastic programming*. A large class of solved problems are these where a decision is to be taken before the realisation of the random parameter is known. We refer to the book of Birge and Louveaux (1997) for a detailed introduction to stochastic programming.

In all decision problems of stochastic programming, uncertainty of parameters is modelled via probability distribution of random variables rather than the random variable itself. Additionally, the *entire* knowledge of the probability distribution is required. As easily seen, usual real-life tasks cannot meet this assumption: the probability distribution is not really known but only estimated, for example, from historical data series. In another class of models the probability distribution is known but far more complicated than we could solve the problem easily; then some approximation (usually discrete one, called *scenarios*) is used instead.

Both of the mentioned methods require another post-optimisation sensitivity analysis: this time with respect to changes in probability distribution. The problem is how to measure these changes: mathematically, the probability distribution is a measure, so we have to quantify “distance” between two measures (i.e., between two abstract

---

<sup>#</sup> This research is supported by the project *Nonlinear Dynamical Economic Systems: Theory and Applications* registered by the Grant Agency of the Czech Republic under No. 402/03/H057 and by the project *Stochastic Decision Approaches in Nonlinear Economic Models* registered by the Grant Agency of the Czech Republic under No. 402/04/1294.

<sup>\*</sup> Mgr. Michal Houda – student of doctoral study; Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics, Charles University in Prague, Ke Karlovu 3, 121 16 Prague 2; <houda@karlin.mff.cuni.cz>.

mathematical objects). Borrowed from the area of functional analysis, a tool known as *probability metric* (a metric on some space of probability measures) is now widely accepted. The problem is still not trivial: choosing a right metric is crucial and cannot be arbitrary done as we show in the next section.

A great attention has already been paid in the literature to the stability in stochastic programming. For example, a notion of *minimal information (m. i.) metric* introduced by Rachev (1991) and Zolotarev (1983) is now considered as a starting point for forthcoming analysis. See Römisch (2003) for a basis and further references.

## Probability metrics

Among a broad set of probability metrics we will now talk about two of them: *Kolmogorov and Wasserstein distances*. Besides that these two fall into the range of metrics with  $\zeta$ -structure (considered as m. i. metric for the stability for a certain specified class of stochastic programs), they both have an unquestionable virtue: they could be quite easily enumerated. This important property serves us to present a numerical illustration at the end of the paper. Next definitions could be found for example in Rachev (1991).

### Kolmogorov metric

The *Kolmogorov metric* is defined on the space of all probability measures as a uniform distance of their distribution functions:

$$K(F, G) = \sup_{z \in \Xi} |F(z) - G(z)|, \quad [1]$$

where  $\Xi \in \mathbb{R}^s$  is a support of the probability measures and  $F, G$  are their distribution functions. Knowing both distribution functions, one could easily calculate the value of the corresponding Kolmogorov distance.

### Wasserstein metric

The *1-Wasserstein metric* for one-dimensional distribution is defined on the space of probability measures having finite first moment as a surface between their distribution functions:

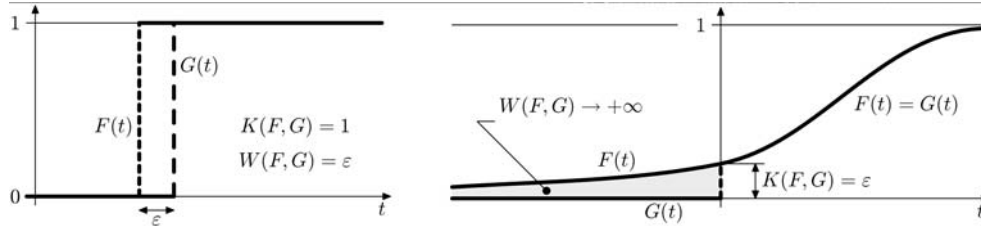
$$W(F, G) = \int_{-\infty}^{+\infty} |F(z) - G(z)| dz. \quad [2]$$

More general definitions for multidimensional Wasserstein metrics are to be found in Rachev (1991). The representation given here is again suitable for numerical treatment.

It turns out that the problem of choosing a right metric is closely related to the properties of the original model and cannot be treated separately from it. Next example illustrates different properties of Kolmogorov and Wasserstein distances.

**Example 2** (Kaňková – Houda, 2002). The left part of Figure 1 represents approximation of a discrete (here degenerated) distribution with unknown mass point. In that case, the Kolmogorov metrics takes the value of one, so does not provide any useful information about how far the two distributions are. The right part of the figure represents a distribution with heavy tails:  $F$  and  $G$  take forms of  $F(z) = \varepsilon/(1-z)$  on  $[-K; 0)$  with  $\varepsilon \in (0; 1)$  and  $K > 0$ , and  $F(z) = G(z)$  are arbitrary on  $[0; +\infty)$ . Now the Wasserstein metric could take arbitrary high value as illustrated.

**Fig. 1: Wasserstein and Kolmogorov metrics as unsuitable distances**



## Problem formulation and stability results

A general decision problem in stochastic programming is formulated for a (say unknown or original) probability distribution  $F$  as

$$\text{minimise } E_F g(x; \xi) \text{ subject to } x \in X, \quad [3]$$

where  $g$  is uniformly continuous function measurable in  $\xi$ ,  $X$  is compact set of feasible solutions. In [3],  $x$  denotes control (or decision) variable,  $\xi$  is random input parameter with distribution  $F$ ,  $E$  is the symbol for mathematical expectation. This class of problems is known as problems with fixed constraint set  $X$ . A generalisation for a “random” constraint set is still possible but not treated in this paper (see e. g. Birge – Louveaux, 1997, Römisch, 2003). For example,  $g$  could be a function representing production costs that we want to minimise; but such cost function  $g$  usually depends on a random element, so rather than minimising  $g$  directly we look at its expected value.

In [3], let replace the original distribution  $F$  with an estimate  $G$  and denote  $\varphi(F)$  and  $\varphi(G)$  optimal values of the problem [3] for the two distributions. Stability theory asks about the difference between  $\varphi(F)$  and  $\varphi(G)$ , i. e. look for an upper bound for the difference  $|\varphi(F) - \varphi(G)|$ . In Houda (2002), such bound is given for  $g$  that is Lipschitz continuous in  $\xi$ , and with respect to the Wasserstein metric:

$$|\varphi(F) - \varphi(G)| \leq L \times W(F, G) \quad [4]$$

where  $L$  is Lipschitz constant for  $g$ . As we can see, the right hand side of [4] is double-structured: first part depends on the model structure by means of properties of  $g$  – here by the constant  $L$ ; second part measures the distance between  $F$  and  $G$  by the Wasserstein metric. This is the simplest result that one can draw up. Similar but more complicated formulas could be obtained for other stochastic programs, with respect to other metrics, for solutions or solution sets of stochastic programs (see Römisch, 2003, and references therein).

## Empirical distributions

Let now consider widespread approximations of probability distributions called empirical distributions. A one-dimensional *empirical distribution function* based on the sample  $\xi_1, \xi_2, \dots$  of random variables with the common distribution function  $F$  is a random variable (function) defined by

$$F_n(z) = F_n(z; \xi) = \frac{1}{n} \sum_{i=1}^n I_{(-\infty; z]}(\xi_i) \quad [5]$$

where  $I_A$  denotes the indicator function of a set  $A$ . It is well known that the sequence of empirical distribution functions converges almost surely to the distribution function  $F$  under rather general conditions. For each realization of the sample  $\xi$ ,  $F_n(z)$  is actually a distribution function. We illustrate now convergence properties of the Wasserstein metric using this notion of the empirical distribution.

### Simulation study and comments

All necessary algorithms and computation outputs have been created using R programming language and interface<sup>1</sup>. Here we only present results for normal distribution (representing a “good” distribution) and modified (cut) Cauchy distribution (representing “bad” distribution with heavy tails). More precisely, we estimate a distribution for Wasserstein distance  $W(F, F_n)$  for a given distribution (normal, Cauchy) and a length  $n$  (100 and 1000). Each estimate is based on the sample of 200 realizations of  $\xi$  for each pair (distribution, length).

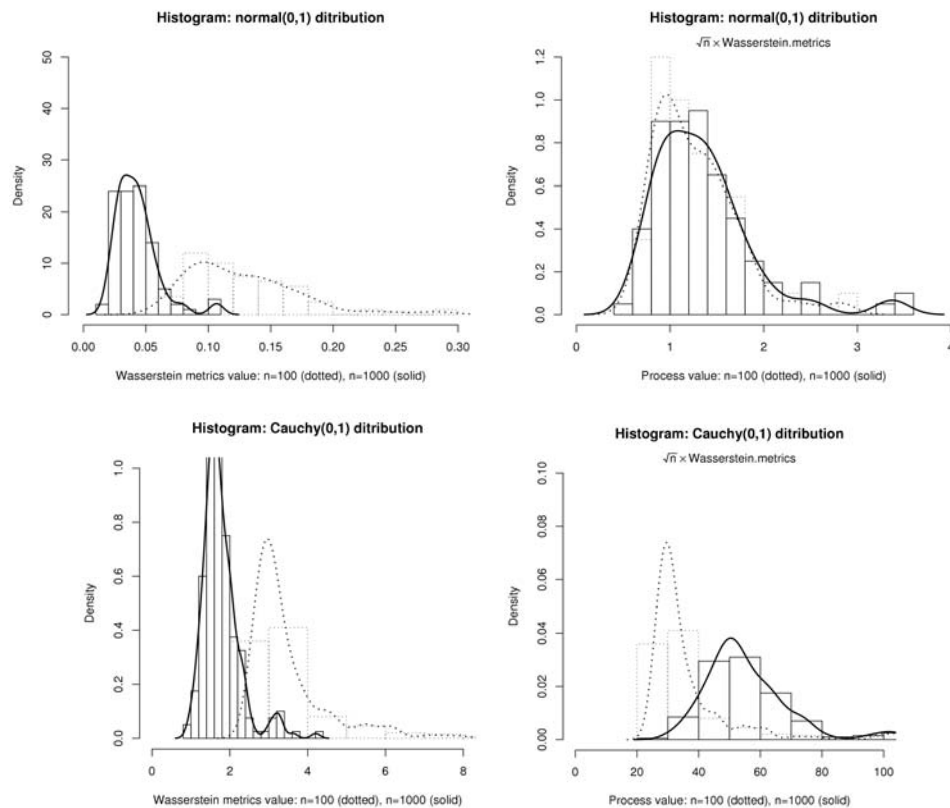
The left-hand sides of Figure 2 illustrate a fact that the Wasserstein metric converges to zero if the sample length  $n$  tends to infinity. But the speed of the convergence is about four times smaller for Cauchy distribution (compare values at horizontal axe). This is that we have actually expected.

The right-hand sides of Figure 2 are estimates for the probability density of the stochastic process  $\sqrt{n}W(F, F_n)$ , representing a convergence rate. Theoretically, the limiting distribution for this process could be explicitly given only for the uniform distribution on  $[0;1]$  (see Shorack – Wellner, 1986, chapter 3, p. 150). Again, simulations show that for (“good”) normal distributions, the probability density stabilizes rapidly – however this is not true for Cauchy distribution. The reader can find analogous graphics for other distributions (uniform, exponential, and beta) in Houda (2004).

From practical point of view, if we derive the Lipschitz constant of the model, we can directly conclude on upper bound in [4] and thus make comments about influence of the estimate of the distribution to the resulting solutions. As seen in Figure 2, these estimates are useful only in case of “good” distribution – i. e., when the used metric and original distribution “coincide”. Otherwise, obtained upper bounds are too high to make reasonable conclusions.

<sup>1</sup> R is freely available software distributed at <http://www.r-project.org/>.

**Fig. 2: Convergence properties of Wasserstein metric**



## Conclusion

We have restated one of the numerous results from the area of quantitative stability in stochastic programming. In order to quantify differences in optimal values and optimal solutions when an estimate or an approximation replaces the original distribution of the random element, a suitable notion of “distance between probability distributions” has to be defined. Example illustrates that the right choice of this distance (or probability metric) is closely related to the properties of the original probability distribution in the model, but also have a considerable consequence to the stability of the model. In real-life applications, this kind of post-optimization analysis – searching for errors with respect to changes in probability distribution – cannot be thoughtlessly automated without keeping in mind effective properties of the model.

## Acknowledgement

The author wishes to thank Vlasta Kaňková (Academy of Science of the Czech Republic) and other people from the Czech Academy of Science and Charles University of

Prague for their invaluable help and suggestions in the area of stochastic optimization and related disciplines.

## References

- [1] BIRGE, J. R. – LOUVEAUX, F. (1997): *Introduction to Stochastic Programming*. New York, Springer-Verlag.
- [2] HOUDA, M. (2002): *Probability metrics and the stability of stochastic programs with recourse*. Bulletin of the Czech Econometric Society, vol. 9, issue 17, pp. 65–77.
- [3] HOUDA, M. (2004): *Wasserstein metrics and empirical distributions in stability of stochastic programs*. In Lukáčik M. (ed.): *Proceedings of the International Conference Quantitative Methods in Economics (Multiple Criteria Decision Making XII)*. Bratislava, University of Economics, pp. 71–77.
- [4] KALL, P. (1976): *Stochastic Linear Programming*. Berlin, Springer-Verlag.
- [5] KAŇKOVÁ, V. – HOUDA, M. (2002): *A note on quantitative stability and empirical estimates in stochastic programming*. In Leopold-Wildburger, U., Rendl, F., Wäshel, G. (eds.): *Operations Research Proceedings 2002*. Heidelberg, Springer-Verlag, pp. 413–418.
- [6] RACHEV, S. T. (1991): *Probability Metrics and the Stability of Stochastic Models*. Chichester, Wiley.
- [7] RÖMISCH, W. (2003): *Stability of Stochastic Programming Problems*. In Ruszczyński, A., Shapiro, A. (eds.): *Stochastic Programming, Handbooks in Operations Research and Management Science*, vol. 10. Amsterdam, Elsevier, pp. 483–554.
- [8] SHORACK, G. R. – WELLNER, J. A. (1986): *Empirical Processes with Applications to Statistics*. New York, John Wiley&Sons.
- [9] ZOLOTAREV, V. M. (1983): *Probability metrics*. Theory of Probability and its Applications, vol. 28, pp. 278–302.

## **Užití metrik ve stabilitě úloh stochastického programování**

*Michal Houda*

### **Abstrakt**

Optimalizační techniky často vstupují jako matematický nástroj do mnoha ekonomických aplikací. Nejistota je v takových modelech reprezentována prostřednictvím rozdělení pravděpodobnosti, které je následně v praktických úlohách aproximováno nebo odhadováno. Do hry tak vstupuje otázka stability získaných řešení, a to vzhledem ke změnám v tomto pravděpodobnostním rozdělení. Práce ilustruje jeden z možných přístupů (použití pravděpodobnostních metrik), možné numerické problémy a zpětný pohled na ekonomickou interpretaci.

**Klíčová slova:** stochastické programování; kvantitativní stabilita; Wassersteinova metrika; Kolmogorovova metrika; simulační studie.

## **Using Metrics in Stability of Stochastic Programming Problems**

### **Abstract**

Optimization techniques enter often as a mathematical tool into many economic applications. In these models, uncertainty is modelled via probability distribution that is approximated or estimated in real cases. Then we ask for a stability of solutions with respect to changes in the probability distribution. The work illustrates one of possible approaches (using probability metrics), underlying numerical challenges and a backward glance to economical interpretation.

**Key words:** stochastic programming; quantitative stability; Wasserstein metrics; Kolmogorov metrics; simulation study.

**JEL classification:** C44