

Comparison of approximations in stochastic and robust optimization programs

Michal Houda

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

MSC 2000: 90C15, 90C25

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

1 Introduction

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

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

1.1 Uncertain convex program

An uncertain convex program is an optimization problem in which the constraints are not precisely known. Formally, consider $\xi \in \Xi \subset \mathbb{R}^s$ (a random, uncertainty,

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

or instance parameter), $X \subset \mathbb{R}^n$ convex and closed set, and a function $f(x; \xi) : X \times \Xi \rightarrow \mathbb{R}$ convex in x for all $\xi \in \Xi$; an *uncertain convex program* (UCP) is then a problem

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

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

1.2 Chance-constrained approach

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

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

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

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

1.3 Robust programming approach

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

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

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

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

2 Sampled convex programs

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

2.1 Chance-constrained sampled program

Consider a set of independent samples ξ_1, \dots, ξ_N distributed according to P , the original distribution of the parameter ξ . We define the *empirical distribution function* as a discrete random variable of the form $P_N := \frac{1}{N} \sum \delta_{\xi_i}$ where δ_{ξ} denotes the Dirac measure placing the unit mass at ξ . The (PCP) problem is now approximated, for the given sample, by replacing the original probability distribution P by P_N and the problem then reads

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

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

There exists many results in the theory of stability of stochastic optimization problems dealing with a question how far is the optimal solution of (PCP_N) from the original optimal solution of (PCP); among all we refer to works [8], [10], [11], [13], [14], references therein, and many of other authors. Here, general stability theorem (Theorem 1 in [8]) can be considered as a base for further special results. As an example we recall the following proposition, formulated as in the original paper for a linear (vector) function $f(x; \xi)$.

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

1. $f(x; \xi) = \xi - Tx$, $T \in \mathbb{R}^n \times \mathbb{R}^s$ is a constant matrix of parameters;
2. P is a logarithmic concave measure, i. e.,
 $P\{\lambda B_1 + (1 - \lambda)B_2\} \geq P(B_1)^\lambda P(B_2)^{1-\lambda}$ is valid for all $\lambda \in [0; 1]$ and all convex Borel $B_1, B_2 \subset \mathbb{R}^s$ such that $\lambda B_1 + (1 - \lambda)B_2$ is also Borel set;
3. the optimal solution set $\Psi(P)$ of (PCP) is nonempty and bounded and $\Psi(P) \cap \operatorname{argmin}_{x \in X} c'x = \emptyset$;
4. there exists an $\bar{x} \in X$ such that $F_P(T\bar{x}) > 1 - \varepsilon$, F_P is the distribution function corresponding to P ;
5. $\log F_P$ is strongly concave on some convex neighbourhood $T\Psi(P)$, i. e., there is a constant $c > 0$ such that
 $\log F_P(\lambda t_1 + (1 - \lambda)t_2) \geq \lambda \log F_P(t_1) + (1 - \lambda) \log F_P(t_2) + \frac{1}{2}c\lambda(1 - \lambda)\|t_1 - t_2\|^2$
is valid for all $\lambda \in [0; 1]$ and all t_1, t_2 from the convex neighbourhood of $T\Psi(P)$.

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

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

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

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

2.2 Robust sampled program

Recently, Calafiore and Campi [4] and de Farias and Van Roy [6] independently proposed the following approximations to (RCP). Consider again a set of independent samples ξ_1, \dots, ξ_N distributed according to P . The (RCP) is then approximated by asking the constraints to be satisfied for all ξ_i :

$$\min_{x \in X} c'x \quad \text{subject to} \quad X[N] := \{x \in X; f(x; \xi_i) \leq 0 \text{ for } i = 1, \dots, N\} \quad (\text{SCP}_N)$$

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

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

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

3 Numerical study

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

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

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

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

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

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

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

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

In the first rank, dotted histograms represent the fact that the optimal solution of the chance-constrained sampled problem (PCP_N) converges, as N goes to infinity, to the solution of (PCP) , marked by a short tickmark on x -axis. This sampling method is useful especially if the number N of samples is high, as the possible error in estimating optimal solution decreases. The optimal solution of the second

mentioned approach, robust sampled problem (SCP_N), goes to the upper boundary of the support of F (i. e. to infinity in our cases), but with rapidly decreasing rate. The tickmark now represents the lower boundary of ε -feasibility set X_ε , i. e. the limiting point for which a solution is feasible for (PCP) with a high probability. You could observe the fact mentioned in Proposition 2.2 – if N is greater than 241, then the optimal value of the (SCP_N) program is feasible in (PCP) with probability of 0.95.

4 Conclusion

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

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

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

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

References

- [1] A. Ben-Tal and A. Nemirovski. Robust convex optimization. *Mathematics of Operations Research*, 23(4):769–805, 1998.
- [2] D. Bertsimas and M. Sim. Tractable approximations to robust conic optimization problems. *Mathematical Programming*, published online, Dec. 2005.
- [3] G. Calafiore and M. C. Campi. Decision making in an uncertain environment: the scenario based optimization approach. In J. Andrýšek, M. Kárný, and J. Kracík, editors, *Multiple Participant Decision Making*, pages 99–111. Advanced Knowledge International, 2004.

- [4] G. Calafiore and M. C. Campi. Uncertain convex programs: randomized solutions and confidence levels. *Mathematical Programming*, 102 (1, Ser. A):25–46, 2005.
- [5] A. Charnes and W. W. Cooper. Chance constrained programming. *Management Science*, 6:73–79, 1959.
- [6] D. P. de Farias and B. Van Roy. On constraint sampling in the linear programming approach to approximate dynamic programming. *Mathematics of Operations Research*, 29(3):462–478, 2004.
- [7] L. El Ghaoui, F. Oustry, and H. Lebret. Robust solutions to uncertain semi-definite programs. *SIAM Journal of Optimization*, 9(1):33–52, 1998.
- [8] R. Henrion and W. Römisch. Metric regularity and quantitative stability in stochastic programs with probabilistic constraints. *Mathematical Programming*, 84(1):792–818, 1999.
- [9] P. Kall. *Stochastic Linear Programming*. Springer-Verlag, Berlin, 1976.
- [10] V. Kaňková. On the convergence rate of empirical estimates in chance constrained stochastic programming. *Kybernetika*, 26:449–461, 1990.
- [11] V. Kaňková. A note on relationship between distribution function estimation and estimation in stochastic programming. In *Transactions of the 12th Prague Conference on Information Theory, Statistical Decision Functions and Random Processes*, pages 122–125. Academy of Sciences of the Czech Republic, Prague, 1994.
- [12] A. Prékopa. *Stochastic Programming*. Akadémiai Kiadó, Budapest, 1995.
- [13] S. T. Rachev and W. Römisch. Quantitative stability in stochastic programming: The method of probability metrics. *Mathematics of Operations Research*, 27(4):792–818, 2002.
- [14] W. Römisch. Stability of stochastic programming problems. In A. Ruszczyński and A. Shapiro, editors, *Stochastic Programming. Handbooks in Operations Research and Management Science*, volume 10, pages 483–554. Elsevier, Amsterdam, 2003.

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

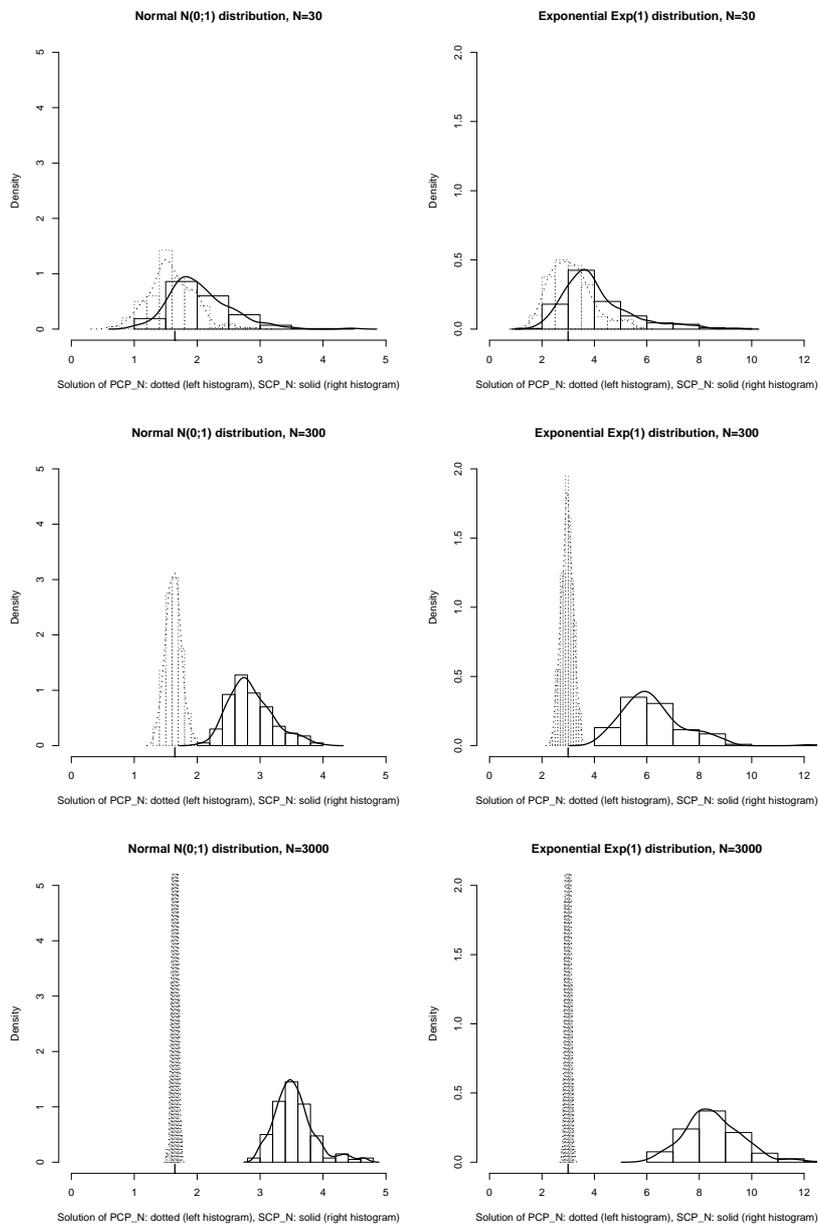


Figure 1: Convergence of optimal values for (SCP_N) and (PCP_N)