Estimation in chance-constrained problem

Michal Houda

Institute of Information Theory and Automation,
Academy of Science of the Czech Republic,
Pod vodárenskou věží 4, CZ-182 08 Praha 8
houda@karlin.mff.cuni.cz

Abstract. Many engineering and economic applications make use of the stochastic programming theory. Major part of models require a complete knowledge of distribution of random parameters, but this assumption is rarely accomplished. We then need to study behaviour of optimal solutions when the distribution changes slightly. In our contribution we consider the chance constrained problem; we recapitulate some known theoretical results about stability and estimation of the problem. We concentrate especially on results from stochastic and robust programming; we try to outline a link between empirical and robust estimates of chance-constrained problem.

1 Introduction

Optimization problems take their important part in both economic and technical disciplines. Input parameters of these problems, modelled as deterministic in the first approach, are random by nature. The randomness of some of them could be freely ignored – to be exact, replaced by some deterministic versions, e.g. average values – without registering any notable modification of resulting solutions. On the other side, there are examples where this is not possible at all – for example, replacing random parameters with their expectations in the classical Kall's linear programming example ([11]) leads to a solution which is not feasible with 75% probability. This is an important task of the post-optimization sensitivity analysis: it has to find a way how to play the game with the randomness.

If the randomness cannot be ignored, we have to incorporate it in the model in a convenient way. Stochastic programming theory is searching for an optimal solution of the optimization model where random parameters are taken into account. If a realization of the random parameter is not known in advance (i. e., before some decision is taken), we cannot work with the random variable itself but its probability distribution. Moreover, a *complete knowledge* of the distribution is required.

[1] This research is supported by the project Nonlinear Dynamical Economic Systems: Theory and Applications registered under No. 402/03/H057, by the project Stochastic Decision Approaches in Nonlinear Economic Models registered under No. 402/04/1294, and by the project Validation of Economic Decision Models and Results registered under No. 402/05/0115 by the Grant Agency of the Czech Republic.

2 Chance-constrained problem

Consider a general constrained problem in the form

$$\min_{x \in X} c'x \quad \text{subject to} \quad f(x;\xi) = \max_{i=1,\dots,m} f_i(x;\xi) \le 0 \tag{1}$$

where $\xi \in \Xi \subset \mathbb{R}^s$ is a random (uncertainty, instance) parameter, $X \subset \mathbb{R}^n$ is a closed convex set, and vector functions $f_i : X \times \Xi \to \mathbb{R}$ are convex for each fixed $\xi \in \Xi$. Without loss of generality, we assume that the objective is linear and independent of ξ.

If a realization ξ of the random parameter is known and fixed, a solution of (1) could be easily obtained by means of deterministic optimization. Two main philosophies are developed in the literature in order to deal with uncertainty of random parameter. First, the *chance-constrained approach* considers ξ as a random variable and checks the constraints in (1) to be fulfilled with a given probability $1 - \beta \in [0;1]$ – the problem then reads

$$\min_{x \in X} c'x \quad \text{subject to} \quad x \in X_\beta := \big\{x \in X;\ \mu\{f(x;\xi) > 0\} \le \beta\big\} \tag{2}$$

where $\mu \in \mathcal{P}(\Xi)$ is the probability distribution of the random vector ξ. The chance-constrained optimization has a long history, starting at least by the work of Charnes and Cooper [7]; an extensive presentation of many results is given in Prékopa's book [12]. Problem (2) is not necessarily a convex optimization problem even if function f is convex in x for all ξ (see [12] again). Another difficulties are arising when we evaluate the probability measure in X_β – it involves a multidimensional integral.

The probability distribution is rarely known completely. Instead, various approximation and estimating techniques are used in practice. Here, let us present an approach using empirical distribution function. Consider a set of iid (independent identically distributed) samples $\xi_1, \dots, \xi_N$ from the original distribution μ. The *empirical distribution function* is a random variable defined by $\mu_N := \frac{1}{N}\sum \delta_{\xi_i}$ where δ_ξ denotes the unit mass at ξ. The problem (2) is now approximated by replacing probability distribution μ by μ_N, and the problem reads

$$\min_{x \in X} c'x \quad \text{subject to} \quad x \in X_N := \big\{x \in X; \frac{1}{N}\operatorname{card}\{i; f(x;\xi_i) > 0\} \le \beta\big\}. \tag{3}$$

Stability theory of stochastic programming answers a question how far is the optimal solution of approximated problem (3) with respect to the optimal solution of the original chance-constrained problem (2). A measurability issues arise when studying convergences properties of the problem. For this reason we establish some notions to overcome these issues. For details, see [13].

A class $\mathcal{F}$ of measurable functions from Ξ to $\bar{\mathbb{R}}$ is called *μ-permissible* if there is a countable subset $\mathcal{F}_0$ of $\mathcal{F}$ such that for each $F \in \mathcal{F}$ there exists a sequence F_j in $\mathcal{F}_0$ converging pointwise to F and such that the integral $\mu F_j := \int_\Xi F_j(\xi)\mathrm{d}\mu(\xi)$

also converge to the integral μF. μ-permissibility of the class is somehow a technical assumption to ensure the measurability of studied objects. Next, define a distance between the probabilities measures $\mu_N(\cdot)$ and μ (Zolotarev pseudometric) by

$$d_{\mathcal{F}}(\mu_N(\cdot),\mu) = d_{\mathcal{F}_0}(\mu_N(\cdot),\mu) = \sup_{F\in\mathcal{F}} |(\mu_N(\cdot)-\mu)F|. \tag{4}$$

$\mathcal{F}$ is a *μ-Glivenko-Cantelli class* if it is μ-permissible and the sequence of distances $\{d_{\mathcal{F}}(\mu_N(\cdot),\mu)\}$ converges to zero almost surely. This is desirable property of the correctly selected probability metric.

Theorem 50 of Römisch work [13] – as a direct consequence of the general stability theorem formulated with respect to a special form of Zolotarev pseudometric – gives us, under some assumptions (μ-permissible class $\mathcal{F}$, metric regularity for X_β, existence and uniqueness of optimal solution) – an upper bound for the convergence rate of distance between the optimal solutions. For example, if $f(x;\xi)$ is linear, then the bound reads

$$\mathrm{P}\big\{\sqrt{N}d(\sup_{x\in X_N^*} d(x;X^*) > \varepsilon\big\} \le \hat{K}\varepsilon^{d_f} e^{-2\min\{\delta,\hat{L}^{-1}\Psi_\mu^{-1}(\varepsilon)\}^2} \tag{5}$$

where X_N^* is the (localised) solution set of estimated problem (3), X^* is the solution set of the chance-constrained problem (2), δ, $\hat{K}$ and $\hat{L}$ are positive constants (arising from the stability theory), $\Psi_\mu^{-1}(\cdot)$ is the inverse associated growth function of the objective (see [13] for a precise definition), and d_f is VC dimension of the constraints $f(\cdot;\cdot)$ (see Section 4). The notable thing is that the approximated solution converge exponentially to the original solution of the chance-constrained problem (2).

3 Robust optimization problem

An alternative approach leading to solve an uncertain program is the so-called "worst-case" analysis which also adopted the name of *robust optimization*. In this approach one looks for a solution which is feasible for all instances of ξ, i. e. solving the problem

$$\min_{x\in X} c'x \quad \text{subject to} \quad f(x;\xi)\le 0 \text{ for all } \xi\in\Xi. \tag{6}$$

This framework was introduced by Ben-Tal and Nemirovski [1], [2], [3] and developed by other authors in various directions (see e. g. [8], [9], [4] and references therein). The robust convex programming problem is convex, but it involves an infinite number of constraints and it is numerically hard to solve again. Robust optimization theory proposes a method of approximation of the problem (6) by the following so-called *sampled problem*; instead of requiring constraints to be satisfied for all instances of ξ, we are satisfied by only finite (but sufficiently large) number of samples $\xi_1,\dots,\xi_N$:

$$\min_{x\in X} c'x \quad \text{subject to} \quad f(x;\xi_i)\le 0 \text{ for } i=1,\dots,N. \tag{7}$$

A question arises for this problem how the conditions of the nominal problem (1) are satisfied. This question of the robust post-optimization analysis can be formulated as follows: we are searching for minimal sample size N such that an optimal solution of (7) is feasible solution of chance constrained problem (2) (i. e. violates conditions of the nominal problem only with a small probability). In this context we speak about two sources of error in the sampling method:

- with a probability α, the feasible (or optimal) solution of the sampled problem (7) is not feasible solution of the chance-constrained problem;
- with a probability β, the feasible solution of (7) can still violate nominal constraints $f(x;\xi) \le 0$.

A result was obtained by Calafiore and Campi [6], [5] for convex functions $f(x;\xi)$. If for given β and α

$$N \ge \frac{2n}{\beta}\ln\frac{1}{\beta} + \frac{2}{\beta}\ln\frac{1}{\alpha} + 2n \tag{8}$$

then the optimal solution of the sampled problem (7) is feasible solution to the chance-constrained problem with the probability $1-\alpha$. Hence the sample size of $\mathcal{O}(n/\beta)$ is enough to achieve this goal.

The complete knowledge of the distribution μ is required in order to generate the samples ξ_i. Again, we only know some approximation to μ in practice. This question is treated by Erdoğan and Iyengar [10]: consider a set of "close" probability distributions $\mathcal{Q} := \{\mu;\ \rho_P(\mu;\mu_0) \le \delta\}$, where $\delta > 0$ is a prescribed constant, $\mu_0 \in \mathcal{P}(\Xi)$ a suitable *central* probability measure and ρ_P is the Prohorov metric defined on $\mathcal{P}(\Xi)$ by

$$\rho_P(\mu;\nu) := \inf\{\varepsilon;\ \mu B \ge \nu(B^\varepsilon) + \varepsilon \ \text{ for all measurable } B \subset \Xi\} \tag{9}$$

where $B^\varepsilon := \{x \in X : \ \inf_{z\in B}\|x-z\| \le \varepsilon\}$. Prohorov metric is a smallest distance "in probability" between two random variables with the distributions μ and ν. See again [10] and references therein for relations to other probability metrics.

Now we are ready to formulate a so-called *ambiguous chance-constrained problem*; the constraint now have to be satisfied with a high probability for all sufficiently close measures:

$$\min_{x\in X} c'x \quad \text{subject to} \quad x \in \bar{X}_\beta := \big\{x \in X : \ \mu\{f(x;\xi) > 0\} \le \beta \quad \forall \mu \in \mathcal{Q}\big\} \tag{10}$$

We approximate this problem by a *robust sampled problem* defined as follows

$$\min_{x\in X} c'x \quad \text{subject to} \quad f(x;z) \le 0 \quad \forall z : \|z-\xi_i\| \le \delta,\ i = 1,\dots,N \tag{11}$$

and again try to find a sufficient sample size N ensuring that an optimal solution of the robust sampled problem (11) is feasible for ambiguous chance-constrained problem (10). This has been done in [10]: the optimal solution of the robust sampled problem (11) is feasible solution to the ambiguous chance-constrained problem (10) with the probability $1-\gamma$ if the sample size is of order $\mathcal{O}(n/(\beta-\gamma))$.

4 VC-dimension and computational learning theory

Another question of robust optimization is the behaviour of all the feasibility set of the problems (7) and (11), not only of their optimal solutions. Here we need in addition a notion of Vapnik-Chervonenkis (VC) dimension and computational learning theory (see e. g. [14]).

Let $x \in X$ be a decision vector, denote $C_x := \{\xi \in \Xi : \ f(x;\xi) \leq 0\}$ (*concept* or *classification*) a set of instances which are "feasible" for this decision. Learning X_β means to find concepts $C_f := \{C_x; \ x \in X\}$ with $\mu(\xi \notin C_x) \leq \beta$, i. e. (in real world), to approximate $f(x;\cdot)$ with a finite number of samples. Denote this set of samples by $S := \{\xi_1, \dots, \xi_N\} \subset \Xi$ and denote $\Pi_f(S) := \{(I_{C_x}(\xi_1), \dots, I_{C_x}(\xi_N)\}$ a set of dichotomies induced by C_f ($I_A(\cdot)$ is an indicator function of the set A). If $\operatorname{card} \Pi_f(S) = 2^N$ we say that S is *shatered* by C_f – this means that $\Pi_f(S)$ contains all possible results and hence no information is given about C_f. *Vapnik-Chervonenkis dimension* d_f of $f(\cdot;\cdot)$ (or C_f) is the maximal cardinality of S that is shatered by C_f.

VC dimension measures the complexity of concepts C_f. We usually assume that $d_f < +\infty$ – this means that some information is finally given with a sufficiently large number of samples. But the VC dimension could be rather large, as seen for example in the third case below:

- in the linear case with a single constraint, $d_f \leq n$;
- in the linear case with multiple constraints, $d_f \leq \mathcal{O}(n^2)$;
- in the case of pointwise maximum from (1), $d_f \leq \mathcal{O}(10^m \max_i d_{f_i})$.

To end the robust programming part of the paper, let apply this construct to our questions of feasibility; one can deduce following sample size orders (see [10]):

- the worst case number of samples required to learn X_β is $\mathcal{O}(d_f/\beta)$.
- the worst case number of samples required to learn $\bar{X}_\beta$ is $\mathcal{O}(d_f/(\beta-\gamma))$.

5 Conclusion and open questions

We have made a short review of two different approaches to estiomation in uncertain optimization problems where some input parameters are considered as random variables. First, ve have mentioned the chance-constrained problems and its approximation through empirical distribution functions. Known stability results relate to the difference between original optimal solution and the optimal solution of the approximated problem. Second approach is based on the robust optimization theory: we replace an infinite number of constraints (required to be satisfied for all instances of random parameters) by a finite one (constraints are satisfied by finite number of samples). Robust optimization then looks for the sample size which is needed to satisfy the constraints with a high probability, instead to measure distance between original and approximated optimal solution.

Some questions arise and are planned to be explored in the future research:

– how the optimal values of (robust) sampled problem relate to the optimal value of the chance-constrained problem or its approximation;
– both result have in common the dependence on the complexity of constraints expressed via VC dimension; it is of interest to estimate an upper bound of VC dimensions for some special types of constraints and problems (e. g. quadratic ones);
– if the independence of samples is crucial and if there is a way to use somehow dependent samples;
– probability metrics can be adopted to the problem via stochastic programming stability theorems; this way we could modify the robust sampled problem for ambiguous chance-constrained problem and make some conclusions about achieved results.

References

1. Ben-Tal, A., and Nemirovski, A.: Robust convex optimization. Mathematics of Operations Research **23** (4) (1998), 769–805.
2. Ben-Tal, A., and Nemirovski, A.: Robust solutions of uncertain linear programs. Operations Research Letters **25** (1) (1999), 1–13.
3. Ben-Tal, A., and Nemirovski, A.: Lectures on Modern Convex Optimization. SIAM, Philadelphia, PA (2001).
4. Bertsimas, D., and Sim, M.: Robust conic optimization. To appear in Mathematical Programming.
5. Calafiore, G., and Campi, M. C.: Decision making in an uncertain environment: the scenario based optimization approach. In: Multiple Participant Decision Making (J. Andrýsek, M. Kárný, and J. Kracík, eds.), Advanced Knowledge International (2004), 99–111.
6. Calafiore, G., and Campi, M. C.: Uncertain convex programs: randomized solutions and confidence levels. Mathematical Programming **102** (1, Ser. A) (2005), 25–46.
7. Charnes, A., and Cooper, W. W.: Chance constrained programming. Management Science **6** (1959), 73–79.
8. El Ghaoui, L., and Lebret, H.: Robust solutions to least-squares problems with uncertain data. SIAM Journal of Matrix Analysis Applications**18** (4) (1997), 1035–1064.
9. El Ghaoui, L., Oustry, F., and Lebret, H.: Robust solutions to uncertain semidefinite programs. SIAM Journal of Optimization**9** (1) (1998), 33–52.
10. Erdoğan, E., and Iyengar, G.: Ambiguous chance constrained problems and robust optimization. CORC Technical Report TR-2004-10, Columbia University (2004). To appear in Mathematical Programming.
11. Kall, P.: Stochastic Linear Programming. Springer-Verlag, Berlin (1976).
12. Prékopa, A.: Stochastic Programming. Kluwer Academic Publishers, Dordrecht, and Akadémiai Kiadó, Budapest (1995).
13. Römisch, W.: Stability of stochastic programming problems. In: Stochastic Programming (A. Ruszczyński, and A. Shapiro eds.), Handbooks in Operations Research and Management Science Vol. 10, Elsevier, Amsterdam (2003), 483–554.
14. Vapnik, V. N.: The Nature of Statistical Learning Theory. Springer Verlag, New York (1995).