Approximations in stochastic and robust programming problems

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

Abstract
Optimization procedures are very useful tools in many economic decision-making problems. We deal with the case of optimization problems where uncertainties in parameters occur. The stochastic programming approach considers the probability distribution of uncertain parameters and seeks for a solution that is feasible up to a certain level of probability (chance-constrained programming). Robust programming techniques search for such a solution that satisfies simultaneously all possible realizations of the parameters. Both methods require some kind of approximation because of computational difficulties. The paper deals with such approximations and illustrates the essential difference between the two above-mentioned methods. Even if a variety of economic problems lead to the same optimization program, one is required to choose a correct method to solve it; the economic background of the problem is crucial for such decision.

Keywords
Chance-constrained problem, empirical distribution function, robust optimization problem, sampled problem.
JEL: C44, C61

1 Introduction: uncertainty of the data

The traditional optimization research became an integral part of the post-war science and many advances and applications in various fields of the area were obtained, including applications in finance, engineering, management, control, etc. The real world carries the uncertainty of the data as a generic property of all the models of mathematical programming. There are many ways to handle the uncertainty and to give applicable results of the optimization procedures.

Consider an optimization problem of the form

\[
\text{minimize } c(x; \xi) \text{ subject to } x \in X, f(x; \xi) \leq 0
\]

where \( \xi \subseteq \mathbb{R}^s \) is a data element of the problem, \( x \in X \subseteq \mathbb{R}^n \) is a decision vector, the dimensions \( n, s, M \) and the mappings \( c : \mathbb{R}^n \times \mathbb{R}^s \to \mathbb{R} \) and \( f : \mathbb{R}^n \times \mathbb{R}^s \to \mathbb{R}^M \) are structural elements of the problem. This is a general framework for a large class of optimization problems which we characterize further by

- insufficient knowledge of the data; all that is known about the data vector \( \xi \) is that it belongs to a given uncertainty set \( \Xi \subseteq \mathbb{R}^s \);
- the constraints of problem (1) are required to be satisfied as much as possible given the actual realization of \( \xi \in \Xi \).

If a realization of \( \xi \) is known and fixed in advance, standard deterministic optimization algorithms can be used to solve problem (1). This is rarely the case; in practice, uncertainty of the data is typical in the modelling framework, for example:

- the data \( \xi \) is not known at the time when the decision (value of \( x \)) have to be made, and will be realized in the future (the data can represent future demands and prices in economy, loads to the bridge in truss construction, weather conditions, etc.);
- the data \( \xi \) cannot be measured or estimated exactly even if it is realized before a concrete decision is taken (material properties, measuring errors, etc.);
- the data is certain and the optimal solution of the problem can be computed exactly, but such solution cannot be implemented exactly due to physical characteristics of the solution (e.g. uncertain production of some commodity, properties of construction, etc.). The last can be easily modeled via uncertainty in the parameters of the model, not in the decision vector \( x \);
\begin{itemize}
  \item the model itself is an approximation of a complicated real-world phenomenon and uncertainty comes directly from the modeling process.
\end{itemize}

Dealing with uncertainty is a kind of bread-and-butter problems that classical optimization try to solve. Several approaches were developed. First, the uncertainty is simply ignored at the stage of building the model and/or finding an optimal solution of it. The data is replaced by some nominal values (e.g., averages, expected values) and the accuracy of the optimal solution is (or should be) inspected ex-post by sensitivity analysis. This is a traditional way to control the stability of the model but it is limited only to an already generated solution. There are examples where the “ignoring uncertainty” approach leads to a solution that is not acceptable in practice (see e.g., Kall’s linear programming example, [15]).

Stochastic programming handles the uncertainty of stochastic nature. More precisely, we consider \( \xi \) to be a random vector and assume that we are able to identify its underlying probability distribution. The idea of stochastic programming approach is to incorporate available information about data through its probability distribution and solve the new model by means of deterministic optimization (the new model was said to be a “deterministic equivalent” in early works on stochastic programming). There are various ways of doing that and there are many papers and books dealing with particular branches of stochastic programming. The stochastic programming community recognizes Dantzig’s paper [7] as the initial work in the area; there are also a large number of books devoted to stochastic programming and its applications ([4], [16], [19], [23], [24], and others).

The concept of so-called robust optimization does not have such a long history. It introduces an alternative way to handle uncertainty in the model by the so-called “worst-case” analysis: we look for such a solution that satisfies the constraints for all possible realizations of \( \xi \) and we optimize the worst-case objective function among all robust solutions. Even if such paradigm is classical in statistical decision theory, the real development in this area of optimization dates only to the last decade starting with the work [2]. On the other hand, robust optimization problems are not new (they are a part of semi-infinite programming problems); also the influence of the robust control theory is evident and not negligible.

2 Mathematical model of uncertainty

2.1 Uncertain convex program

An uncertain convex program (UCP) is a family of convex optimization programs (1) parameterized by \( \xi \in \Xi \).

Without loss of generality we consider the following form of (UCP):

\[
\text{minimize } c^t x \text{ subject to } x \in X, \; f(x; \xi) \leq 0, \tag{2}
\]

where \( \xi \in \Xi \subset \mathbb{R}^n, \; X \subset \mathbb{R}^n \) is convex and closed set, the objective is linear and the scalar-valued function \( f : X \times \Xi \rightarrow \mathbb{R} \) is convex in \( x \) for all \( \xi \in \Xi \). In fact, the linearity of the objective can be imposed by considering the problem minimize \( t \) subject to \( x \in X, \; c(x; \xi) \leq t, \; f(x; \xi) \leq 0 \) instead of problem (1); and multiple valued convex constraint functions \( f_i(x; \xi) \) can be converted into a single scalar-valued function of the form \( f(x; \xi) := \max_{i=1, \ldots, M} f_i(x; \xi) \). If the realization of \( \xi \) is known and fixed, we use the deterministic optimization to solve problem (2). This corresponds to the approach of ignoring uncertainty as described above. In many cases, such solution is very sensitive to perturbations of \( \xi \) and one of the following methods must be used.

2.2 Chance (probability) constrained program

A chance (or probability) constrained program (PCP) is a particular variant of stochastic programming problem. We assume that \( \xi \) is a random vector defined on some probability space \( (\Omega, \mathcal{A}, \mathbb{P}) \) with known probability distribution \( \mathbb{P} \in \mathcal{P}(\Xi) \) where \( \mathcal{P}(\Xi) \) is the space of all probability measures defined on \( \Xi \). Next, we require the constraints of (UCP) to be fulfilled with a prescribed level of probability \( \varepsilon \). The problem reads

\[
\text{minimize } c^t x \text{ subject to } x \in X_\varepsilon := \{ x \in X; \; P\{ f(x; \xi) > 0 \} \leq \varepsilon \}. \tag{3}
\]

The feasible solution of this problem allows to violate the original constraints with a small level of “risk”; it is sometimes called \( \varepsilon \)-feasible solution. The problem (PCP) need not be convex even if \( f \) is convex in \( x \) for all \( \xi \); another issue arises when we try to evaluate the probability in the definition of \( X_\varepsilon \) because it usually involves multidimensional integrals. One of the possible approaches overcoming this issue is the subject of Section 3.

The term of “chance-constrained” optimization is coined with an early work of Charnes and Cooper [6]; most of the relevant literature is resumed in [19]. Main results of the topic concern conditions under which (3) is a convex program and how to convert the probability constraints into an explicit deterministic form.
2.3 Robust convex program

In the robust convex program (RCP) we look for a solution to optimization program (2) that satisfies the constraints for all possible realizations of $\xi$, i.e., that is feasible for any member of problems belonging to the family (UCP). The problem can be rewritten as

$$\text{minimize } c'x \text{ subject to } x \in X, f(x; \xi) \leq 0 \text{ for all } \xi \in \Xi. \quad (4)$$

The symbol $\Xi$ is overloaded here: first, in (2), it represents the uncertainty set which the (unknown) parameter $\xi$ belongs to, and second, in (4), it is the set of parameters for which the constraints must be fulfilled. In fact, this overloading is not too much important: the two sets usually coincide for the reason that information we have about the uncertain parameter is the same as the risk we want to hedge against.

(PCP) is a convex program but it is numerically hard to solve because of an infinite number of constraints. There are several relaxation techniques to deal with this issue, see e.g. [2], [3], and references therein. We describe a so-called “randomized” approach [5] in the following.

3 Approximation to stochastic and robust optimization programs

The probability distribution $P$ of $\xi$ is rarely known precisely as required by the chance-constrained methods solving (3). Similarly, numerically intractable problems arising from (4) are common in practice. Sampling techniques are useful in this context for both approaches dealing with the uncertain convex program, stochastic and robust programming. Let us introduce ideas of sampling for both presented approaches.

3.1 Chance-constrained “sampled” problem

Let $\xi_1, \ldots, \xi_N$ be a set of independent samples distributed according to $P$ (distribution of the random vector $\xi$). We define the empirical distribution function as a discrete random vector of the form

$$P_N := \frac{1}{N} \sum_{i=1}^{N} \delta_{\xi_i},$$

where $\delta_{\xi_i}$ denotes the Dirac measure placing the unit mass at $\xi$. We approximate, for the given sample, the problem (PCP) by replacing the original distribution $P$ in (3) by $P_N$. We denote the resulting problem as (PCP$_N$):

$$\text{minimize } c'x \text{ subject to } x \in X[\varepsilon, N]:= \{ x \in X; \frac{1}{N} \text{ card}\{i; f(x; \xi_i) > 0\} \leq \varepsilon \}. \quad (5)$$

The main idea behind this program is that the relative frequency of constraint violations is approximately the desired level $\varepsilon$ of infeasibility in (PCP). An extensive literature on stability in stochastic programming deals with the question how far the resulting optimal solution of (PCP$_N$) is from the optimal solution of (PCP), see e.g. [12], [17], [18], [21], [22] and others. If the assumptions of the general stability theorem (Theorem 1 in [12]) are fulfilled, the distance between optimal solutions of the two problems are expected to converge to zero.

3.2 Robust sampled convex problem

Consider again $\xi_1, \ldots, \xi_N$ a set of independent samples from a given probability distribution $P$. Note that the original (RCP) problem does not involve any information about stochastic nature of parameter $\xi$. This is why this approach adopted the name randomized program or robust sampled convex program (SCP$_N$):

$$\text{minimize } c'x \text{ subject to } X[N]:= \{ x \in X; f(x; \xi_i) \leq 0 \text{ for all } i = 1, \ldots, N \}. \quad (6)$$

This is a relaxation of the original robust problem: we does not require the original constraints to be satisfied for all realizations of $\xi \in \Xi$ but only for a certain finite but sufficiently large number of samples. This approach has several favorable impacts:

- the problem is convex, it has a finite number of constraints and it is effectively computable;
- it incorporates weights to the individual parameter instances of $\xi$ – their absence was also the subject of criticism of the common robust framework;
- in addition, realizations of $\xi$ used in (SCP$_N$) are that which are most probably to happen.
The randomized approach to (RCP) was proposed in [5] and [8]; in [9] the idea was extended to the case of the so-called ambiguous chance-constrained programming where the distribution \( P \) is known only approximately. For a survey of these results see [13].

The solution of (6) approximates the solution of (RCP): the higher the number of samples, the closer the solutions are. In order to find an optimal solution of (SCP\(_N\)) that is sufficiently close to the optimal solution of (RCP) one need a rather high number of samples to be generated. The authors of above-cited papers have sought for a rule on the sample size \( N \) that assures the optimal solution of (SCP\(_N\)) to be \( \varepsilon \)-feasible, i.e. feasible in the chance-constrained problem (PCP). But one cannot expect that this solution is near to the optimal solution of (PCP). In fact, in [14] we gave a comparative numerical study on a simple optimization problem where this conclusion is approved. We develop this idea in detail from the economical and practical point of view in the following section.

4 Application issues of stochastic and robust programming problems

4.1 Simple example

The following example is taken from [14] where it is examined in more details. Consider a simple uncertain convex program

\[
\text{minimize } x \text{ subject to } x \geq \xi, \quad x \in X
\]

where \( \xi \in \mathbb{R} \) is distributed according to the standard normal distribution. The solution to the chance-constrained program (PCP) related to (7) with \( \varepsilon = 0.05 \) is 0.95 quantile of normal distribution (approximately 1.64). The number of samples assuring that the optimal solution of (SCP\(_N\)) is \( \varepsilon \)-feasible is about 240. Figure 1 shows how optimal values of both (PCP\(_N\)) (dotted line) and (SCP\(_N\)) (solid line) behave for sequences of 240 and 3000 samples. On the one hand you should note the convergence property of solutions of (PCP\(_N\)) to the optimal value of (PCP); on the other hand robust sampled solutions are getting away from this point as far as the number of samples increases. We are going to discuss this feature in a practical point of view.

4.2 Applications of the chance-constrained problems

There is a huge number of applications in stochastic programming due to the long history of the subject. A collection of the most important ones is given in the Wallace and Ziemba’s book [24]. We give here only a short overview of selected particular tasks that was solved in real-world applications. Other items include applications in agriculture, power generation and electricity distribution, military, production control, telecommunications, transportation and many others.

- Chemical engineering ([11]). A continuous distillation process is frequently very dependent on a controlled rate of its inflow; if the last is of stochastic nature, it cannot be processed immediately but has to be stored in a feed tank. The objective is to find the optimal feed control with the prescribed lower and upper level of the inflow preventing the feed tank to be empty or full, together with the fact that costs compensating possible level violations are difficult to model.
• Finance – portfolio selection. The objective is to select the optimal portfolio of bonds in order to maximize final amount of money and to cover necessary payments in all years. The last is modeled via liquidity constraints we want to satisfy with some high level of probability.

• Water management ([20]): one of the very beginning application of chance-constrained problems. A number of reservoirs must be designed in order to control flooding due to random stream inflows.

These models have in common that we estimate the probability distribution (needed to solve the optimization problem) by means of observations from the past. The resulting solution in (PCP_N) is an approximation to the (unknown) solution of (PCP) and the approximation is better as the number of samples (observations) is higher. Furthermore, our solution of the chance-constrained “sampled” problem is only approximatively $\varepsilon$-feasible for a given level $\varepsilon$. On the other hand, this level is usually not crucial for real applications if our preferences are pointed towards costs saving solutions – this is also the case of all mentioned applications.

4.3 Applications of robust sampled problem

The number of real-world applications of the robust programming is a little more sparse. The most important are the robust truss topology design and the robust portfolio selection problems; there are also other applications, especially in finance (e.g. option modeling), management (supply chain management), or engineering (power supply).

• Robust Truss Topology Design ([1]). The problem is to select the optimal configuration of a structural system (mechanical, aerospace, ...) that is subjected to one or more given loads (nominal loads) and an unspecified set of small uncertain loads. The goal is to find such configuration that the construction is rigid to all of the prescribed loads.

• Robust portfolio selection ([10]). Here, the uncertain parameters are the modeling errors in the estimates of market parameters and they are assumed to lie in a known and bounded uncertainty set. The robust portfolio is the solution to an optimization problem where the worst-case behaviour of parameters is assumed.

The optimal solutions to robust problems hedge against the worst-case realization of uncertain parameters regardless their “importance”. The randomized (sampled) approach incorporate the information about importance to the model via probability distribution of samples so that the optimal solution of the sampled problem does not have to satisfy the constraints for all possible realizations of parameter. At the same time, the probability of such violation is small and for a given $\varepsilon$ one could easily compute the number of samples to generate in order to have an $\varepsilon$-feasible solution. Indeed, if the number of samples is significantly greater than required, the optimal solution of the sampled problem also hedges against the parameters with the smaller probability of occurrence. This could be the task if the risk of constraint violation has to be minimized as much as possible and costs of doing that are of smaller importance. This is usually the case of the truss constructions mentioned above.

5 Conclusion

Many economic and engineering applications lead to the same optimization model that incorporate uncertainty parameter. One can handle this parameter in a variety of ways: including ignoring uncertainty, using stochastic or robust programming. But the decision about how to deal with uncertainty cannot be left out of an economical analysis – the desired optimal solution of the problem is closely related to the economical and managerial background of the problem.

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.

References


