

Introduction to Optimization with Stochastic Uncertainties

Irina Nishakova

14. Juni 2010

Preface

The area of stochastic programming was created in the middle of the last century, following fundamental achievements in linear and nonlinear programming. While it has been quickly realized that the presence of uncertainty in optimization models creates a need for new problem formulations, many years have passed until the basic stochastic programming models have been formulated and analyzed. Today, stochastic programming theory offers a variety of models to address the presence of random data in optimization problems: chance-constrained models, two- and multistage models, etc. It is based on advanced mathematical tools such as abstract optimization, probability theory, statistical techniques.

1 Motivation¹

Technically, stochastic programs are much more complicated than the corresponding deterministic programs. Hence, at least from a practical point of view, there must be very good reasons to turn to the stochastic models.

We start with an example illustrating that these reasons exist. In fact, we shall demonstrate that alternative deterministic approaches do not even look for the best solutions. Deterministic models may certainly produce good solutions for certain data set in certain models, but there is generally no way you can conclude that they are good, without comparing them to solutions of stochastic programs. In many cases, solutions to deterministic programs are very misleading.

1.1 A numerical example

You own two lots of land. Each of them can be developed with necessary infrastructure and a plant can be built. In fact, there are nine possible decisions (see Figure 1).

In each of the plants, it is possible to produce one unit of some product. It can be sold at a price p . The price p is unknown when the land is developed. The cost structure is given in the Table 1.

	Developing the land	Building the plant	Building the plant later
Lot 1	600	200	220
Lot 2	100	600	660

Table 1. The cost structure. For each lot of land we give the cost of developing the land and building the plant before and after p becomes known.

¹ This section is taken from [1].

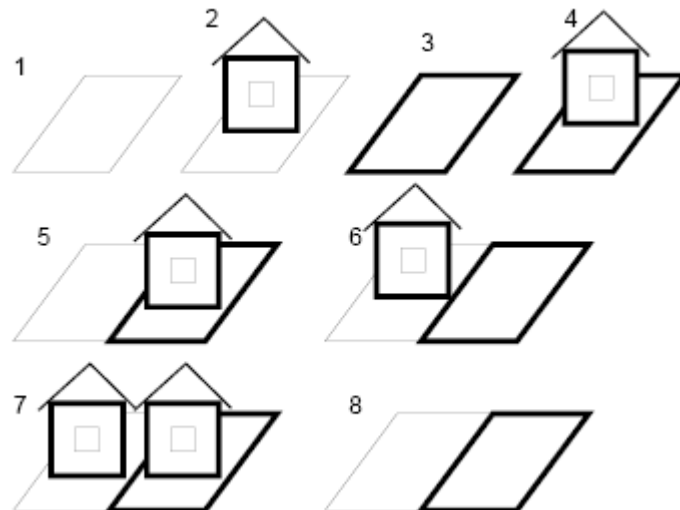


Figure 1 Eight of the nine possible decisions. The areas surrounded by thin lines correspond to Lot 1, the areas with thick lines to Lot 2. For example, Decision 6 is to develop both Lots, and build a plant on Lot 1. Decision 9 is to do nothing.

Also, if the plant on Lot 1, say, is to be built at its lowest cost, given as 200 in the table, that must take place before p becomes known. However, it is possible to delay the building of the plant until after p becomes known, but at a 10% penalty. That is given in the last column of the table. This can only take place if the lot is already developed. There is not enough time to both develop the land *and* build a plant after p has become known.

1.2 Scenario analysis

A common way of solving problems of this kind is to perform scenario analysis, also sometimes referred to as simulation. The idea is to construct or sample possible futures (values of p in our case) and solve the corresponding problem for these values. After having obtained a number of possible decisions this way, we either pick the best of them or try to find good combinations of the decisions.

In our case it is simple to show that there are only three possible scenario solutions. These are given as follows:

Interval for p	Decision number*
$p < 700$	9
$700 \leq p < 800$	4
$p \geq 800$	7

*Decision numbers refer to Figure 1.

So whatever scenarios are constructed or sampled, these are the only possible solutions. Now, assume for simplicity that p can take on only two values, namely 210 and 1250, each with a probability of 0.5. This is a very extreme choice, but it has been made only for convenience. We could have made the same points with more complicated (for example continuous) distributions, but nothing would have been gained by doing that, except make the calculations more complicated. Hence, the expected price equals 730.

1.3 Using the expected value of p

A common solution procedure for stochastic problems is to use the expected value of all random variables. This is sometimes done very explicitly, but more often it is done in the following fashion: The modeler collects data, either by experiments or by checking an existing process over time, and then calculates the mean, which is then said to be the best available estimate of the parameter. In this case we would then use 730, and from the list of scenario solutions above, we see that the optimal solution will Decision 4 with a profit of

$$-700 + 700 = 30.$$

We call this the expected value solution. We can also calculate the expected value of using the expected value solution. That is, we can use the expected value solution, and then see how it performs under the possible futures. We get

$$-700 + \frac{1}{2}210 + \frac{1}{2}1250 = 30.$$

1.4 Maximizing the expected value of the objective

We can also calculate the expected value of using any of the possible scenario solutions. We find that for doing nothing (Decision 9), the expected value is 0, and for Decision 7 the expected value equals

$$-1500 + \frac{1}{2}420 + \frac{1}{2}2500 = -40.$$

In other words, the expected value solution is the best of the three scenario solutions in terms of having the best expected performance. But is this the solution with the best expected performance? Let us answer this question by simply listing all possible solutions, and calculate their expected value. In all cases, if the land is developed before p becomes known, we will consider the option of building the plant at the 10% penalty if that is profitable. The results are given in Table 2.

Decision	Investment	Income if $p = 210$	Income if $p = 1250$	Expected profit
1	-600	-	$\frac{1}{2}1030$	-85
2	-800	$\frac{1}{2}210$	$\frac{1}{2}1250$	-70
3	-100	-	$\frac{1}{2}590$	195
4	-700	$\frac{1}{2}210$	$\frac{1}{2}1250$	30
5	-1300	$\frac{1}{2}210$	$\frac{1}{2}2280$	-55
6	-900	$\frac{1}{2}210$	$\frac{1}{2}1840$	125
7	-1500	$\frac{1}{2}420$	$\frac{1}{2}2500$	-40
8	-700	-	$\frac{1}{2}1620$	110
9	0	0	0	0

Table 2 The expected value of all nine possible solutions. The income is the value of the product if the plant is already built. If not, it is the value of the product minus the construction cost at 10% penalty.

As we see from Table 2, the optimal solution is to develop Lot 2, then wait to see what the price turns out to be. If the price turns out to be low, do nothing, if it turns out to be high, build plant 2. The solution that truly maximizes the expected value of the objective function will be called the *stochastic solution*. Note that also two more solutions are substantially better than the expected value solution. All three solutions that are better than the expected value solution are solutions with *options* in them. That is, they mean that we develop some land in anticipation of high prices. Of course, there is a chance that the investment will be lost. In scenario analysis, as outlined earlier, options have no value, and hence, never show up in a solution. It is important to note that the fact that these solutions did not show up as scenario solutions is *not* caused by few scenarios, but by the very nature of a scenario, namely that it is deterministic. It is incorrect to assume that if you can obtain enough scenarios, you will eventually come upon the correct solution.

1.5 Hindsight

In hindsight, that is, after the fact, it will always be such that one of the scenario solutions turn out to be the best choice. In particular, the expected value solution will be optimal for any $700 < p \leq 800$. (We did not have any probability mass there in our example, but we could easily have constructed such a case.) The problem is that it is not the same scenario solution that is optimal in all cases. In fact, most of them are very bad in all but the situation where they are best. The stochastic solution, on the other hand, is normally *never* optimal after the fact. But, at the same time, it is also hardly ever really bad. In our example, with the given probability distribution, the decision of doing nothing (which has an expected value of zero) and the decision of building both plants (with an expected value of -40) both have a probability of 50% of being optimal after p has become known. The stochastic solution, with an expected value of 195, on the other hand, has zero probability of being optimal in hindsight. This is an important observation. If you base your decisions on stochastic models, you will normally never do things really well. Therefore, people who prefer to evaluate after the fact can always claim that you made a bad decision. If you base your decisions on scenario solutions, there is a certain chance that you will do really well. It is therefore possible to claim that in certain cases the most risky decision one can make is the one with the highest expected value, because you will then always be proven wrong after the fact.

2 Models

2.1 Uncertainties in the objective function²

In deterministic optimization, a decision x must be found, which minimizes a known cost function $f(x)$ among all possible candidates x lying in the feasible set D

$$\min f(x), \quad x \in D.$$

In stochastic optimization, the cost function is not exactly known at the time when the decision is made. Only a stochastic model $F(x, \zeta)$ for the costs is known, where ζ is some random vector. The crucial point is that although the particular value of ζ is unknown, its distribution is completely known. This situation is usually called the *uncertainty problem*. Since we do not know the actual value of ζ , we cannot minimize the cost function for each value of ζ separately, but have to minimize some real functional \mathbb{F} , which summarizes the random costs $F(x, \zeta)$ in an appropriate manner.

² This section is taken from [2].

The uncertainty problem reads

$$\min \{f(x) = \mathbb{F}[F(x, \xi)]\}, \quad x \in D.$$

Any functional \mathbb{F} , which maps a set of distribution functions on \mathbb{R} onto the real line may be used as summarizing functional. We call such functionals *risk functionals*. Examples for risk functional are the expectation, the variance, the median, etc.

In case of the *expectation functional* \mathbb{E} as the summarizing functional the problem reads as

$$\min \{f(x) = \mathbb{E}[F(x, \xi)]\}, \quad x \in D.$$

The above formulation of a stochastic programming problem assumes implicitly that the expected value is taken with respect to a known probability distribution P on (Ω, \mathcal{F}) , where Ω is a sample space equipped with a sigma algebra \mathcal{F} , and that the expected value operator

$$\mathbb{E}[F(x, \xi)] = \int_{\Omega} F(x, \xi) dP(\xi)$$

is well defined. We call the function $f(x)$ the *expectation* or *expected value* function.

In some cases the expectation can be explicitly computed, but generally for the numerical treatment one has to approximate this integral. The approximation depends on the distribution of random variable, for example, for multivariate truncated normal distribution the integral can be efficiently evaluated by a Gaussian quadrature³, where the quadrature points are the roots of a polynomial belonging to a class of orthogonal polynomials. The other way to simplify the complex problem is to apply a Taylor series expansion about a nominal set point ξ^0 which is at the same time the expected value of the random variable ξ . The Taylor approximation of second order of F gives

$$\hat{F}(x, \xi) := F(x, \xi^0) + \frac{\partial F(x, \xi^0)}{\partial \xi} (\xi - \xi^0) + \frac{1}{2} (\xi - \xi^0)^T \frac{\partial^2 F(x, \xi^0)}{\partial \xi^2} (\xi - \xi^0).$$

Integrating this, we observe

$$\int_{\Omega} \hat{F}(x, \xi) dP(\xi) = F(x, \xi^0) + \frac{1}{2} \sum_{i=1}^k \frac{\partial^2 F(x, \xi^0)}{\partial \xi_i^2} \text{Var}(\xi_i)$$

where $\text{Var}(\xi_i)$ is the variance of the i -th component of ξ . This representation leads to the deterministic optimization problem

$$\min F(x, \xi^0) + \frac{1}{2} \sum_{i=1}^k \frac{\partial^2 F(x, \xi^0)}{\partial \xi_i^2} \text{Var}(\xi_i).$$

The assumption that the distribution P , i.e., $P(A) = \mathbb{P}\{\xi \in A\}$ of the random part ξ of the cost function is exactly known is rarely fulfilled. Typically, the probability measure P is unknown (this is called the *ambiguity problem*) and only some information about it is available.

In the ambiguity situation, the stochastic optimization problem gets an additional difficulty: the unknown probability P has to be guessed through the available information and this gives us an additional source of error.

³ The source is the paper of V. Schulz and C. Shillings, "On the nature and treatment of uncertainties in aerodynamic design", University of Trier, Germany.

Example⁴

Now we illustrate the difference between the solutions for different distributions and also the difference between stochastic and worst-case formulations of the robust approximation problem. Consider the least-squares problem

$$\min \|A(u)x - b\|_2^2,$$

where $u \in \mathbf{R}$ is an uncertain parameter and $A(u) = A_0 + uA_1$. We consider a specific instance of the problem, with $A(u) \in \mathbf{R}^{20 \times 10}$, $\|A_0\| = 10$, $\|A_1\| = 1$, and u in the interval $[-1, 1]$. (So, roughly speaking, the variation in the matrix A is around $\pm 10\%$.)

We find four approximate solutions:

- *Nominal optimal.* The optimal solution x_{nom} is found, assuming $A(u)$ has its nominal value A_0 :

$$x_{nom} = (A_0^T A_0)^{-1} A_0^T b.$$

- *Stochastic robust approximation with different distributions.* We find x_{stoch1} , which minimizes $\mathbf{E} \|A(u)x - b\|_2^2$, assuming the parameter u is **uniformly distributed** on $[-1, 1]$. The solution is:

$$x_{stoch1} = \left(A_0^T A_0 + \frac{1}{3} A_1^T A_1 \right)^{-1} A_0^T b.$$

And also we find x_{stoch2} , which minimizes $\mathbf{E} \|A(u)x - b\|_2^2$, assuming the parameter u is **normally distributed** on $[-1, 1]$ (truncated normal distribution) with zero mean and assuming the variance is equal to 1. In this case the solution is:

$$x_{stoch2} = (A_0^T A_0 + A_1^T A_1)^{-1} A_0^T b.$$

- *Worst-case robust approximation.* The solution x_{wc} minimizes

$$\sup_{-1 \leq u \leq 1} \|A(u)x - b\|_2 = \max\{\|(A_0 - A_1)x - b\|_2, \|(A_0 + A_1)x - b\|_2\}.$$

This is an SOCP since it is equivalent to:

$$\begin{aligned} & \min t \\ & \text{subject to } \|(A_0 - A_1)x - b\|_2 \leq t, \\ & \qquad \qquad \|(A_0 + A_1)x - b\|_2 \leq t. \end{aligned}$$

For each of these four values of x , we plot the residual $r(u) = \|A(u)x - b\|_2$ as a function of the uncertain parameter u , in figure 1. These plots show how sensitive an approximate solution can be to variation in the parameter u . The nominal solution achieves the smallest residual when $u = 0$, but is quite sensitive to parameter variation: it gives much larger residuals as u deviates from 0, and approaches -1 or 1 .

The worst-case solution has a larger residual when $u = 0$, but its residuals do not rise much as u varies over the interval $[-1, 1]$. The solution for normally distributed parameter with variance equal to one gives us very similar result. The stochastic robust approximate solution for uniform distribution is in between.

⁴ The problem in this example is taken from [4].

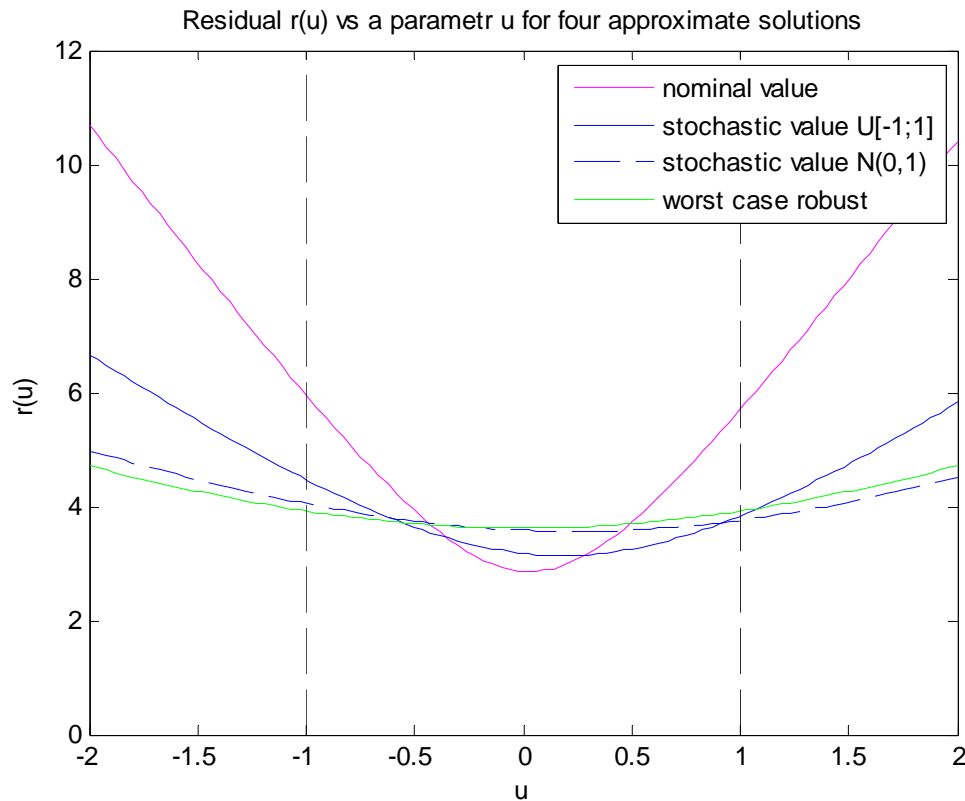


Figure 1. The residual $r(u) = \|A(u)x - b\|_2$ as a function of the uncertain parameter u for four approximate solutions x :

- (1) the nominal least-squares solution x_{nom} ;
- (2) the solution of the stochastic robust approximation problem x_{stoch1} (assuming u is uniformly distributed on $[-1, 1]$);
- (3) the solution of the stochastic robust approximation problem x_{stoch2} (assuming u is normally distributed on $[-1, 1]$);
- (4) the solution of the worst-case robust approximation problem x_{wc} , assuming the parameter u lies in the interval $[-1, 1]$.

2.2 Probabilistic constraints⁵

Chance Constrained Programming belongs to the major approaches for dealing with random parameters in optimization problems. Typical areas of application are engineering and finance, where uncertainties like product demand, meteorological or demographic conditions, currency exchange rates etc. enter the inequalities describing the proper working of a system under consideration. The main difficulty of such models is due to (optimal) decisions that have to be taken prior to the observation of random parameters. In this situation, one can hardly find any decision which would definitely exclude later constraint violation caused by unexpected random effects. Sometimes, such constraint violation can be balanced afterwards by some compensating decisions taken in a second stage.

In many applications, however, compensations simply do not exist or cannot be modeled by costs in any reasonable way. In such circumstances, one would rather insist on decisions guaranteeing feasibility 'as much as possible'. This loose term refers once more to the fact that constraint violation can almost never be avoided because of unexpected extreme events. On the other hand,

⁵ This section is composed from [5] and [3].

when knowing or approximating the distribution of the random parameter, it makes sense to call decisions feasible (in a stochastic meaning) whenever they are feasible with high probability, i.e., only a low percentage of realizations of the random parameter leads to constraint violation under this fixed decision. A generic way to express such a *probabilistic* or *chance constraint* as an inequality is

$$P(h(x, \xi) \geq 0) \geq p$$

The optimization problem with probabilistic constraints can take the form

$$\min \{f(x) \mid P(h_j(x, \xi) \geq 0, (j = 1 \dots m)) \geq p\}, p \in [0, 1].$$

Here, x and ξ are decision and random vectors, respectively, " $h(x, \xi) \geq 0$ " refers to a finite system of inequalities and P is a probability measure. The value $p \in [0, 1]$ is called the probability level, and it is chosen by the decision maker in order to model the safety requirements. Sometimes, the probability level is strictly fixed from the very beginning (e.g., $p=0.95, 0.99$ etc.). In other situations, the decision maker may only have a vague idea of a properly chosen level.

Probabilistic constraints may be given in different forms. The *individual* form is:

$$P(h_j(x, \xi) \geq 0) \geq p_j, j = 1 \dots m.$$

The term 'individual' relates to the fact that each of the constraints is transformed into a chance constraint individually. A different - and more realistic - *joint* form is

$$P(h_j(x, \xi) \geq 0 (j = 1 \dots m)) \geq p.$$

If the random variables $h_1(x, \xi), h_2(x, \xi), \dots, h_m(x, \xi)$ are independent of each other, then the constraint becomes simpler:

$$P(h_j(x, \xi) \geq 0 (j = 1 \dots m)) = P(h_1(x, \xi) \geq 0)P(h_2(x, \xi) \geq 0) \dots P(h_m(x, \xi) \geq 0) \geq p.$$

In some cases the use of the individual constraint instead of the joint constraint may be legitimate from the point of view of model construction. In general, however, the simplification of this type distorts the problem.

Not surprisingly, there does not exist a general solution method for chance constrained programs. The choice strongly depends on:

- Form of the probabilistic constraints (joint or individual);
- Distribution of the random vector (continuous, discrete, independence);
- Properties of the constraint function h (linear, convex, separable).

Of particular interest is the application of algorithms from convex optimization. Convexity of chance constraints, however, does not only depend on convexity properties of the constraint function h but also of the distribution of the random parameter ξ . The question of whether this distribution is continuous or discrete is another crucial aspect for algorithmic treatment. The biggest challenges from the algorithmic and theoretical points of view arise in chance constraints where random and decision variables cannot be decoupled.

In contrast to conventional optimization problems, those including chance constraints introduce the presence of two kinds of approximations: first, the distribution of the random parameter is almost never known exactly and has to be estimated from historical data. Secondly, even a given multivariate distribution (such as multivariate normal) cannot be calculated exactly in general but has to be approximated by simulations or bounding arguments. Both types of imprecision motivate the discussion of stability in programs with chance constraints.

3 Concluding remarks

The optimization under uncertainty is developing rapidly with contributions from many disciplines such as operations research, economics, mathematics, probability, statistics, etc. This brief introduction could neither provide a survey on the whole subject nor give a representative list of references.

References

- [1] P. Kall and Stein W. Wallace, *Stochastic Programming*, John Wiley & Sons, Chichester, 1994.
- [2] A. Ruszczyński and A. Shapiro, *Stochastic Programming*, Handbooks in Operations Research and Management Science, Vol.10, Elsevier, Amsterdam, 2003.
- [3] A. Prekopa, *Stochastic Programming*, Kluwer, Dordrecht, 1995.
- [4] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.
- [5] R. Henrion, *Introduction to Chance-Constrained Programming*, 2004.