Optimization under Uncertainty

Kartikey Sharma - kartikey.sharma@zib.de

Contents

1 Course Organization

The course is divided into two sections. The first part of the course covers Stochastic Programming i.e., optimization under uncertainty models which leverage distributional knowledge about the uncertainty. The second part of the course focuses on Robust Optimization which looks at situations where we have more limited knowledge about the uncertainty primarily about upper and lower bounds.

There will be biweekly problem sets provided. While the problem sets will not be graded, I am happy to discuss the problem with you before or after the class and through online appointments.

1.1 Book

I am primarily using my notes but the following two books can be used as supplementary material.

1. Introduction to Stochastic Programming by John Birge
2. Robust Optimization by Aharon Ben-Tal, Laurent El Ghaoui and Arkadi Nemirovski

1.2 Goals

At the end of the course, you should be able to

1. Given an application, develop an optimization model that takes into account the uncertainty in the problem.
2. Identify suitable models for the uncertainty.
3. Choose appropriate methods to solve the optimization problem.

2 Motivation

Why do we need to solve optimization problems while accounting for uncertainty? Let us start with some examples. The first example we are going to consider is the problem of a Döner seller. Every day they buy the ingredients for Döners in the morning to sell for the rest of the day. For the sake of simplicity, we assume that 1 unit of ingredients is sufficient to make 1 Döner which can then be sold. We can then express the relevant information as follows

1. Units of Döner ingredients purchased: x
2. Purchase cost per unit of ingredients: c
3. Selling price per Döner: p
4. Demand for Döners: d

The decision we need to make is how many ingredient units to purchase to purchase to maximize the profit.

We can write a simplified model for this problem as follows

$$\begin{aligned}
&\max_x f(x,d) := -cx + p\min(x,d)\\
&\text{s.t. } x \geq 0
\end{aligned}$$

In the above optimization problem, objective $f(x,d)$ consists of the cost of the Döner ingredients $-cx$ and the revenue made from selling the Döners $p\min(x,d)$. Note that since we have x units of ingredients we can make x Döners to sell. However, we can only sell as much as the demand. Hence the total Döners sold is $\min(x,d)$. For the above problem, if we know the demand d then the optimal number of Döners to purchase is d leading to a profit of $(p-c)d$.

However, the demand may not be known since when purchasing the ingredients in the morning the demand for the day is in the unknown future. The question then becomes how to solve the problem for uncertain d

The first strategy we attempt is to use a point estimate of d from historical data. We denote such a point estimate by $\bar{d}$. In order to evaluate the performance of the optimal solution obtained using such a point estimate we consider the following example.

Example

Let $c = 10$ and $p = 15$. We assume that d is a random variable as follows

$$\tilde{d} = \begin{cases} 10 \text{ w.p. } 0.3 \\ 20 \text{ w.p. } 0.7 \end{cases}$$

Then point estimate of d or equivalently the expected value of d is

$$\bar{d} = \mathbb{E}[\tilde{d}] = 0.3 \times 10 + 0.7 \times 20 = 17$$

Considering the earlier discussion, the optimal solution is then given as $x^* = \bar{d} = 17$. Since $\tilde{d}$ is a random variable then the objective function $f(x, \tilde{d})$ is also a random variable. We then evaluate the performance of the solution x^* by evaluating the objective function at this point. This is then given by $\mathbb{E}[f(x^*, \tilde{d})]$. We can do so by evaluating the expected profit which is given by

1. w.p. 0.3 it is $-10 * 17 + 15 * 10 = -20$
2. w.p. 0.7 it is $-10 * 17 + 15 * 17 = 85$

The expected profit for $x^* = 17$ is $0.3 \times -20 + 0.7 \times 85 = 53.5$

Is this the best we can do?

If you set $x = 20$ you get an expected profit of 55. Hence, the solution obtained using the point estimate is not necessarily optimal. As such we need to take uncertainty into account during the optimization model itself. Thus a better model for the optimization problem would be

$$\begin{aligned} &\max_{x} \mathbb{E}[f(x, d)] \\ &\text{s.t. } x \geq 0. \end{aligned}$$

Here, we try to maximize the expected value of the objective function.

2.1 Two stage example

The example considered in the previous section involved making a decision before observing the uncertainty realization. However, in many problems, the decision-making is divided into stages with some decisions being made without knowledge of the uncertainty and others being made after observing the uncertainty. The following example illustrates such a problem.

Consider the problem of the Döner seller. Previously they were only purchasing the ingredients in the morning to sell for the rest of the day. However, in this new situation, we assume that the vendor can make purchases in the morning as well as in the middle of the day. The purchases made in the middle of the day are made with knowledge of the demand during the morning. This can provide some indication of the demand for the rest of the day. For the sake of simplicity, we assume that the demand after midday is equal to the demand in the morning. We can then write the resulting problem as

$$\begin{aligned} &\max_{x_1} \; -cx_1 + \mathbb{E}[p\min(x_1, d_1) + Q(x_1, d_1)] \\ &\text{s.t. } x_1 \geq 0 \end{aligned}$$

where $Q(x_1, d_1)$ is a function that reflects the objective function value of the vendor problem for the 2nd half of the day as a function of the decision made in the morning x_1 and the morning demand d_1. We can write the 2nd stage problem as

$$\begin{aligned} Q(x_1, d_1) = &\max_{x_2, s_1} \; -cx_2 + p\min(x_2 + s_1, d_2) \\ &\text{s.t. } s_1 = \max(x_1 - d_1, 0) \\ &\qquad x_2 \geq 0 \end{aligned}$$

The second stage problem is the same as the first stage problem but we also take into account an inventory left over from the first part of the day. This is captured by the variable s_1. Note that there is no uncertainty in the demand for the second half of the day because we assume that the demand is the same as the morning i.e. $d_2 = d_1$.

The above problem is a two-stage stochastic optimization problem.

3 Stochastic Programs

We start with a brief discussion of the notation of random variables.

$\tilde{\xi}$: random vector

ω: sample point

Ω: sample space

ξ^ω: specific realization of the random vector

ξ: generic realization of the random vector

Ξ: support of $\tilde{\xi}$

Given this above notation, we can write the general formulation of a Stochastic Program is given by

$$\begin{aligned} \min_x \ & \mathbb{E}[f_0(x, \tilde{\xi})] \\ \text{s.t. } & \mathbb{E}[f_i(x, \tilde{\xi})] \leq 0, \quad i = 1, \ldots, m \\ & x \in \mathcal{X} \end{aligned}$$

The above optimization problem has the random variable $\tilde{\xi}$ in the objective and in the m constraints. However, the constraint set $x \in \mathcal{X}$ is deterministic and involves no random variables.

The above formulation is a bit too general to be useful. As such we focus on problems with problems with a linear objective and linear constraints.

As an illustration, consider the following mathematical Program

$$\begin{aligned} \min_x \ & \tilde{c}^\top x \\ \text{s.t. } & \tilde{A}x \geq \tilde{b} \\ & x \geq 0 \end{aligned}$$

The above optimization problem is ill-defined as the terms $\tilde{c}, \tilde{A}$ and $\tilde{b}$ are random variables and are not known. They can take multiple possible values so it is not clear what specific value of the random variable should be used to write the problem. As such we can directly solve it as a mathematical program on a computer

In order to solve the above problem in a well-defined fashion we have two options.

1. One option is to observe a realization of the random variables and then solve the problem. The resulting problem can be expressed as

$$\begin{aligned} \min_x \ & (c^\omega)^\top x \\ \text{s.t. } & A^\omega x \geq b^\omega \\ & x \geq 0 \end{aligned}$$

   This is a valid formulation is c^ω, A^ω and b^ω have known deterministic values. However, it comes with the limitation that the decision is being made after knowing c, A and b which may not be possible in many situations.

2. Another option we replace the random variables with their expectation or some operation.

$$\begin{aligned}
&\min_x \mathbb{E}[\tilde{c}^\top x] \\
&\text{s.t. } \mathbb{E}[\tilde{A}x - \tilde{b}] \geq 0 \\
&\quad\quad x \geq 0
\end{aligned}$$

The above is a well-defined formulation. We discuss it and its variants in the upcoming section.

3.1 Approaches to Feasibility

There are several approaches to express the feasible region for constraints with random variables.

3.1.1 Guaranteed Feasibility

In this model, we ensure the solutions satisfy the constraints for every possible realization of the random variable. We can express this feasible region as

$$\mathcal{F} = \bigcap_{\omega \in \Omega} \{x \mid A^\omega x \geq b^\omega, \;\; x \geq 9\}$$

If $x \in \mathcal{F}$ then this ensures that x satisfies the constraint for every realization ξ^ω of $\tilde{\xi}$.

Advantages Any solution which lies in the above feasible region is feasible for any scenario that is realized.

Disadvantages There are several limitations to the above model

1. Due to the need to satisfy the constraints for every scenario, the solution tends to be pessimistic and highly conservative.
2. It is also possible that the feasible region is empty.
3. The solutions are dominated by extreme realization even if they have low probability.

3.1.2 Chance Constrained Models

A key limitation of the previous approach was that the solution needed to be feasible for every scenario. An alternative approach is to ensure that the solution is feasible with high probability. This means that the solution can avoid the need to be feasible in the case of low-probability extreme scenarios which leads to highly conservative solutions. We can write a stochastic program with chance-constrained feasible regions as follows

$$\begin{aligned}
&\min_x c^\top x \\
&\text{s.t. } P(\tilde{A}x \geq \tilde{b}) \geq \alpha \\
&\quad\quad x \geq 0
\end{aligned}$$

where $\alpha \in [0, 1]$. The above is a stochastic program with a deterministic objective but stochastic constraints. The constraints need to be satisfied with probability α which is prespecified. Since all the constraints need to be *jointly* satisfied with probability α the above problem is also referred to as a joint chance-constrained stochastic program.

An alternative to the joint chance-constrained program is an individual chance-constrained model which can be expressed as

$$\begin{aligned}
&\min_x c^\top x \\
&\text{s.t. } P(\tilde{A}_{i,:}x \geq \tilde{b}_i) \geq \alpha_i, \;\; i = 1, \ldots, m \\
&\quad\quad x \geq 0
\end{aligned}$$

Here instead of all constraints being jointly satisfied with probability α each constraint only needs to be satisfied with probability α_i.

Advantages

1. The model is not primarily driven by low-probability worst-case scenarios.
2. Can naturally model reliability constraints which are common in practical applications.

Disadvantages

1. Does not capture the magnitude of the violation. To a chance-constrained program, it does not matter if the constraints are violated by a small amount or a large amount.
2. Feasible region is not necessarily convex which makes it difficult to solve the problem.

3.1.3 Penalty Based Method

The chance-constrained models are limited in the sense that they do not take into account the extent of the violation of a constraint. However, for most problems, constraints can be divided into *soft* constraints and *hard* constraints. While violating the latter may lead to an invalid solution, the violation of the former (especially if by small amounts) only leads to minor penalties.

Suppose we have a vector $q = (q_1, \ldots, q_m) \geq 0$ which denotes the per unit penalty for the violation of each constraint. Let $s^+ = \max(s, 0)$ applied to a vector component-wise. We can write down a stochastic model which takes into account the violation of the constraints as follows

$$\begin{aligned} &\min_x c^\top x + \mathbb{E}[q(\tilde{b} - \tilde{A}x)^+] \\ &\text{s.t. } x \geq 0 \end{aligned}$$

In the above problem, we have moved the constraint to the objective and are penalizing its violation.

Advantages The penalty for constraint violation grows with the magnitude of the violation. This also limits the impact of extreme scenarios.

Disadvantages

1. It does not allow the fine-tuned control of violation probabilities like the chance-constrained models.
2. While the penalty value may arise naturally for some constraints from real-world problems that does not hold always. This leads to challenges when the penalty does not have a good real-world equivalent.

3.2 Approaches to Objective

So far we have focused on stochastic constraints. In this section, we discuss different ways to model objectives with random variables.

3.2.1 Expectation Cost

The most popular methodology is to minimize the expected cost of the objective function. This is illustrated in the following problem

$$\begin{aligned} \min_x \ & \mathbb{E}[\tilde{c}^\top x] \\ \text{s.t. } & Ax = b \\ & x \geq 0 \end{aligned}$$

This methodology is popular but it does not take into account the variability of the random variable. This can lead to solutions which have high expected value but which are highly variable.

3.2.2 Markowitz Mean-Variance Model

In order to take into account the variability in the optimal objective we can simultaneously minimize the expected value and the variance. We use a parameter $\lambda \geq 0$ to balance between the expectation and the variance.

$$\begin{aligned} \min_x \; & \mathbb{E}[\tilde{c}^\top x] + \lambda(Var[\tilde{c}^\top x])^{\frac{1}{2}} \\ \text{s.t. } & Ax = b \\ & x \geq 0 \end{aligned}$$

The above model is popular in portfolio optimization applications. However, it has a limitation in that the variable is symmetric. So it penalizes both upsides and downsides. However, a decision-maker might only care about minimizing the downsides.

3.2.3 Minimize Probability of Expeceeding a Threshold

In order to obtain an asymmetric model we can minimize only the probability of exceeding a cost threshold c_0.

$$\begin{aligned} \min_x \; & \mathbb{P}[\tilde{c}^\top x > c_0] \\ \text{s.t. } & Ax = b \\ & x \geq 0 \end{aligned}$$

This risk measure deals with the issue of asymmetric risks but it does not take into account the magnitude of the cost.

3.2.4 Utility Functions

The most general approach is to model the objective using an expected utility function which can balance both the magnitude and variability of the objective.

$$\begin{aligned} \min_x \; & \mathbb{E}[u(\tilde{c}^\top x)] \\ \text{s.t. } & Ax = b \\ & x \geq 0 \end{aligned}$$

Utility functions are quite general and capture the previously presented models. For example, using an indicator function as the utility function allows us to capture the probabilistic objective. Usually, a good utility function for the objective is *increasing* and *convex*. The former ensures that the objective increases with the cost $\tilde{c}^\top x$ and the latter makes the problem more tractable.

4 Two-stage Stochastic Optimization Formulation

So far all the problems that we have considered have involved static decisions i.e., decisions which do not depend on the uncertainty. However, in many problems, we have to take multiple decisions. Some of these decisions are taken before the uncertainty is realized (and hence do not depend on the uncertainty). However, there are other decisions which are taken after the uncertainty is realized and hence can depend on the uncertainty.

Consider an optimization problem where we have to take 2 decisions x and y. The decision x is taken before the uncertainty $\tilde{\xi}$ is realized and the decision y is taken after the uncertainty is realized. We can write such a problem as

$$\begin{aligned} \min_x \; & c^\top x + \mathbb{E}[h(x, \tilde{\xi})] \\ \text{s.t. } & Ax = b \\ & x \geq 0 \end{aligned}$$

where

$$\begin{aligned} h(x, \tilde{\xi}) = \min_y \ & \tilde{f}^\top y \\ \text{s.t. } & \tilde{D}y = \tilde{B}x + \tilde{d} \\ & y \geq 0 \end{aligned}$$

Here the parameters $\tilde{f}, \tilde{D}, \tilde{B}, \tilde{d}$ are random variables. We use $\tilde{\xi}$ to jointly denote all of them together. The decision y is taken with the knowledge of the values taken by these variables and is different for each realization.

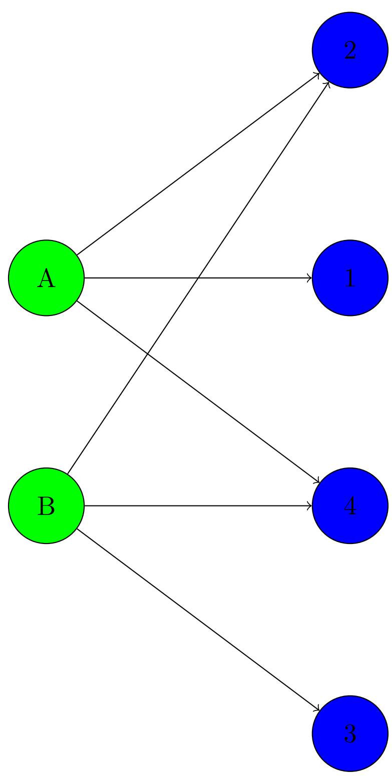
Advantages This model is dynamic and we can model decisions that respond to uncertainty realizations. So we can have a first-stage decision before the uncertainty is realised. This is followed by a second stage decision that responds to the uncertainty.

Disadvantages

1. We are only considering the expectation and not taking into account the variability of the second-stage decisions.
2. We only focus on two stages.

4.1 Illustrative Example

As an example of the two-stage stochastic optimization problem, we consider a transportation problem. Figure 4.1 shows a network with facilities (green) and customers (blue). Each customer has an uncertain demand. We need to make two decisions. The first stage decision is to determine the capacity of A and B. The second stage decision is to transport goods from the facilities to the customers while ensuring that the customer's demand is satisfied and that facility capacity is not exceeded.



We can write such a problem as a two-stage stochastic program. For this setup, we define the following things

Sets

$\mathcal{I} = \{A, B\}$: Set of all facilities

$\mathcal{J} = \{1, \ldots, 4\}$: set of all customers

Ω: Index set for the scenarios

Parameters

$c_i \;\; i \in \mathcal{I}$: per unit cost to expand capacity of facility i

$h_{ij} \;\; i \in \mathcal{I},\ j \in \mathcal{J}$: per unit cost to move good from facility i to customer j

$u_i \;\; i \in \mathcal{I}$: maximum possible capacity of facility i

$p^\omega \;\; \omega \in \Omega$: probability of scenario ω

$d_j^\omega \;\; j \in \mathcal{J},\ \ \omega \in \Omega$: demand of customer j in scenario ω.

Variables

$x_i \;\; i \in \mathcal{I}$: Capacity of facility i

$y_{ij}^\omega \;\; i \in \mathcal{I},\ j \in \mathcal{J},\ \ \omega \in \Omega$: Amount transported from from facility i to customer j in scenario ω.

We can then write the problem as

$$\begin{aligned} \min_x \ & \sum_{i \in \mathcal{I}} c_i x_i + \sum_{\omega \in \Omega} p^\omega h(x_i, d^\omega) \\ \text{s.t. } & 0 \leq x_i \leq u_i \ \ \forall i \in \mathcal{I} \end{aligned}$$

where

$$\begin{aligned} h(x_i, d^\omega) = \min_y \ & \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}} h_{ij} y_{ij}^\omega \\ \text{s.t. } & \sum_{i \in \mathcal{I}} y_{ij}^\omega = d_j^\omega \ \ \forall j \in \mathcal{J} \ \forall \omega \in \Omega \\ & \sum_{j \in \mathcal{J}} y_{ij}^\omega \leq x_i \ \ \forall i \in \mathcal{I} \ \forall \omega \in \Omega \end{aligned}$$