IE 331 OR: Optimization **KAIST, Spring 2023**
Lecture #18: Introduction to optimization under uncertainty May 4, 2023
Lecturer: Dabeen Lee

1 Outline

In this lecture, we cover

- newsvendor problem,
- scenarios,
- optimization under data uncertainty,
- risk measures.

2 Optimization under uncertainty

Consider the optimization model

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \le b_i, \quad i \in [m]. \end{aligned}$$

In this section, we examine the **data** that goes into optimization models.

- For linear and integer programs, these are the objective coefficients c, the constraint matrix A, and the right-hand side vector b.
- We will call the data ξ. We may think of this as a vector of a matrix that stores problem information (costs, resource amounts, prices, demands, etc).

In many situations, some or all of the data are **uncertain**, meaning that we have to **estimate** the values but we do not know them exactly. We will study how to model such problems.

2.1 Example: newsvendor problem

A newsvendor must choose a quantity $x \in \mathbb{Z}_+$ of newspapers to print on a particular day. The cost of printing a copy is $c > 0$. Each copy of newspapers is sold at price $p > c$. Each copy of unsold newspapers must be discarded at cost $h > 0$.

The demand for a newspaper on a particular day is $\xi \in \mathbb{Z}_+$. However, the demand value ξ is unknown to the newsvendor at the time it must decide on x. How do we model this problem?

The objective of the newsvendor is to maximize its profit. The profit is given by

$$f(x, \xi) = p \cdot \min\{x, \xi\} - c \cdot x - h \cdot \max\{0, x - \xi\}.$$

How can we determine the optimal quantity x without known the demand ξ?

2.2 Scenarios

Consider again the generic optimization model with data made explicit.

$$\begin{aligned} \min \quad & f(x, \xi) \\ \text{s.t.} \quad & g_i(x, \xi) \le b_i, \quad i \in [m]. \end{aligned}$$

While we do not know ξ **exactly**, we need to have **some knowledge** of ξ. Assume that we have access to a **finite** set of **scenarios** $\xi_1, \ldots, \xi_N$. These are **possible outcomes** for ξ. Scenario ξ_i may come with probability $p_i \ge 0$ of occurence, i.e., we have

$$\sum_{i \in [N]} p_i = 1, \quad p \ge 0.$$

Where do scenarios come from? Typically, historical data. For the newsvendor problem, it may be past observations of demands.

Why do we only consider finitely many scenarios? Probability distributions such as normal, uniform, exponential distributions have infinitely many outcomes. In constrast, historical data is typically **finite**. Even if the uncertainty ξ is drawn from a continuous distribution, it is hard to process this in an optimization model. Therefore, a classical approach is **simulation**. Basically, we generate samples $\xi_1, \ldots, \xi_N$ from the continuous distribution, and use the **empirical distribution** as an approximation. The empirical distribution is defined by

$$\mathbb{P}[\xi = \xi_i] = \frac{1}{N}, \quad i \in [N].$$

2.3 Optimization models with data uncertainty

Consider again the optimization model

$$\begin{aligned} \min \quad & f(x, \xi) \\ \text{s.t.} \quad & g_i(x, \xi) \le b_i, \quad i \in [m]. \end{aligned}$$

Assume that we have finitely many **scenarios** $\xi_1, \ldots, \xi_N$ for ξ. Think of the decision-making process as occurring in **two steps** as follows.

1. We build a **proxy optimization mode** using the scenarios $\xi_1, \ldots, \xi_N$ to help us make a decision $\hat{x}_N$. For the newsvendor problem, we use historical demand data to choose how many newspapers to print.

2. The **"true" data ξ is revealed**, but we must implement our decision $\hat{x}_N$ before and accept its consequences. For the newsvendor problem, the actual demand ξ for that day is revealed, and we make profit of

$$f(\hat{x}_N, \xi) = p \cdot \min\{\hat{x}_N, \xi\} - c \cdot \hat{x}_N - h \cdot \max\{0, \hat{x}_N - \xi\}.$$

   Note that ξ may or may not be part of the scenario list $\{\xi_1, \ldots, \xi_N\}$.

Another approach is to obtain N copies of the original optimization model with $\xi = \xi_i$ plugged in. Then we obtain $x_1, \ldots, x_N$ different decisions, and we may attempt to aggregate these decisions somehow. However, this is **different** from the above approach.

For simplicity, we consider a model with a **single uncertain constraint**. We may extend our discussion for the single constraint setting to multiple uncertain constraints and uncertain objectives in a straightforward manner.

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g(x,\xi) \le 0, \\ & x \in \mathcal{X} \end{aligned}$$

where $\mathcal{X}$ is a set containing the exact constraints. Given a fixed decision x and scenarios $\xi_1, \ldots, \xi_N$, we may consider the N different values

$$g(x,\xi_1), \ldots, g(x,\xi_N).$$

Here, the constraint may be satisfied for some, but not for others. How can we decide on a single decision x based on the N different values?

3 Risk measures

Higher values of $g(x,\xi)$ are "more risky" than lower values, i.e., they come closer to violating the constraint $g(x,\xi) \le 0$ or may already violate it.

When we are given scenarios, some values of $g(x,\xi_1), \ldots, g(x,\xi_N)$ are more risky than others. Therefore, to meaningfully compare decisions x and x', we must **measure the aggregat risk** across all scenarios $g(x,\xi_1), \ldots, g(x,\xi_N)$ in some way, and then do the same for $g(x',\xi_1), \ldots, g(x',\xi_N)$.

A **risk measure** is simply the name of a formal framework to think about this idea. We can easily think of several ways of aggregating $g(x,\xi_1), \ldots, g(x,\xi_N)$ in order to make our decision, which are all technically risk measures.

Formally, a risk measure $\rho : \mathbb{R}^N \to \mathbb{R}$ maps a vector of different possible outcomes $g(x,\xi_1), \ldots, g(x,\xi_N)$ to a single number

$$\rho\left(g(x,\xi_1), \ldots, g(x,\xi_N)\right).$$

We will use the following shorthand notation

$$\rho_g\left(x; \{\xi_i\}_{i\in[N]}\right) := \rho\left(g(x,\xi_1), \ldots, g(x,\xi_N)\right).$$

We then find $\hat{x}_N$ by solving

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & \rho_g\left(x; \{\xi_i\}_{i\in[N]}\right) \le 0, \\ & x \in \mathcal{X} \end{aligned}$$

Then the next question is, **which risk measures should we use?**

We want risk measures to have a **basic monotonicity property**. That is,

$$g(x,\xi_i) \le g(x',\xi_i) \quad \forall i \in [N] \quad \Rightarrow \quad \rho_g\left(x; \left\{\xi_i\right\}_{i\in[N]}\right) \le \rho_g\left(x'; \left\{\xi_i\right\}_{i\in[N]}\right).$$

In other words, if the risk at **every scenario** is greater, then the **aggregate risk** should be greater. There are other reasonable properties that we can think of, but we will not explore these in detail. Instead, we will focus on specific risk measures that can be modeled using linear constraints.

3.1 Risk measure 1: expectation

Assume that there exists a **distribution over the scenarios**, i.e., weights $p_1, \ldots, p_N \geq 0$ such that $\sum_{i\in[N]} p_i = 1$, capturing the likelihood of each scenario. Define a distribution $\hat{P}_N$ as the one that satisfies

$$\mathbb{P}_{\xi\sim\hat{P}_N}[\xi = \xi_i] = p_i, \quad i \in [N].$$

The **expectation risk measure** is simply the expectation of $g(x, \xi)$ over the distribution $\hat{P}_N$:

$$\rho_g\left(x; \{\xi_i\}_{i\in[N]}\right) = \mathbb{E}_{\xi\sim\hat{P}_N}[g(x,\xi)] = \sum_{i\in[N]} p_i \cdot g(x, \xi_i).$$

When $p_1, \ldots, p_N$ correspond to the empirical distribution, then we have $p_i = 1/N$.

Lemma 18.1. *If $g(x, \xi_i)$ for each $i \in [N]$ is linearly representable, then so is the expectation risk measure.*

3.2 Risk measure 2: worst-case value (robust optimization)

The **worst-case value** risk measure is defined by

$$\rho_g\left(x; \{\xi_i\}_{i\in[N]}\right) = \max_{i\in[N]} g(x, \xi_i).$$

There is **no need** for likelihood weights p_i. If $g(x, \xi_i)$ is linearly representable, then $\rho_g\left(x; \{\xi_i\}_{i\in[N]}\right)$ is also linearly representable. Why? Note that

$$\rho_g\left(x; \{\xi_i\}_{i\in[N]}\right) \leq 0$$

is equivalent to

$$g(x, \xi_i) \leq 0, \quad i \in [N].$$

3.3 Issues with expectation and worst-case value

Expectation and worst-case value as a risk measure are useful, but there are some shortcomings. Consider two scenarios, 1 and 2, and two possible decisions x and x'.

i	1	2
p_i	0.5	0.5
$g(x, \xi_i)$	-1	1
$g(x', \xi_i)$	-100	100

Then it follows that

$$\mathbb{E}_{\xi\sim\hat{P}_N}[g(x,\xi)] = \frac{1}{2}\cdot(-1) + \frac{1}{2}\cdot 1 = 0, \quad \mathbb{E}_{\xi\sim\hat{P}_N}\left[g(x',\xi)\right] = \frac{1}{2}\cdot(-100) + \frac{1}{2}\cdot 100 = 0.$$

However, it is not intuitive at all to perceive x and x' as the same. In the worst case, x' has value 100 which is much worse than 1. On the other hand, focusing on the worst-case is sometimes **too conservative**. If we only have 0.001% chance of a bad value, then maybe it is not too dangerous to ignore the worst case.

Is there a systematic way to better account for risk values while being less conservative than the worst-case value?