STOCHASTIC PROGRAMMING

November 24, 2023

Today:
- Introduction
- Two-stage stochastic programs

Logistics

Course webpage:

https://www.lamsade.dauphine.fr/~croyer/teachSP.html

My email:

clement.royer@lamsade.dauphine.fr

Five lectures

1. Nov 24 1.45-5pm
2. Nov 30: 8.30-11.45am
3. Dec 7: 8.30-11.45am
4. Dec 14: 8.30-11.45am
5. Dec 21: 8.30-11.45am

Assessment

100% exam (January 15, 2024)

2 hours, 1 sheet of notes allowed

INTRODUCTION

↳ Stochastic programming is a subfield of (mathematical) optimization

Optimization = Making the best decision out of a set of alternatives

↳ Mathematically, an optimization problem is written as

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \; f(x) \quad \text{s.t.} \quad x \in X$$

- minimize: Find the lowest value of the objective
- f: $f: \mathbb{R}^n \to \mathbb{R}$ objective function — quantifies how good a decision is
- s.t.: "subject to"
- $x \in X$: Constraints of the problem (conditions on the decision variables) $X \subseteq \mathbb{R}^n$
- $x \in \mathbb{R}^n$: Decision variables (typically concatened into a vector) — Parameters that characterize a decision

↳ We will also consider maximization problems

$$\underset{x \in \mathbb{R}^n}{\text{maximize}} \; f(x) \quad \text{s.t.} \quad x \in X$$

which is equivalent

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \; (-f)(x) \quad \text{s.t.} \quad x \in X$$

<u>Terminology and definitions</u> (for problem $\underset{x \in \mathbb{R}^n}{\text{minimize}} \; f(x)$ s.t. $x \in X$)

- $X \subseteq \mathbb{R}^n$ is called the <u>feasible set</u>
- A point $x \in X$ is feasible for the problem.
- A point $x \notin X$ is called <u>infeasible</u>

- A solution of the problem is a vector $x^* \in \mathbb{R}^n$ such that
  - $x^* \in X$ (feasibility)
  - $\forall x \in X,\ f(x^*) \leq f(x)$ (optimality for minimization)

NB: Solutions of minimization problems are also called global minima.

The set of solutions of the problem is denoted by

$$\underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \{ f(x) \text{ s.t. } x \in X \} \subseteq \mathbb{R}^n$$

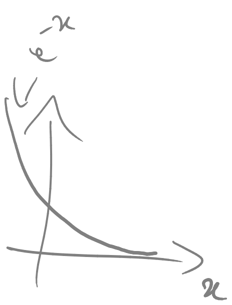
The optimal value of the problem is denoted by

$$\min_{x \in \mathbb{R}^n} \{ f(x) \text{ s.t. } x \in X \} \in \overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$$

- A problem is called infeasible if $X = \emptyset$

  In that case, $\underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \{ f(x) \text{ s.t. } x \in X \} = \emptyset$

  and $\min_{x \in \mathbb{R}^n} \{ f(x) \text{ s.t. } x \in X \} = +\infty$



- A problem is called unbounded if it does not have a solution but the feasible set is not empty. In that case,

$$\underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \{ f(x) \text{ s.t. } x \in X \} = \emptyset$$

$$\min_{x \in \mathbb{R}^n} \{ f(x) \text{ s.t. } x \in X \} = -\infty$$

↳ We define similarly the notions of argmax( ) and max( )

$$\underset{x \in \mathbb{R}^n}{\operatorname{argmax}} \{ f(x) \text{ s.t. } x \in X \} = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \{ (-f)(x) \text{ s.t. } x \in X \}$$

$$\max_{x \in \mathbb{R}^n} \{ f(x) \text{ s.t. } x \in X \} = - \min_{x \in \mathbb{R}^n} \{ (-f)(x) \text{ s.t. } x \in X \}$$

Stochastic programming problem

$$(P) \quad \underset{x \in \mathbb{R}^n}{\text{minimize}} \; F(x, \xi) \quad \text{s.t.} \quad x \in X(\xi)$$

ξ : vector of uncertainties (unknown)

↳ Stochastic programming = when the deterministic equivalent of (P) can be solved using mathematical optimization techniques

↳ As stated, (P) is not really solvable, because the value of ξ is unknown.

- One approach consists in fixing ξ to $\bar{\xi}$ and solve the deterministic problem
$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \; F(x, \bar{\xi}) \quad \text{s.t.} \quad x \in X(\bar{\xi})$$
$\Rightarrow$ Can be very far from other possible values of ξ

- Another (robust) approach consists in assuming that $\xi \in \Xi$ and solving

Ξ: all possible values for ξ

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \; \max_{\xi \in \Xi} F(x, \xi)$$
$$\text{s.t.} \quad x \in \bigcap_{\xi \in \Xi} X(\xi)$$

↳ The most classical formulation in stochastic programming is the static stochastic programming formulation

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \ \mathbb{E}_\xi\left[F(x,\xi)\right] \quad \text{s.t.} \quad x \in X \subseteq \mathbb{R}^n$$

↑ Expectation with respect to ξ

$$\mathbb{E}_\xi\left[F(x,\xi)\right] = \int_{\Xi} F(x,\xi)\, p(\xi)\, d\xi$$

where $\xi \in \Xi$ and $p : \Xi \to [0,1]$ is the probability distribution of ξ

⚠ $\mathbb{E}_\xi\left[F(x,\xi)\right] \neq F(x, \mathbb{E}_\xi[\xi])$ in general

→ Uncertainty only affects the objective function and it is handled using the expected value operator

→ Here-and-now decision : To evaluate F, one selects x first then one observes $F(x, \xi_1)$ for a realization of ξ

$\Rightarrow$ $\mathbb{E}_\xi\left[F(x,\xi)\right]$ cannot be evaluated directly in general

Approximations of the static formulation

↳ Since $\mathbb{E}_\xi\left[F(x,\xi)\right]$ is expensive to compute as a function of x, we approximate it using samples of the variable ξ

↳ Two main techniques

- stochastic approximation
  - → Approach used in AI/ML these days
  - → Algorithm uses one sample at every iteration

Ex for F differentiable

$$\forall k \geq 0 \qquad x^{k+1} = x^k - P_X\left[x^k - \alpha_k \nabla F(x^k, \xi^k)\right]$$

$\nabla F(x^k, \xi^k)$: gradient of $F(\cdot, \xi^k)$ ← fixed realization

$P_X[\cdot]$: projection onto X

- Sample Average Approximation

$\xi_1, \dots, \xi_N$ samples of ξ (typically independent)

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad \frac{1}{N} \sum_{i=1}^{N} F(x, \xi_i) \quad \text{s.t. } x \in X$$

$\approx \mathbb{E}_\xi[F(x,\xi)]$ (approximation via a discrete average)

↳ This course is about the main principles behind solving static stochastic programming problems, and its main variations

Static ⟹ Dynamic (two-stage, multi-stage)

$\mathbb{E}_\xi[\cdot] \Rightarrow$ Risk measures

$X \Rightarrow X(\xi)$ (chance constraints)

ξ with a distribution p ⟹ Family of distributions

Most classical SP example: Newsvendor problem

$n = 1$

A vendor of newspapers purchases a quantity $x \in \mathbb{R}$ of newspapers ahead of knowing the demand D.

- The per-unit cost of purchase is $c_p > 0$
  (Purchase x ⟹ cost $c_p x$)
- The back-ordering cost is $c_b > 0$ ($c_b > c_p$ typically)
  (Demand $D > x$ ⟹ cost $c_b(D - x)$)

- The holding cost is $c_h > 0$

  (Demand $D < x \Rightarrow$ cost $c_h(x-D)$)

Total cost for the newsvendor when purchasing x

$$F(x, D) = c_p x + c_b \max(D-x, 0) + c_h \max(x-D, 0)$$

Newsvendor problem

$$\underset{x \in \mathbb{R}}{\text{minimize}} \quad \mathbb{E}_D[F(x, D)]$$

$$\text{s.t.} \quad x \geq 0$$

$$X = \{x \mid x \geq 0\}$$

TWO-STAGE STOCHASTIC PROGRAMMING

Motivation (through the newsvendor problem)

- In the newsvendor problem's objective, the term $c_p x$ does not depend on the uncertain demand D, and can be computed exactly and deterministically

$$\mathbb{E}_D[F(x, D)] = c_p x + \mathbb{E}_D[c_b \max(D-x, 0) + c_h \max(x-D, 0)]$$

$\Rightarrow$ this structure is called a two-stage structure

Def: A two-stage stochastic program (aka a stochastic program with recourse) is a stochastic program of the form

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x) + \mathbb{E}_\xi[Q(x, \xi)] \quad \text{s.t. } x \in X$$

$f: \mathbb{R}^n \to \mathbb{R}$ that does not depend on ξ

and $Q(x, \xi)$ is given as the solution to another optimization problem:

$$Q(x, \xi) = \min_{y \in \mathbb{R}^{m_1}} g(y, \xi) \quad \text{s.t.} \quad y \in \mathcal{Y}(x, \xi)$$

x: Fixed when evaluating Q

ξ: Fixed when evaluating Q

$g(\cdot, \xi) : \mathbb{R}^{m_1} \to \mathbb{R} \quad \forall \xi$

$\forall (x, \xi), \; \mathcal{Y}(x, \xi) \subseteq \mathbb{R}^{m_1}$

- x is the here-and-now decision
- y is the wait-and-see decision, that can only be computed for known x and ξ

Two-stage process: x (decide x) $\longrightarrow$ ξ (is realized) $\longrightarrow$ y (decided)

Newsvendor problem

↳ Could model the problem as

$$\underset{x \in \mathbb{R}}{\text{minimize}} \quad c_p x + \mathbb{E}_D\left[Q(x, D) \right]$$

$$\text{s.t.} \quad x \geq 0$$

where

$$Q(x, D) = \begin{cases} \min_{y \in \mathbb{R}^2} & c_b y_1 + c_h y_2 \\ \text{s.t.} & y_1 \geq D - x \\ & y_1 \geq 0 \\ & y_2 \geq x - D \\ & y_2 \geq 0 \end{cases}$$

we can solve this problem for y: ←

$y_1^* = \max(D - x, 0)$

$y_2^* = \max(x - D, 0)$

The problem is equivalent to

$$\text{minimize}_{\substack{x \in \mathbb{R} \\ y \in \mathbb{R}^2}} \quad c_p x + \mathbb{E}_D [c_b y_1 + c_h y_2]$$

$$\text{s.t.} \quad x \geq 0, \quad y_1 \geq D - x, \quad y_1 \geq 0, \quad y_2 \geq x - D, \quad y_2 \geq 0$$

All decision variables

→ We can isolate x in the first stage so as to optimize over x

→ The value of y will depend on x and D

uncertainty in the constraints

↳ The same reformulation would apply with costs c_b and c_h that are dependent on D

The main class of two-stage stochastic programs is the class of linear two-stage stochastic programs

$\forall (u,v) \in (\mathbb{R}^n)^2$

$u^T v = \sum_{i=1}^{n} u_i v_i$

Linear program in x →

$$\text{minimize}_{x \in \mathbb{R}^n} \quad c^T x + \mathbb{E}_\xi [Q(x, \xi)]$$

$$\text{s.t.} \quad Ax = b, \quad x \geq 0$$

$$\text{where} \quad Q(x, \xi) = \min_{y \in \mathbb{R}^{m_1}} q^T y \quad \text{s.t.} \quad Tx + Wy = h, \quad y \geq 0$$

$\xi = (q, T, W, h)$

↓ For fixed (q, T, W, h, x), linear program in y

Key property: This problem can be reformulated as a single linear program jointly on x and y, but with stochastic coefficients in the objective and the constraints

$$\underset{\substack{x \in \mathbb{R}^n \\ y \in \mathbb{R}^{n_1}}}{\text{minimize}} \quad c^T x + q^T y \quad \text{s.t.} \quad \begin{aligned} Ax &= b \\ Tx + Wy &= h \\ x &\geq 0 \\ y &\geq 0 \end{aligned}$$

⚠ This notational convenience can hide the stochastic nature of certain quantities

↳ Deterministic linear programs can be solved numerically to high accuracy in very large dimensions ($n \simeq 10^9$, even more for certain problems) using state-of-the-art solvers

→ For an LP $\underset{x \in \mathbb{R}^n}{\text{minimize}} \; c^T x$ s.t. $Ax = b$, $x \geq 0$, can find $\bar{x}$ such that $\|A\bar{x} - b\| < 10^{-12}$, $\bar{x} \geq 0$

and $c^T \bar{x} - \min_{x \in \mathbb{R}^n} \{ c^T x \text{ s.t. } Ax = b \} < 10^{-12}$

↳ The main algorithms for two-stage LPs are based on

- a good linear programming solver (or algorithm)
- scenarios for the stochastic quantities

Scenario approach

Suppose that $\xi = (q, T, W, h)$ has a finite number of possible values $\{ \xi_k = (q_k, T_k, W_k, h_k) \}_{k=1..K}$ and that

probability $p(\xi_k)$ → $\mathbb{P}(\xi = \xi_k) = p_k \geq 0$ with $\sum_{k=1}^{K} p_k = 1$

Then, the two-stage LP can be rewritten as a deterministic LP by using one variable for every scenario

$$\underset{\substack{x \in \mathbb{R}^n \\ y_1, \dots, y_K \in \mathbb{R}^{m_1}}}{\text{minimize}} \quad c^T x + \sum_{k=1}^{K} p_k q_k^T y_k$$

↳ Average of $q_k^T y_k$

$\forall y, \sum_{k=1}^{K} p_k q_k^T y = \mathbb{E}_\xi [q^T y]$

$$\text{s.t.} \quad Ax = b$$

$$T_k x + W_k y_k = h_k \quad k = 1..K \quad \Big) \text{ K constraints}$$

$$x \geq 0 \quad \longrightarrow \quad x \geq 0 \iff [x]_i \geq 0 \ \forall i = 1..n$$

$$y_k \geq 0 \quad k = 1..K$$

(+) The K terms in the sum are independent (they depend on different y_ks)

(+) Also true for the constraints