CS 524: Introduction to Optimization
Lecture 34 : Stochastic Programming

Michael Ferris

Computer Sciences Department
University of Wisconsin-Madison

November 22, 2023

Stochastic Programming

$64 Question

- What does "Programming" mean in "Mathematical Programming", "Linear Programming", etc...?

- Answer: Planning.
- Mathematical Programming (Optimization) is about decision making, or planning.
- Stochastic Programming is about decision making *under uncertainty*.
- View it as "Mathematical Programming with random parameters"

Dealing With Randomness

- Typically, randomness is ignored, or it is dealt with by
  - Sensitivity analysis
    - Look at effects of certain perturbations of the data
    - For large-scale problems, sensitivity analysis is useless
  - "Careful" determination of instance parameters
    - No matter how careful you are, you can't get rid of inherent randomness.
- Stochastic Programming is the way![1]

[1] This is not necessarily true, but humor me for the time being

Applications of Optimization under Uncertainty

- Generation of electrical power
  - Utilities must decide how much power to produce before knowing demand
  - Wind supply is also highly stochastic
- Reservoir operation
  - Flood control or recreation or irrigation or power supply
  - How much water can/should be released?
- Portfolio Selection: Returns are uncertain
- Inventory management: our newsvendor problem today

Three Ways for Modeling Uncertainty

1. Deterministic modeling: The uncertainty plays no role.
   - Often, modelers know there are obviously important uncertainties, but can use only a "point estimate" (e.g. mean values) of the uncertainty.
   - Not enough is known about the uncertainty
   - They don't have software available for getting numerical solutions
2. Stochastic Modeling
   - Uncertain elements are treated as random variables, so probability theory may be applied
   - The elements must have a "known" probability distribution
   - Sometimes the distribution comes from a statistical model, or it is an educated guess (subjective probabilities): mathematical treatment is the same
3. Range Modeling (robust optimization)
   - Uncertain probabilities: The actual distribution will be chosen by "an adversary" (Mother Nature) from a limited range of distributions.
   - Range modeling can be combined with stochastic modeling, by assuming that a probability distribution is present, but not completely known.

Hot Off the Presses

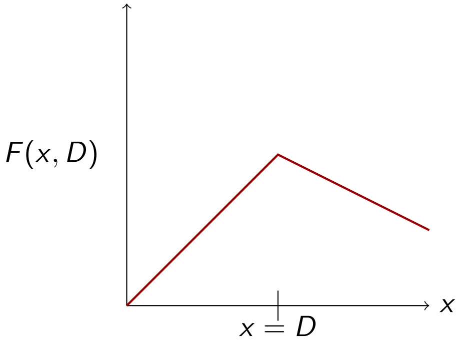
- A paperboy (newsvendor) needs to decide how many papers x to buy in order to maximize his profit.
- He doesn't know at the beginning of the day how many papers he can sell (his demand).
  - Each newspaper costs c.
  - He can sell each newspaper for a price of q.
  - He can return each unsold newspaper at the end of the day for r.
  - The demand is D.

Newsvendor Profit

$$F(x, D) = \begin{cases} (q-c)x & \text{if } x \leq D \\ qD + r(x-D) - cx & \text{if } x > D \end{cases}$$

Profit Function

- Marginal profit: $(q - c)$ if can sell all: $x \leq D$
- Marginal loss: $(c - r)$ if have to salvage



$$F(x, D) = \min\{(q - c)x, (q - r)D + (r - c)x\}$$

What Should We Do?

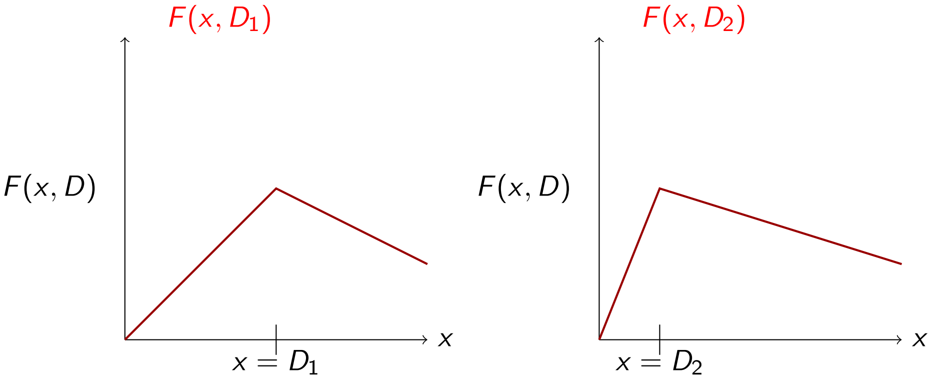
- Optimize, silly:

$$\max_{x \geq 0} F(x, D).$$

http://en.wikipedia.org/wiki/Chewbacca_defense

- This problem does not make sense!
- You can't optimize something random!

The Function is Random



- One x can't simultaneously optimize both functions

(Silly) Idea

- Let's assume that D is a random variable with cumulative distribution function (cdf) $H(t) \stackrel{\text{def}}{=} \text{Prob}(D \leq t)$

Plan for Average Case

- Let $\mu \stackrel{\text{def}}{=} \mathbb{E}[D]$ be the mean value of demand
- In this case: (proof by picture)

$$\max_{x \geq 0} F(x, \mu) \Rightarrow x^* = \mu.$$

- In this case, the optimal policy is to purchase μ

The Flaw of Averages

- We will see that this can be *far* from optimal when your problem takes more uncertainty into account

Scenario Modeling

- The most common representation of uncertainty (in stochastic programming) is via a list of scenarios, which are specific representations of how the future will unfold.
- Think of these as $\xi^1, \xi^2, \ldots \xi^S$, with $\xi^j \in \Xi$

What we CAN'T do

- Planners often generate a solution for each scenario generated—"What-if" analysis.
- Each solution yields a prescription of what should be done if the scenario occurs, but there is no theoretical guidance about the compromise between those prescriptions
- Is there a systematic way to generate one solution or policy to follow that will achieve a "good" objective with respect to all scenarios?

Stages and Decisions

- The newsvendor problem is a classical "recourse problem":

1. We make a decision now (first-stage decision)
2. Nature makes a random decision ("stuff" happens)
3. We make a second-stage decision that attempts to repair the havoc wrought by nature in (2) (recourse)

Key Idea

The evolution of information is of paramount importance

Newsvendor SP

- Suppose that
$$\Omega = \{d_1, d_2, \ldots d_{|S|}\}$$
- So there are a finite set of scenarios S, each with associated probability p_s. ($\sum_{s \in S} p_s = 1$)

Parameters

- d_s: Demand for newspapers in scenario s
- p_s: Probability of scenario s

Variables

- x: Number to purchase
- y_s: Number to sell in scenario s
- z_s: Number to salvage in scenario s

Newsvendor Stochastic LP (`newsvendor.gms`)

Note: in our example first stage decision is x, random decision is D and second-stage decisions are amount to sell y and amount to return z

$$\max -cx + \sum_{s \in S} p_s(qy_s + rz_s)$$

s.t.

$$\begin{aligned} d_s &\geq y_s && \forall s \in S \\ x - y_s &= z_s && \forall s \in S \\ x &\geq 0 && \\ y_s, z_s &\geq 0 && \forall s \in S \end{aligned}$$

Newsvendor Again: Profit

$$F(x, D) = \min\{(q - c)x, (q - r)D + (r - c)x\}$$

- $q = 2, c = 0.3, r = 0.05$
- D a random variable with cdf $H_D(t)$
- Demand: Normally distributed. $\mu = 100, \sigma = 20$

- Mean Value Solution
  - Buy 100. (Duh!)
  - TRUE long run profit ≈ 155
- Stochastic Solution
  - Buy $123 = H^{-1}(2 - 0.3/(2 - 0.05))$ (e.g. `stolib.icdfnormal`)
  - TRUE long run profit ≈ 163
  - Note: we can show that

$$x^* = H^{-1}\left(\frac{q - c}{q - r}\right).$$

- The difference between the two solutions $(163 - 155)$ is called the *value of the stochastic solution*.

Out of sample testing

- Make sure you kick the model around by testing on similar but not identical data (sensitivity)
- If you solve a stochastic program, ensure that you a-posteori take a completely new sample and evaluate your solution on that sample: this is called out-of-sample testing
- Once we have a solution x^* we can evaluate the true solution or an out of sample estimate of it by:
  1. generate lots of new samples $\hat{d}_1, \hat{d}_2, \ldots, \hat{d}_N$
  2. evaluate
  $$\frac{1}{N}\sum_{i=1}^{N} F(x^*, \hat{d}_i)$$
  (no optimization, just evaluation of F)
- this is an estimate of the true solution (as reported above)

Put Another Way

- We could write the objective for the newsvendor problem in the form:

$$F(x, D) = -cx + \mathbb{E}Q(x, D),$$

where

$$Q(x, D) = \max_{y \geq 0, z \geq 0} \{qy + rz \mid y \leq D, y + z = x\}.$$

- $Q(x, D)$ is the optimal recourse function: Given that we have chosen x and observed demand D, what should I do to maximize profit?

It's Not Always So Easy

- For the newsvendor the recourse function: $Q(x, D)$ has a simple closed form:

$$Q(x, D) = \min\{qx, qD + r(x - D)\}$$

  but it general it will not. For two-stage stochastic linear programming, it will be the optimal value of a linear program (as we demonstrated even in this case above)
- Note that sampling in GAMS is demonstrated in the `newsvendor.gms` file, and more information is available at `https://www.gams.com/latest/docs/UG_ExtrinsicFunctions.html`

A Take Away Message

The "Flaw" of Averages

- The flaw of averages occurs when uncertainties are replaced by "single average numbers" planning.
  - Did you hear the one about the statistician who drowned fording a river with an average depth of three feet.

Point Estimates

- If you are planning a point estimates, then you are planning sub-optimally
- It doesn't matter how carefully you choose the point estimate – it is impossible to hedge against future uncertainty by considering one realization of the uncertainty in your planning process