Introduction to One and Two-Stage Stochastic and Robust Optimization

Michel De Lara
Cermics, École nationale des ponts et chaussées
IP Paris, France

January 14, 2025

Outline of the presentation

In decision-making, risk and time are bedfellows,
but for the fact that an uncertain outcome is revealed after the decision

The talk moves along the number of decision stages: 1,2, more

Working out static examples

Two-stage linear stochastic programs

Two-stage stochastic programs

Two-stage stochastic programs with risk

Outline of the presentation

Working out static examples

Two-stage linear stochastic programs

Two-stage stochastic programs

Two-stage stochastic programs with risk

Working out classical examples

We will work out classical examples in Stochastic Optimization

- the blood-testing problem

  static, only risk

- the newsvendor problem

  static, only risk

Outline of the presentation

The blood-testing problem (R. Dorfman)

- A large number N (say, $N = 1,000$) of possibly diseased individuals are subjected to a blood test
- Blood-testing method:
  the blood samples of k individuals are pooled together
  and analyzed together
  - If the pool test is negative,
    this one test suffices for the k individuals
  - If the pool test is positive,
    each of the $k > 1$ individuals must be tested separately,
    and $k + 1$ tests are required, in all

The blood-testing problem is a static stochastic optimization problem

- Data:
  - A large number N of individuals are subjected to a blood test
  - The probability that the test is positive is $p \in]0,1[$, (small, say $p = 0.01$) the same for all individuals (a positive test means that the target individual has a specific disease; the prevalence of the disease in the population is p)
  - Individuals are stochastically independent
- Blood-testing method:
  the blood samples of k individuals are pooled and analyzed together
  - If the test is negative, this one test suffices
  - If the test is positive, $k + 1$ tests are required, in all
- Optimization problem:
  - Find the value of k which minimizes the expected number of tests
  - Find the minimal expected number of tests

What is a possible stochastic model?

- Sample space Ω (describes all possible outcomes)
- Primitive random variables (a way to describe relevant outcomes)
- Probability $\mathbb{P}$ on Ω (assigns weights to all possible outcomes)

Once equipped with a stochastic model,

- the number of diseased individuals in a group is a random variable, which depends on the number k of individuals
- hence, the total number of tests is a random variable

$$T_k : \Omega \to \mathbb{N}$$

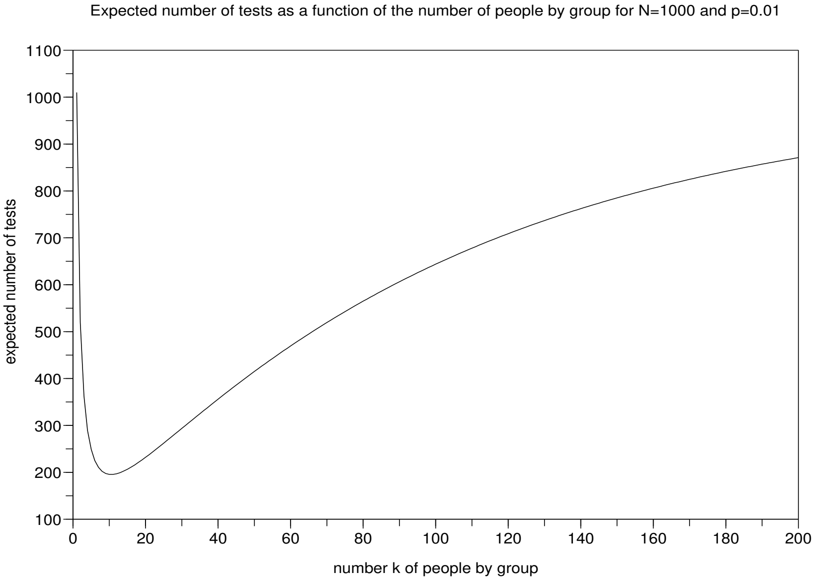
which depends on the number k of individuals,
with probability distribution $\mathbb{P} \circ T_k^{-1}$ on $\mathbb{N}$,
hence mathematical expectation $\mathbb{E}(T_k)$

What is the expected number $\mathbb{E}(T_k)$ of tests?

- For the first pool $\{1, \ldots, k\}$, the test is
  - negative with probability $(1-p)^k$ (by independence) $\to$ 1 test
  - positive with probability $1-(1-p)^k \to k+1$ tests
- When the pool size k is small,
  compared to the number N of individuals,
  the blood samples $\{1, \ldots, N\}$ are split in approximately N/k groups,
  so that the expected number of tests is

$$\mathbb{E}(T_k) = J(k) \approx \frac{N}{k}[1 \times (1-p)^k + (k+1) \times (1-(1-p)^k)]$$

The expected number $\mathbb{E}(T_k)$ of tests displays a marked hollow



In army practice, R. Dorfman achieved savings up to 80%

- The expected number of tests is

$$J(k) \approx \frac{N}{k}[1 \times (1-p)^k + (k+1) \times (1-(1-p)^k)]$$

- For small p,

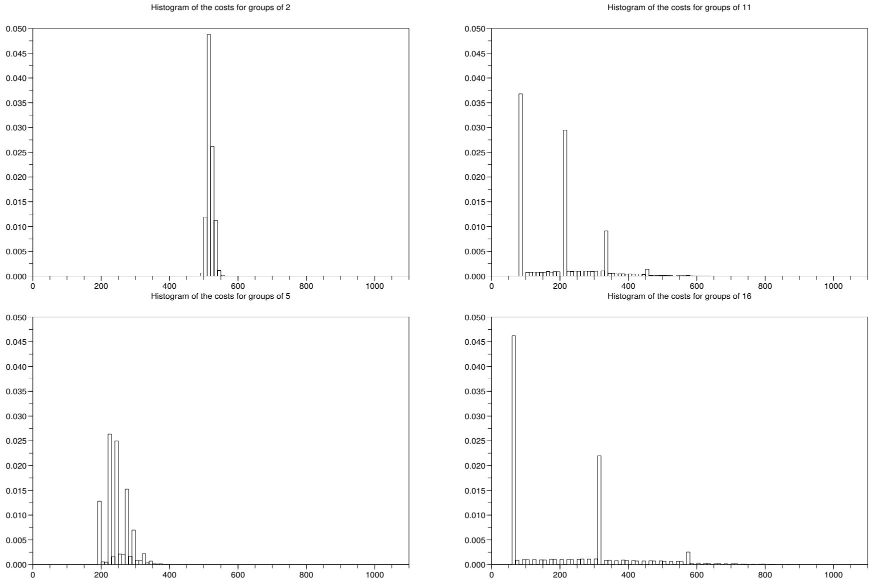
$$J(k)/N \approx 1/k + kp$$

  - so that the optimal number of individuals per group is $k^* \approx 1/\sqrt{p}$
  - and the minimal expected number of tests is about

$$J^* \approx J(k^*) \approx 2\sqrt{p} \times N < N$$

- William Feller reports that, in army practice,
  R. Dorfman achieved savings up to 80%,
  compared to making N tests (the worst case solution)
  (take $p = 1/100$, giving $k^* = 11 \approx 1/\sqrt{1/100} = 10$ and $J^* \approx N/5$)

The optimal number T_{k^*} of tests is a random variable



What about risk?

- The optimal number of individuals per group is 11
  if one minimizes the mathematical expectation $\mathbb{E}$
  of the number of tests
  (see also the top right histogram above)
- But if one minimizes the Tail Value at Risk at level $\lambda = 5\%$
  of the number of tests (more on $TVaR_\lambda$ later),
  numerical calculation show that, in the range from 2 to 33,
  the optimal number of individuals per group is 5
  (see also the bottom left histogram above)
- The bottom left histogram is more tight (less spread)
  than the top right histogram

Outline of the presentation

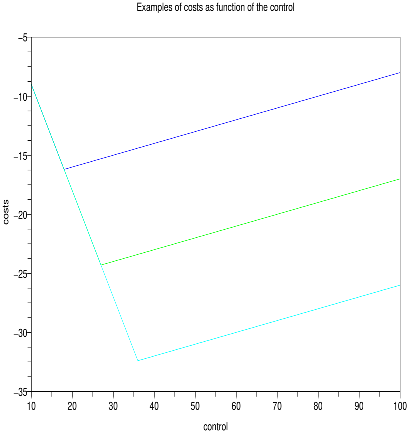
The "news*boy* problem" is now coined the "news*vendor* problem" ;-)

The (single-period) newsvendor problem stands as a classic in stochastic optimization

- Each morning, the newsvendor must decide how many copies $u \in \mathcal{U} = \{0, 1, 2 \ldots\}$ of the day's paper to order: u is the decision variable
- The newsvendor will meet a demand $w \in \mathcal{W} = \{0, 1, 2 \ldots\}$: the variable w is the uncertainty
- The newsvendor faces an economic tradeoff
  - she pays the unitary purchasing cost c per copy
  - she sells a copy at price p
  - if she remains with an unsold copy, it is worthless (perishable good)
- The newsvendor's costs $j(u, w)$ depend both on the decision u and on the uncertainty w:

$$j(u,w) = \underbrace{cu}_{\text{purchasing}} - \underbrace{p\min\{u,w\}}_{\text{selling}} = \max\{cu - pu, cu - pw\}$$

What is an "optimal" solution to the newsvendor problem?



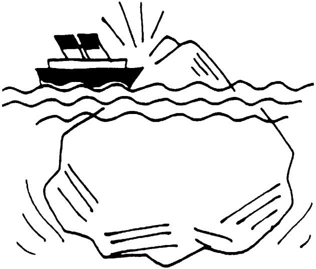
If you solve

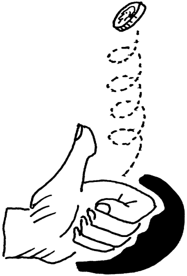
$$\min_{u \in \mathcal{U}} j(u, w)$$

the optimal solution is $u^* = w$,
which depends...
on the unknown quantity w!

So, what would you propose for an "optimal" solution?

For you, Nature is rather random or hostile?





The newsvendor reveals her attitude towards risk in how she aggregates outcomes with respect to uncertainty

- In the robust or pessimistic approach, the (paranoid?) newsvendor minimizes the worst costs

$$\min_{u \in \mathcal{U}} \underbrace{\max_{w \in \overline{\mathcal{W}}} j(u, w)}_{\text{worst costs } J(u)}$$

  as if Nature were malevolent

- In the stochastic or expected approach, the newsvendor solves

$$\min_{u \in \mathcal{U}} \underbrace{\mathbb{E}_{\mathbf{W}}[j(u, \mathbf{W})]}_{\text{expected costs } J(u)}$$

  as if Nature played stochastically (casino)

If the newsvendor minimizes the worst costs

- We suppose that
  - the demand w belongs to a set $\overline{\mathcal{W}} = [\![w^\flat, w^\sharp]\!]$
  - the newsvendor knows the set $[\![w^\flat, w^\sharp]\!]$
- The worst costs are

$$J(u) = \max_{w \in \overline{\mathcal{W}}} j(u,w) = \max_{w \in [\![w^\flat, w^\sharp]\!]} [cu - p\min\{u,w\}] = cu - p\min\{u, w^\flat\}$$

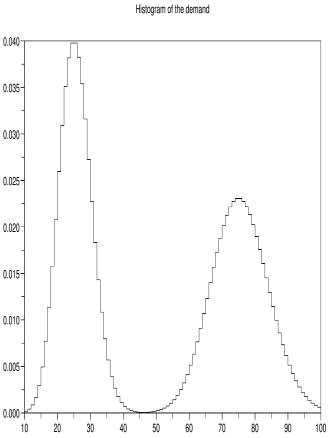
- Show that the order $u^* = w^\flat$ minimizes the above expression $J(u)$
- Once the newsvendor makes the optimal order $u^* = w^\flat$, the optimal costs are

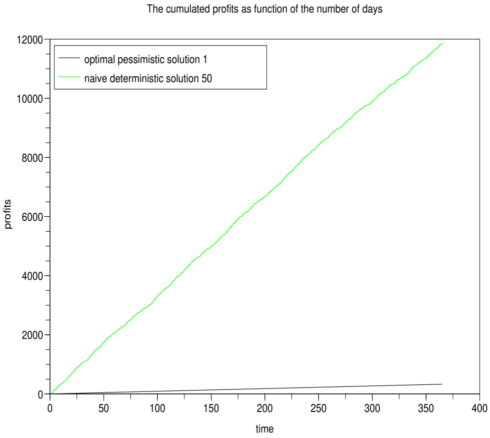
$$j(u^*, \cdot) : w \in [\![w^\flat, w^\sharp]\!] \mapsto -(p-c)w^\flat$$

which, here, are no longer uncertain

Does it pay to be so pessimistic?

Not if demands are drawn independently from a probability distribution





If the newsvendor minimizes the expected costs

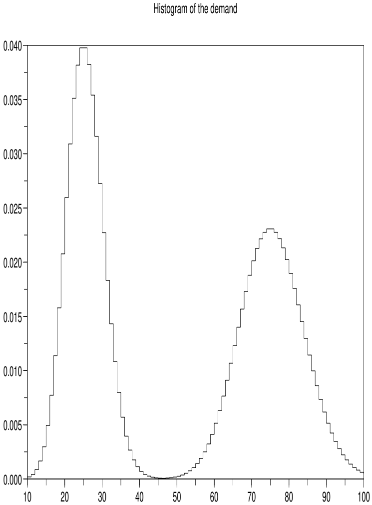
- We suppose that
  - the demand is a random variable, denoted $\mathbf{W}$
  - the newsvendor knows
    the probability distribution $\mathbb{P}_{\mathbf{W}}$ of the demand $\mathbf{W}$
- The expected costs are

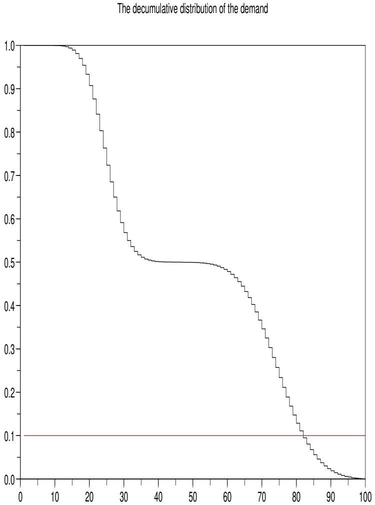
$$J(u) = \mathbb{E}_{\mathbf{W}}[j(u, \mathbf{W})] = \mathbb{E}_{\mathbf{W}}[cu - p \min\{u, \mathbf{W}\}]$$

- Find an order u^* which minimizes the above expression $J(u)$
  - the optimal order u^* can be characterized
  - using the *decumulative distribution function* $u \mapsto \mathbb{P}(\mathbf{W} > u)$

$$\mathbb{P}(\mathbf{W} > u^\star) \approx \frac{c}{p}$$

Here is an example of probability distribution and of decumulative distribution for the demand





Comments

- The optimal order u^* depends on both
  - the cost to price ratio $\frac{c}{p}$
  - the probability distribution $\mathbb{P}_{\mathbf{W}}$ of the demand $\mathbf{W}$
- How does the order u^* vary when
  - cost to price ratio $\frac{c}{p}$ increases?
  - demand $\mathbf{W}$ increases?
- How does one prove the result?
  (simpler in the continuous case)

"Greenwashing" the (single-period) newsvendor problem

- We formulate the determination of the level of energy reserves in a day-ahead market as a one stage stochastic optimization problem
- A decision has to be made at night of day D: which quantity $u \in \mathcal{U} = \mathbb{R}_+$ of energy has to be mobilized to meet a demand that will materialize at morning of day $D+1$?
- Demand is a random variable $\mathbf{W} \in \mathcal{W} = \mathbb{R}_+$ with density f on $\mathbb{R}_+$

$$\mathbb{P}_{\mathbf{W}}([a,b]) = \mathbb{P}\big(\mathbf{W} \in [a,b]\big) = \int_a^b f(w)dw \; , \;\; \forall \, 0 \leq a \leq b \leq +\infty$$

- The vendor's costs $j(u,w)$ depend both on the decision $u \in \mathbb{R}_+$ and on the uncertainty $w \in \mathbb{R}_+$:

$$\begin{aligned} j(u,w) &= \overbrace{cu}^{\text{purchasing}} - \overbrace{p\min\{u,w\}}^{\text{selling}} \\ &= (cu - pu)\mathbf{1}_{u \leq w} + (cu - pw)\mathbf{1}_{u > w} \end{aligned}$$

Proof in the continuous case

The function

$$J(u) = \mathbb{E}_{\mathbb{P}}[j(u, \mathbf{W})] = (cu - pu) \int_u^{+\infty} f(w)dw + \int_0^u (cu - pw) f(w) dw$$

has derivative

$$J'(u) = c - p \underbrace{\int_u^{+\infty} f(w)dw}_{\mathbb{P}(\mathbf{W} > u^\star)}$$

which is an increasing function of u, hence the function J is convex

As $J'(0) = c - p < 0$ and $\lim_{u \to +\infty} J'(u) = c > 0$,
the function J has a minimum $u^\star$ at

$$\mathbb{P}(\mathbf{W} > u^\star) = \frac{c}{p}$$

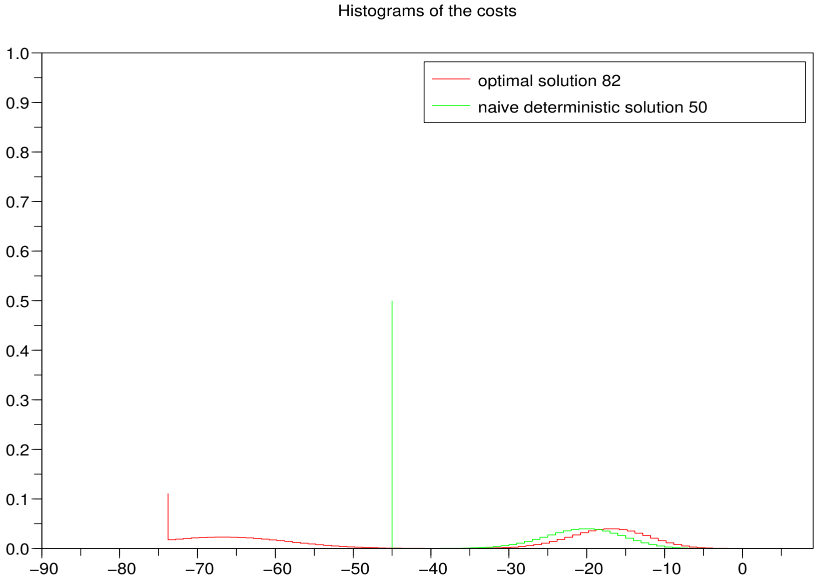
Extension

- Unsold unit costs $h \geq 0$ (*holding* cost)
- Undelivered unit costs $b \geq 0$ (*unsatisfaction* cost)

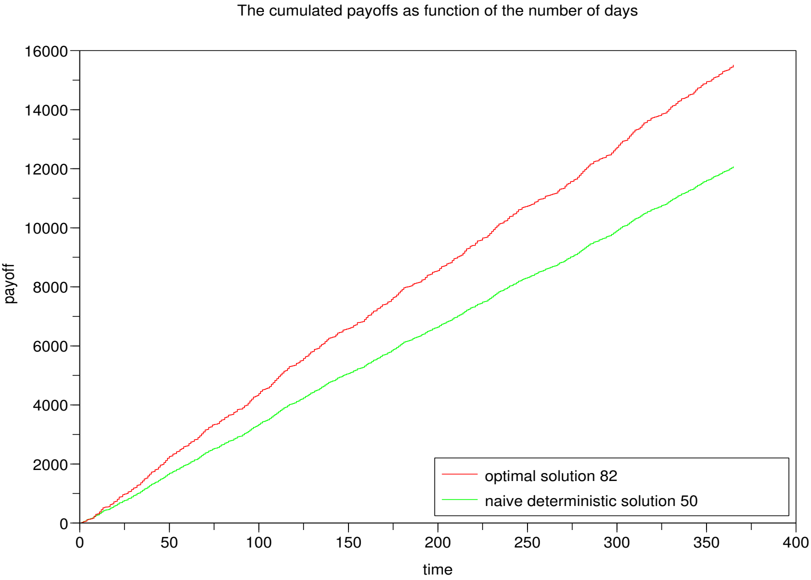
$$j(u, w) = cu - p \min\{u, w\} + h \max\{0, u - w\} + b \max\{0, -u + w\}$$

$$J'(u) = (c + h) - (p + b + h) \int_{u}^{+\infty} f(w)dw$$

The distribution of the optimal costs displays lower costs than with the naive deterministic solution $u = \mathbb{E}[\mathbf{W}]$



The cumulated *profits* over 365 days reveal that it pays to do stochastic optimization



Outline of the presentation

The "deterministic" solution is optimal for the "deterministic" criterion

When you insert the mean value $\overline{\mathbf{W}} = \mathbb{E}_{\mathbf{W}}[\mathbf{W}]$ into the cost function

$$j(u, w) \hookrightarrow j(u, \overline{\mathbf{W}})$$

- you obtain the "deterministic" criterion

$$\overline{J}(u) = j(u, \overline{\mathbf{W}})$$

- hence the "deterministic" optimization problem

$$\min_{u \in \mathcal{U}} \overline{J}(u) = \min_{u \in \mathcal{U}} j(u, \overline{\mathbf{W}})$$

- and a "deterministic" optimal solution $\overline{u}$ that solves

$$\overline{J}(\overline{u}) = j(\overline{u}, \overline{\mathbf{W}}) = \min_{u \in \mathcal{U}} j(u, \overline{\mathbf{W}})$$

The "stochastic" solution is optimal for the "stochastic" criterion

When you insert the random variable $\mathbf{W}$ into the cost function

$$j(u,w) \hookrightarrow j(u,\mathbf{W})$$

- you obtain the "stochastic" criterion

$$\widetilde{J}(u) = \mathbb{E}_{\mathbf{W}}[j(u,\mathbf{W})]$$

- hence the "stochastic" optimization problem

$$\min_{u \in \mathcal{U}} \widetilde{J}(u) = \min_{u \in \mathcal{U}} \mathbb{E}_{\mathbf{W}}[j(u,\mathbf{W})]$$

- and a "stochastic" optimal solution $\widetilde{u}$ that solves

$$\widetilde{J}(\widetilde{u}) = \mathbb{E}_{\mathbf{W}}[j(\widetilde{u},\mathbf{W})] = \min_{u \in \mathcal{U}} \mathbb{E}_{\mathbf{W}}[j(u,\mathbf{W})]$$

Optimality is relative to a criterion

	solution	
	"deterministic" $\overline{u}$	"stochastic" $\widetilde{u}$
"deterministic" criterion $\overline{J}$	optimal	suboptimal
"stochastic" criterion $\widetilde{J}$	suboptimal	optimal

Optimality is relative to a criterion

	solution	
	"deterministic" $\overline{u}$	"stochastic" $\widetilde{u}$
"deterministic" criterion $\overline{J}$	$j(\overline{u}, \overline{\mathbf{W}})$ $\leq$	$j(\widetilde{u}, \overline{\mathbf{W}})$
"stochastic" criterion $\widetilde{J}$	$\mathbb{E}_{\mathbf{W}}[j(\overline{u}, \mathbf{W})]$ $\geq$	$\mathbb{E}_{\mathbf{W}}[j(\widetilde{u}, \mathbf{W})]$

Interpretation problems occur when one compares values $\overline{J}(u)$ and $\widetilde{J}(u)$, instead of solutions $\overline{u}$ and $\widetilde{u}$

Optimality is relative to a criterion

- The "deterministic" optimal solution $\overline{u}$ achieves lower "deterministic" costs than the "stochastic" optimal solution $\widetilde{u}$

$$j(\overline{u},\overline{\mathbf{W}}) = \min_{u \in \mathcal{U}} j(u,\overline{\mathbf{W}}) \leq j(\widetilde{u},\overline{\mathbf{W}})$$

- The "stochastic" optimal solution $\widetilde{u}$ achieves lower "expected" costs than the "deterministic" optimal solution $\overline{u}$

$$\mathbb{E}_{\mathbf{W}}[j(\widetilde{u},\mathbf{W})] = \min_{u \in \mathcal{U}} \mathbb{E}_{\mathbf{W}}[j(u,\mathbf{W})] \leq \mathbb{E}_{\mathbf{W}}[j(\overline{u},\mathbf{W})]$$

- Interpretation problems occur when one confuses solutions and criteria

When the solution of a deterministic optimization problem looks (wrongly) optimistic

- The "deterministic" optimal solution $\overline{u}$ seems to achieve less costs than the "stochastic" optimal solution $\widetilde{u}$ because

$$\underbrace{j(\overline{u}, \overline{\mathbf{W}}) = \min_{u \in \mathcal{U}} j(u, \overline{\mathbf{W}})}_{-44.968856} \leq \underbrace{\mathbb{E}_{\mathbf{W}}[j(\widetilde{u}, \mathbf{W})] = \min_{u \in \mathcal{U}} \mathbb{E}_{\mathbf{W}}[j(u, \mathbf{W})]}_{-41.259519}$$

- But this (true) inequality cannot sustain a comparison between solutions because the criterion has changed

$$\overbrace{\underbrace{j(\overline{u}, \overline{\mathbf{W}}) = \min_{u \in \mathcal{U}} j(u, \overline{\mathbf{W}})}_{\text{"deterministic" criterion}}}^{\text{"deterministic" solution}} \leq \overbrace{\underbrace{\mathbb{E}_{\mathbf{W}}[j(\widetilde{u}, \mathbf{W})] = \min_{u \in \mathcal{U}} \mathbb{E}_{\mathbf{W}}[j(u, \mathbf{W})]}_{\text{"stochastic" criterion}}}^{\text{"stochastic" solution}}$$

To asses the solutions of a stochastic optimization problem you need a proper stochastic benchmark

- In fact, the "deterministic" optimal solution $\overline{u}$ achieves lower expected costs than the "stochastic" optimal solution $\widetilde{u}$ because

$$\underbrace{\mathbb{E}_{\mathbf{W}}[j(\widetilde{u},\mathbf{W})] = \min_{u\in\mathcal{U}} \mathbb{E}_{\mathbf{W}}[j(u,\mathbf{W})]}_{-41.259519} \leq \underbrace{\mathbb{E}_{\mathbf{W}}[j(\overline{u},\mathbf{W})]}_{-32.498824}$$

- and the full picture is the following

$$\underbrace{j(\overline{u},\overline{\mathbf{W}}) = \min_{u\in\mathcal{U}} j(u,\overline{\mathbf{W}})}_{-44.968856} \leq \underbrace{\mathbb{E}_{\mathbf{W}}[j(\widetilde{u},\mathbf{W})] = \min_{u\in\mathcal{U}} \mathbb{E}_{\mathbf{W}}[j(u,\mathbf{W})]}_{-41.259519} \leq \underbrace{\mathbb{E}_{\mathbf{W}}[j(\overline{u},\mathbf{W})]}_{-32.498824}$$

When deterministic optimization is (wrongly) optimistic

Let $\mathbf{W}$ be a random variable with mean $\overline{\mathbf{W}} = \mathbb{E}_{\mathbf{W}}[\mathbf{W}]$, and suppose that $w \mapsto j(u, w)$ is convex, for all decision u. Then, by Jensen inequality,

$$\underbrace{\inf_{u \in \mathcal{U}} j(u, \mathbb{E}_{\mathbf{W}}[\mathbf{W}])}_{\text{"deterministic" optimization problem}} \leq \underbrace{\inf_{u \in \mathcal{U}} \mathbb{E}_{\mathbf{W}}[j(u, \mathbf{W})]}_{\text{"stochastic" optimization problem}}$$

- If we suppose that the infima are minima, this gives

$$\underbrace{j(\overline{u}, \overline{\mathbf{W}})}_{\substack{\text{"deterministic"} \\ \text{optimal solution}}} = \min_{u \in \mathcal{U}} j(u, \overline{\mathbf{W}}) \leq \underbrace{\mathbb{E}_{\mathbf{W}}[j(u^*, \mathbf{W})]}_{\substack{\text{"stochastic"} \\ \text{optimal solution}}} = \min_{u \in \mathcal{U}} \mathbb{E}_{\mathbf{W}}[j(u, \mathbf{W})]$$

- we immediately deduce that the "deterministic" optimal costs are less than the "expected" optimal costs

$$\overbrace{j(\overline{u}, \overline{\mathbf{W}}) \leq}^{\text{overly optimistic}} \ \mathbb{E}_{\mathbf{W}}[j(u^*, \mathbf{W})] \overbrace{\leq \mathbb{E}_{\mathbf{W}}[j(\overline{u}, \mathbf{W})]}^{\text{wrongly optimistic}}$$

Thus, with an improper benchmark, you may jump to wrong conclusions

Where do we stand after having worked out two examples?

- When you move from deterministic optimization
  to optimization under uncertainty,
  you come accross the issue of risk attitudes
- Risk is in the eyes of the beholder ;-)
  and materializes in the a priori knowledge on the uncertainties
  - either probabilistic/stochastic
    - independence and Bernoulli distributions in the blood test example
    - uncertain demand faced by the newsvendor
      modeled as a random variable
  - or set-membership
    - uncertain demand faced by the newsvendor modeled by a set
- In the end, when doing stochastic (cost) minimization,
  selecting a "good" decision among many
  resorts to selecting a "good" histogram of costs among many

Where have we gone till now? And what comes next

- We have seen two examples of optimization problems
  with a single deterministic decision variable,
  and with a criterion including a random variable
- Now, we will turn to optimization problems
  with two decision variables,
  the first one deterministic and the second one random

Outline of the presentation

What awaits us

- We will lay out two ways to move
  from one-stage deterministic optimization problems
  to two-stage stochastic linear programs
  - in one, we start from a deterministic convex piecewise linear program
    (without constraints)
  - in the other, we start from a deterministic linear program
    with constraints
- We will outline the L-shaped method
  to solve such two-stage linear stochastic programs

Outline of the presentation

We revisit the newsvendor problem

Writing the newsvendor problem as a linear program, in three steps

- We consider the stochastic optimization problem

$$\min_{u \in \mathbb{R}} J(u) = \mathbb{E}_{\mathbb{P}}[j(u, \mathbf{W})]$$

- where the decision variable u takes continuous real values, and

$$j(u, w) = cu - p \min\{u, w\}$$

- and we show in three steps how to rewrite this problem as a linear program

Step 1: exploiting convex piecewise linearity of the criterion

First, we write

$$\begin{aligned} j(u,w) &= cu - p\min\{u,w\} \\ &= \max\{cu - pu, cu - pw\} \\ &= \min_{v\in\mathbb{R}}\{v \mid v \geq cu - pu \;,\;\; v \geq cu - pw\} \end{aligned}$$

Step 2: exploiting convexity of the mathematical expectation

- We suppose that the demand $\mathbf{W}$ can take a finite number S of possible values $\{w^s, s \in \mathcal{S}\}$
- where s denotes a scenario in the finite set $\mathcal{S}$ (S=card($\mathcal{S}$))
- and we denote π^s the probability of scenario s, with

$$\sum_{s\in\mathcal{S}} \pi^s = 1 \text{ and } \pi^s \geq 0 \ , \ \ \forall s \in \mathcal{S}$$

Step 2: exploiting convexity of the mathematical expectation

Second, we deduce

$$
\begin{aligned}
J(u) &= \mathbb{E}_{\mathbb{P}}[j(u, \mathbf{W})] \\
&= \sum_{s \in \mathcal{S}} \pi^s j(u, w^s) \\
&= \sum_{s \in \mathcal{S}} \pi^s \min_{v^s \in \mathbb{R}} \{ v^s \mid v^s \geq cu - pu \ , \ \ v^s \geq cu - pw^s \} \\
&= \min_{(v^s)_{s \in \mathcal{S}} \in \mathbb{R}^{\mathcal{S}}} \sum_{s \in \mathcal{S}} \pi^s v^s \\
&\quad \text{under the constraints} \\
&\quad v^s \geq cu - pu \ , \ \ v^s \geq cu - pw^s \ , \ \ \forall s \in \mathcal{S}
\end{aligned}
$$

Step 3: exploiting $\min \min = \min$

Third, we minimize with respect to the original decision $u \in \mathbb{U}$

$$\begin{aligned}\min_{u \in \mathbb{U}} J(u) = \min_{u \in \mathbb{U}, (v^s)_{s \in \mathcal{S}} \in \mathbb{R}^{\mathcal{S}}} & \sum_{s \in \mathcal{S}} \pi^s v^s \\ & v^s \geq cu - pu \;, \;\; \forall s \in \mathcal{S} \\ & v^s \geq cu - pw^s \;, \;\; \forall s \in \mathcal{S}\end{aligned}$$

This is a linear program

The revisited newsvendor problem example
is a special case of a general mechanism

From convex piecewise linear to linear programming

- The convex piecewise linear program (polyhedral)

$$\min_{x \in \mathbb{R}^n} \max_{i=1,\ldots,m} \langle c_i \mid x \rangle + b_i$$

- can be written as the linear program

$$\min_{x \in \mathbb{R}^n} \min_{v \in \mathbb{R}} v$$

$$v \geq \langle c_i \mid x \rangle + b_i \ , \qquad i = 1, \ldots, m$$

From stochastic convex piecewise linear programming to stochastic linear programming

- The stochastic convex piecewise linear program

$$\min_{x\in\mathbb{R}^n} \sum_{s\in\mathcal{S}} \pi^s \max_{i=1,\dots,m} \langle c_i^s \mid x \rangle + b_i^s$$

- can be written as the stochastic linear program

$$\min_{x\in\mathbb{R}^n} \min_{(v^s)_{s\in\mathcal{S}}\in\mathbb{R}^{\mathcal{S}}} \sum_{s\in\mathcal{S}} \pi^s v^s$$
$$v^s \geq \langle c_i^s \mid x \rangle + b_i^s \ , \quad i = 1, \dots, m \ , \ \ s \in \mathcal{S}$$

Outline of the presentation

We revisit the newsvendor problem
when she/he is offered the possibility
to adjust after observing the demand

We change the newsvendor problem by adding a constraint

- We consider the stochastic optimization problem

$$\min_{\substack{u \in \mathbb{R} \\ u \geq \mathbf{W}}} J(u) = \mathbb{E}_{\mathbb{P}}[j(u, \mathbf{W})]$$

- where the decision variable u takes continuous real values and must satisfy the constraint $u \geq \mathbf{W}$
- and where the cost function is now

$$j(u, w) = cu - pw$$

The solution is over conservative

- If we suppose that the demand $\mathbf{W}$ can take a finite number S of possible values w^s, $s \in \mathcal{S}$
  - where s denotes a *scenario* in the finite set $\mathcal{S}$ (S=card($\mathcal{S}$))
  - and we denote π^s the probability of scenario s, with

$$\sum_{s \in \mathcal{S}} \pi^s = 1 \text{ and } \pi^s > 0 \; , \;\; \forall s \in \mathcal{S}$$

- then the stochastic optimization problem becomes

$$\min_{u \in \mathbb{R}} \sum_{s \in \mathcal{S}} \pi^s j(u, w^s)$$

under the constraints

$$u \geq w^s \; , \;\; \forall s \in \mathcal{S}$$

- with (pessimistic) solution $u^* = \max_{s \in \mathcal{S}} w^s$

One way out consists in offering the newsvendor a second (recourse) decision

- In the morning,
  the newsvendor can order a quantity $u_0 \in \mathbb{R}_+$ of product,
  at unitary cost $c_0 > 0$
- In the afternoon,
  the newsvendor can order a quantity $u_1 \in \mathbb{R}_+$ of product,
  at unitary cost $c_1 > c_0 > 0$
- The constraints are now

$$u_0 + u_1 \geq \mathbf{W}$$

- and the cost function is now

$$j(u_0, u_1, w) = c_0 u_0 + c_1 u_1 - pw$$

Writing the newsvendor problem with recourse

- In the formulation

$$\min_{\substack{u_0 \in \mathbb{R} \\ \{u_1^s\}_{s \in \mathcal{S}} \in \mathbb{R}^{\mathcal{S}}}} \sum_{s \in \mathcal{S}} \pi^s j(u_0, u_1^s, w^s)$$

under the constraints

$$u_0 + u_1^s \geq w^s \ , \quad \forall s \in \mathcal{S}$$

- we express the fact that
  - the decision u_0 is the first one, made before the demand materializes
  - the decisions u_1^s are the second ones, made after the demand materializes

The revisited newsvendor problem example
is a special case of a general mechanism

From linear to stochastic programming

- The linear program

$$
\begin{aligned}
\min_{x \in \mathbb{R}^n} & \langle c \mid x \rangle \\
Ax + b & \geq 0 \quad (\in \mathbb{R}^m)
\end{aligned}
$$

- becomes a stochastic program

$$
\begin{aligned}
\min_{x \in \mathbb{R}^n} & \sum_{s \in \mathcal{S}} \pi^s \langle c^s \mid x \rangle \\
A^s x + b^s & \geq 0 \,, \qquad \forall s \in \mathcal{S}
\end{aligned}
$$

- We observe that there are as many (vector) inequalities as there are possible scenarios $s \in \mathcal{S}$

$$
A^s x + b^s \geq 0 \,, \;\; \forall s \in \mathcal{S}
$$

and these inequality constraints can delineate an empty domain for optimization

Recourse variables need be introduced for feasability issues

- We introduce a recourse variable $y = \{y^s\}_{s \in \mathcal{S}}$ and the program

$$\min_{x, \{y^s\}_{s \in \mathcal{S}}} \sum_{s \in \mathcal{S}} \pi^s \Big(\langle c^s \mid x \rangle + \langle p^s \mid y^s \rangle \Big)$$

$$\begin{aligned} y^s &\geq 0 \;, \quad \forall s \in \mathcal{S} \\ A^s x + b^s + y^s &\geq 0 \;, \quad \forall s \in \mathcal{S} \end{aligned}$$

- so that the inequality $A^s x + b^s + y^s \geq 0$ is now possible, at (unitary recourse) price vector $p = \{p^s\}_{s \in \mathcal{S}}$
- Observe that such stochastic programs are huge problems, with solution $\big(x, \{y^s\}_{s \in \mathcal{S}}\big)$, but remain linear

Two-stage stochastic programs with recourse can become deterministic non-smooth convex problems

- The following function of x is convex, but nonsmooth

$$\underbrace{Q^s(x)}_{\text{value function}} = \min\{\langle p^s \mid y\rangle, y \geq 0, A^s x + b^s + y \geq 0\}$$

- The original two-stage stochastic program with recourse

$$\min_{x,\{y^s\}_{s\in\mathcal{S}}} \sum_{s\in\mathcal{S}} \pi^s \big[\langle c^s \mid x\rangle + \langle p^s \mid y^s\rangle\big]$$

$$\begin{aligned} y^s &\geq 0\,, \quad \forall s \in \mathcal{S} \\ A^s x + b^s + y^s &\geq 0\,, \quad \forall s \in \mathcal{S} \end{aligned}$$

  now becomes the deterministic nonsmooth convex program

$$\min_{x} \sum_{s\in\mathcal{S}} \pi^s \big[\langle c^s \mid x\rangle + Q^s(x)\big]$$

- An optimal solution is now more likely to be an inner solution (more robust)

Outline of the presentation

Roger Wets example

http://cermics.enpc.fr/~delara/TEACHING_PAST/CEA-EDF-INRIA_2012/Roger_Wets1.pdf

Robustification and convexification

A linear problem in a deterministic framework

Two (normalized) actions x_1, x_2 of decarbonization, with

- $(x_1, x_2) \in \Delta = \{(x_1, x_2) \mid 0 \leq x_1, x_2 \;,\;\; x_1 + x_2 \leq 1\}$ (simplex) (third action $x_3 \geq 0$ corresponds to the statu quo, with $x_1 + x_2 + x_3 = 1$)
- respective unitary costs c_1, c_2
- respective unitary emissions reductions e_1, e_2
- emissions reduction target $e^{\#}$

$$\begin{aligned} \min_{(x_1,x_2)\in\Delta} \quad & c_1 x_1 + c_2 x_2 \\ \text{s.t.} \quad & e_1 x_1 + e_2 x_2 \geq e^{\#} \qquad \text{(emissions reductions)} \end{aligned}$$

For instance, in a taxi company, x_1 and x_2 represent fractions of vehicles switched from thermal to electric or hybrid

Solutions (extreme) of the deterministic approach

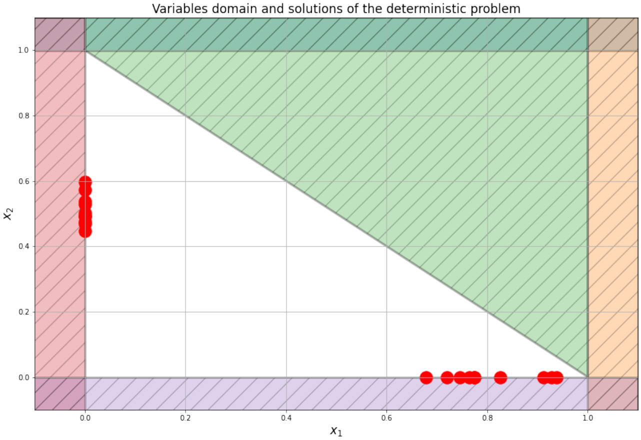


Figure: Variables domain and solutions of the deterministic approach

Fomulation of the multi-scenario approach

- We consider
  - a finite set $\mathcal{S}$ of scenarios (future uncertainties)
  - a family $\{e_1^s, e_2^s, c_1^s, c_2^s, p^s\}_{s \in S}$ of possible values for unitary emissions reduction factors e_1^s, e_2^s, unitary costs c_1^s, c_2^s, and for the price p^s of CO_2 emission rights
  - a family $\{\pi^s\}_{s \in S}$ of nonnegative numbers summing to one, where π^s represents the probability of the scenario s
- and we set the stochastic optimization problem, with a new recourse decision variable q^s, representing buying emission rights after uncertainty is resolved

$$
\begin{aligned}
\min_{(x_1,x_2)\in\Delta, \{q^s\}_{s\in S}\in\mathbb{R}_+^S} \quad & \sum_{s\in\mathbb{S}} \pi^s \Big[c_1^s x_1 + c_2^s x_2 + p^s \overbrace{q^s}^{\text{emission rights}} \Big] \\
\text{s.t.} \quad & e_1^s x_1 + e_2^s x_2 + q^s \geq e^{\#} \ , \quad \forall s \in \mathbb{S}
\end{aligned}
$$

Fomulation of the multi-scenario approach

- We consider
  - a finite set $\mathcal{S}$ of scenarios (future uncertainties)
  - a family $\{e_1^s, e_2^s, c_1^s, c_2^s, p^s\}_{s \in S}$ of possible values for unitary emissions reduction factors e_1^s, e_2^s, unitary costs c_1^s, c_2^s, and for the price p^s of CO_2 emission rights
  - a family $\{\pi^s\}_{s \in S}$ of nonnegative numbers summing to one, where π^s represents the probability of the scenario s
- and we set the stochastic optimization problem, with a new recourse decision variable q^s, representing buying emission rights after uncertainty is resolved

$$\min_{(x_1,x_2)\in\Delta, \{q^s\}_{s\in S}\in\mathbb{R}_+^S} \quad \sum_{s\in\mathbb{S}} \pi^s [c_1^s x_1 + c_2^s x_2 + p^s \overbrace{q^s}^{\text{emission rights}}]$$

$$\text{s.t.} \quad e_1^s x_1 + e_2^s x_2 + q^s \geq e^{\#} \ , \quad \forall s \in \mathbb{S}$$

$$\Updownarrow$$

$$\min_{(x_1,x_2)\in\Delta} \bar{c}_1 x_1 + \bar{c}_2 x_2 + \sum_{s\in\mathbb{S}} \pi^s p^s \overbrace{[e^{\#} - e_1^s x_1 - e_2^s x_2]_+}^{\text{convexification term}}$$

Solution (inner) of the stochastic approach

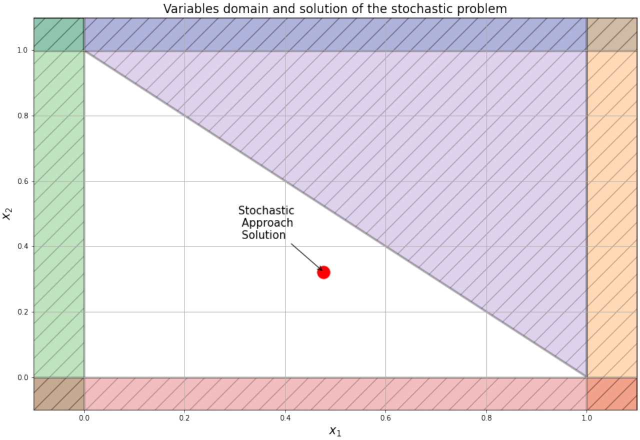


Figure: Variables domain and solution of the stochastic approach

A quadratic toy problem

A quadratic toy problem

Let $c > 0$, $d_1 \geq 0$, $d_2 \geq 0$

- Show that the (worst case) optimization problem

$$\begin{aligned} &\min_{x \in \mathbb{R}} \frac{1}{2} c x^2 \\ &x \geq d_1 \\ &x \geq d_2 \end{aligned}$$

  has (worst case) solution

$$\bar{x} = \max\{d_1, d_2\}$$

- What happens if we allow room for recourse?

A quadratic toy problem with recourse

Let $c > 0$, $d_1 \geq 0$, $d_2 \geq 0$, $p_1 > 0$, $p_2 > 0$

- Show that the (stochastic) optimization problem

$$\begin{aligned} \min_{(x,y_1,y_2)\in\mathbb{R}^3} \frac{1}{2}\Big(cx^2 + p_1 y_1^2 + p_2 y_2^2\Big) \\ x + y_1 = d_1 \\ x + y_2 = d_2 \end{aligned}$$

  has a solution x^* given by

$$x^* = \frac{p_1}{c + p_1 + p_2} d_1 + \frac{p_2}{c + p_1 + p_2} d_2 + \frac{c}{c + p_1 + p_2} 0$$

- Therefore, x^* belongs to the convex generated by $\{0, d_1, d_2\}$, that is,

$$x^* \in [0, \max\{d_1, d_2\}]$$

- Compare with the (worst case) solution $\bar{x} = \max\{d_1, d_2\}$

Two stage stochastic optimization for fixing energy reserves

Two stage stochastic optimization for fixing energy reserves

- We formulate the determination of the level of energy reserves in a day-ahead market as a two stage stochastic optimization problem
- A decision has to be made at night of day D: which quantity of the cheapest energy production units (reserve) has to be mobilized to meet a demand that will materialize at morning of day $D+1$?
- Excess reserves are penalized
- Demand unsatisfied by reserves has to be covered by costly extra units (recourse variables)

Hence, there is a trade-off to be assessed by optimization

Stages

There are two stages, represented by the letter t (for time)

- $t = 0$ corresponds to night of day D
- $t = 1$ corresponds to morning of day $D + 1$

Probabilistic model

- Demand, materialized on the morning of day $D+1$, takes a finite number S of possible values w^s, where s denotes a scenario in the finite set $\mathcal{S}$ (S=card($\mathcal{S}$))
- π^s is the probability of scenario s

$$\forall s \in \mathcal{S} \; , \;\; \pi^s > 0 \; , \;\; \sum_{s \in \mathcal{S}} \pi^s = 1$$

- Notice that we do not consider scenarios with zero probability

Decision variables

- The decision variables are
  - the scalar Q_0 (reserve)
  - the finite family $(Q_1^s)_{s \in \mathcal{S}}$ of scalars (recourse variables)

  where
  - at stage $t = 0$, the energy reserve is Q_0
  - at stage $t = 1$, a scenario s materializes
    and the demand w^s is observed,
    so that one decides of the recourse quantity Q_1^s
    knowing the demand w^s
- The decision variables can be considered as indexed by a tree with
  - one root (corresponding to the index 0):
    Q_0 is attached to the root of the tree
  - and as many leafs as scenarios in $\mathcal{S}$
    (each leaf corresponding to the index $1, s$) :
    each Q_1^s is attached to the leaf corresponding to s

Optimization problem formulation

- The balance equation between supply and demand is

$$Q_0 + Q_1^s = w^s \ , \ \ \forall s \in \mathcal{S}$$

- Energies mobilized at stages $t = 0$ and $t = 1$ differ in terms of capacities and costs
  - at stage $t = 0$, the energy production
    - has maximal capacity $Q_0^\sharp$
    - costs $c_0(Q_0)$ to produce the quantity Q_0
  - at stage $t = 1$, the energy production
    - has unbounded capacity
    - costs $c_1(Q_1)$ to produce the quantity Q_1

Optimization problem formulation

We formulate the stochastic optimization problem

$$\begin{aligned}
\min_{Q_0, \{Q_1^s\}_{s \in \mathcal{S}}} \quad & \sum_{s \in \mathcal{S}} \pi^s \left[c_0(Q_0) + c_1(Q_1^s) \right] \\
\text{s.t.} \quad & 0 \leq Q_0 \leq Q_0^\sharp \\
& 0 \leq Q_1^s && \forall s \in \mathcal{S} \\
& w^s = Q_0 + Q_1^s && \forall s \in \mathcal{S}
\end{aligned}$$

- Here, we look for energy reserve Q_0 and recourse energy Q_1^s so that the balance equation is satisfied (at stage $t = 1$) at minimum expected cost
- By weighing each scenario s with its probability π^s, the optimal solution $(Q_0^*, (Q_1^{s*})_{s \in \mathcal{S}})$ performs a compromise between scenarios

Outline of the presentation

Stochastic linear program

- We write the stochastic linear program

$$\min_{x,\{y^s\}_{s\in\mathcal{S}}} \sum_{s\in\mathcal{S}} \pi^s \Big(\langle c^s \mid x \rangle + \langle p^s \mid y^s \rangle \Big)$$
$$x \geq 0$$
$$Ax = b$$
$$T^s x + W^s y^s = h^s \ , \quad \forall s \in \mathcal{S}$$

- as a one-stage program

$$\min_{x} \sum_{s\in\mathcal{S}} \pi^s \Big(\langle c^s \mid x \rangle + Q^s(x) \Big)$$
$$x \geq 0$$
$$Ax = b$$

- where the second-stage value function Q^s is given by

$$\forall s \in \mathcal{S} \ , \quad Q^s(x) = \min_{y^s} \langle p^s \mid y^s \rangle$$
$$T^s x + W^s y^s = h^s$$

See the slides for the L-shaped method by Vincent Leclère

Where have we gone till now? And what comes next

- We have arrived at optimization problems
  with two decision variables
  - a first one deterministic
  - a second one random (as it is indexed by the scenarios)
- We have presented a resolution method adapted to the linear case
- No, we move to possibly
  nonlinear two stage stochastic optimization problems
- We will present resolution methods that, somehow surprisingly,
  relax the assumption that the first decision variable is deterministic

Outline of the presentation

What awaits us

- We present a general form of two-stage stochastic programs and we discuss different forms of the nonanticipativity constraint
- We show a scenario decomposition resolution method adapted to two-stage stochastic programs that are strongly convex
- We outline the Progressive Hedging resolution method, adapted to two-stage stochastic linear programs

Outline of the presentation

Finite scenarios case
Nonanticipativity constraint

The finite scenarios case

- Probability space $\left(\mathcal{S}, 2^{\mathcal{S}}, \{\pi^s\}_{s\in\mathcal{S}}\right)$,
  where s denotes a scenario in the finite set $\mathcal{S}$
  and π^s is the probability of scenario s, with

$$\sum_{s\in\mathcal{S}} \pi^s = 1 \text{ and } \pi^s > 0 \; , \quad \forall s \in \mathcal{S}$$

- Decision random variables
  $\mathbf{U}_0 : \mathcal{S} \to \mathcal{U}_0$, $\mathbf{U}_1 : \mathcal{S} \to \mathcal{U}_1$, that is,
  $\mathbf{U}_0 = \{u_0^s\}_{s\in\mathcal{S}} \in \mathcal{U}_0^{\mathcal{S}}$, $\mathbf{U}_1 = \{u_1^s\}_{s\in\mathcal{S}} \in \mathcal{U}_1^{\mathcal{S}}$

Nonanticipativity constraint (finite scenarios case)

- Probability space $\left(\mathcal{S}, 2^{\mathcal{S}}, \{\pi^s\}_{s\in\mathcal{S}}\right)$
- Real-valued decision random variables
  $\mathbf{U}_0 : \mathcal{S} \to \mathcal{U}_0 = \mathbb{R}^{n_0}$, $\mathbf{U}_1 : \mathcal{S} \to \mathcal{U}_1 = \mathbb{R}^{n_1}$, that is,
  $\mathbf{U}_0 = \{u_0^s\}_{s\in\mathcal{S}} \in \mathcal{U}_0^{\mathcal{S}}$, $\mathbf{U}_1 = \{u_1^s\}_{s\in\mathcal{S}} \in \mathcal{U}_1^{\mathcal{S}}$

$$
\begin{aligned}
& \text{Nonanticipativity constraint} \\
\iff & \text{the random variable } \mathbf{U}_0 \text{ is deterministic} \\
\iff & \mathbf{U}_0 = \mathbb{E}(\mathbf{U}_0) \\
\iff & u_0^s = \sum_{s'\in\mathcal{S}} \pi^{s'} u_0^{s'} \ , \ \forall s \in \mathcal{S} \\
\iff & u_0^s = u_0^{s'} \ , \ \forall s \in \mathcal{S} \ , \ \forall s' \in \mathcal{S} \\
\iff & \exists u_0 \in \mathcal{U}_0 \ , \ u_0^s = u_0 \ , \ \forall s \in \mathcal{S}
\end{aligned}
$$

We formulate a two-stage stochastic optimization problem on a tree

- Data

$$\text{Criterion} \quad j : \underbrace{\mathcal{U}_0}_{\substack{\text{initial}\\ \text{decision}}} \times \overbrace{\mathcal{U}_1}^{\substack{\text{recourse}\\ \text{variable}}} \times \underbrace{\mathcal{S}}_{\text{scenario}} \to \mathbb{R} \cup (+\infty)$$

  and set-valued mapping $\mathcal{U}_1 : \mathcal{U}_0 \times \mathcal{S} \to 2^{\mathcal{U}_1}$

- Stochastic optimization problem

$$\begin{aligned} &\min_{u_0, \{u_1^s\}_{s \in \mathcal{S}}} \sum_{s \in \mathcal{S}} \pi^s j^s\big(u_0, u_1^s\big) \\ &u_0 \in \mathcal{U}_0 \\ &u_1^s \in \mathcal{U}_1^s\big(u_0\big) \ , \quad \forall s \in \mathcal{S} \end{aligned}$$

- Solutions $\big(u_0, \{u_1^s\}_{s \in \mathcal{S}}\big)$ are naturally indexed by a tree
  - with one root
  - and $S = \text{card}(\mathcal{S})$ leaves

Outline of the presentation

We start with a two-stage stochastic optimization problem formulated on a tree

$$\text{Criterion} \quad j : \underbrace{\mathcal{X}}_{\substack{\text{initial}\\\text{decision}}} \times \overbrace{\mathcal{Y}}^{\substack{\text{recourse}\\\text{variable}}} \times \underbrace{\mathcal{S}}_{\text{scenario}} \to \mathbb{R} \cup (+\infty)$$

and set-valued mapping $\mathcal{Y} : \mathcal{X} \times \mathcal{S} \to 2^{\mathcal{Y}}$

- Stochastic optimization problem

$$\begin{aligned} &\min_{x, \{y^s\}_{s \in \mathcal{S}}} \sum_{s \in \mathcal{S}} \pi^s j^s(x, y^s) \\ &x \in \mathcal{X} \\ &y^s \in \mathcal{Y}^s(x) \;, \quad \forall s \in \mathcal{S} \end{aligned}$$

- Solutions $(x, \{y^s\}_{s \in \mathcal{S}})$ are naturally indexed by a tree
  - with one root
  - and $S = \mathrm{card}(\mathcal{S})$ leaves

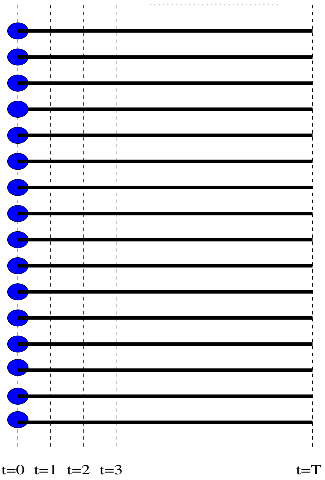
We transform the two-stage stochastic optimization problem by extending the solution space

- We consider initial decisions $\{x^s\}_{s\in\mathcal{S}}$ and the problem

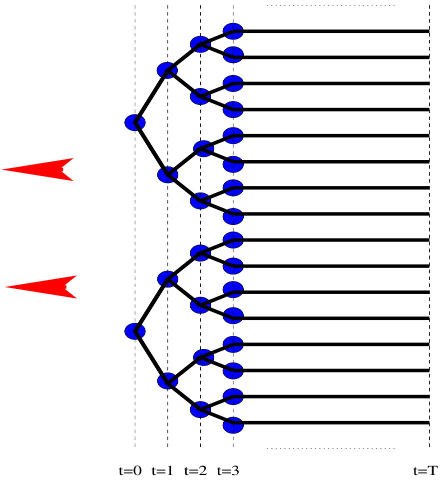
$$
\begin{aligned}
&\min_{x,\{x^s\}_{s\in\mathcal{S}},\{y^s\}_{s\in\mathcal{S}}} \sum_{s\in\mathcal{S}} \pi^s j^s\big(x^s, y^s\big) \\
&x^s \in \mathcal{X} \ , \ \ \forall s \in \mathcal{S} \\
&y^s \in \mathcal{Y}^s(x^s) \ , \ \ \forall s \in \mathcal{S} \\
&x^s = x \ , \ \ \forall s \in \mathcal{S} \\
&x \in \mathcal{X}
\end{aligned}
$$

- This problem has the same solutions $\big(x, \{y^s\}_{s\in\mathcal{S}}\big)$ as the original one

Scenarios can be organized like a fan or like a tree



N scenarios



Scenarios tree

We transform the two-stage stochastic optimization problem from a tree to a fan

- We consider initial decisions $\{x^s\}_{s \in \mathcal{S}}$ and the problem

$$\min_{\{x^s\}_{s\in\mathcal{S}}, \{y^s\}_{s\in\mathcal{S}}} \sum_{s\in\mathcal{S}} \pi^s j^s(x^s, y^s)$$
$$x^s \in \mathcal{X} \;, \;\; \forall s \in \mathcal{S}$$
$$y^s \in \mathcal{Y}^s(x^s) \;, \;\; \forall s \in \mathcal{S}$$
$$x^s = \sum_{s' \in \mathcal{S}} \pi^{s'} x^{s'} \;, \;\; \forall s \in \mathcal{S}$$

- Solutions $\{x^s, y^s\}_{s \in \mathcal{S}}$ are naturally indexed by a fan

Primal and dual problems

- The primal problem is

$$\begin{aligned}&\min_{\{x^s,y^s\}_{s\in\mathcal{S}}}\ \max_{\{\lambda^s\}_{s\in\mathcal{S}}} \sum_{s\in\mathcal{S}} \pi^s \Big(j^s(x^s, y^s) + \lambda^s \big(x^s - \sum_{s'\in\mathcal{S}} \pi^{s'} x^{s'}\big)\Big)\\ &x^s \in \mathcal{X}\ ,\ \ \forall s \in \mathcal{S}\\ &y^s \in \mathcal{Y}^s(x^s)\ ,\ \ \forall s \in \mathcal{S}\end{aligned}$$

- The dual problem is

$$\begin{aligned}&\max_{\{\lambda^s\}_{s\in\mathcal{S}}}\ \min_{\{x^s,y^s\}_{s\in\mathcal{S}}} \sum_{s\in\mathcal{S}} \pi^s \Big(j^s(x^s, y^s) + \lambda^s \big(x^s - \sum_{s'\in\mathcal{S}} \pi^{s'} x^{s'}\big)\Big)\\ &x^s \in \mathcal{X}\ ,\ \ \forall s \in \mathcal{S}\\ &y^s \in \mathcal{Y}^s(x^s)\ ,\ \ \forall s \in \mathcal{S}\end{aligned}$$

We can translate the multipliers λ^s in the dual problem

- Denote by $\mathbf{X} : \mathcal{S} \to \mathcal{X}$ the random variable $\mathbf{X}(s) = x^s$, $s \in \mathcal{S}$
- Denote by $\mathbf{\Lambda} : \mathcal{S} \to \mathbb{R}$ the random variable $\mathbf{\Lambda}(s) = \lambda^s$, $s \in \mathcal{S}$

$$
\begin{aligned}
& \sum_{s \in \mathcal{S}} \pi^s \lambda^s \big(x^s - \sum_{s' \in \mathcal{S}} \pi^{s'} x^{s'} \big) \\
= & \mathbb{E}\big[\mathbf{\Lambda} \big(\mathbf{X} - \mathbb{E}[\mathbf{X}] \big) \big] \\
= & \mathbb{E}\big[\mathbf{\Lambda} \mathbf{X} \big] - \mathbb{E}[\mathbf{\Lambda}] \mathbb{E}[\mathbf{X}] \\
= & \mathbb{E}\big[\big(\mathbf{\Lambda} - \mathbb{E}[\mathbf{\Lambda}] \big) \mathbf{X} \big] \\
= & \sum_{s \in \mathcal{S}} \pi^s \underbrace{\big(\lambda^s - \sum_{s' \in \mathcal{S}} \pi^{s'} \lambda^{s'} \big)}_{\text{projected multiplier } \overline{\lambda}^s} x^s
\end{aligned}
$$

Restricting the multiplier

Then the dual problem is

$$\max_{\{\lambda^s\}_{s\in\mathcal{S}}} \min_{\{x^s,y^s\}_{s\in\mathcal{S}}} \sum_{s\in\mathcal{S}} \pi^s \Big(j^s(x^s, y^s) + (\lambda^s - \sum_{s'\in\mathcal{S}} \pi^{s'} \lambda^{s'}) x^s \Big)$$
$$x^s \in \mathcal{X} \,, \ \forall s \in \mathcal{S}$$
$$y^s \in \mathcal{Y}^s(x^s) \,, \ \forall s \in \mathcal{S}$$

The dual problem can be decomposed scenario by scenario

- The dual problem

$$
\begin{aligned}
&\max_{\{\lambda^s\}_{s\in\mathcal{S}}} \min_{\{x^s,y^s\}_{s\in\mathcal{S}}} \sum_{s\in\mathcal{S}} \pi^s \Big(j^s(x^s, y^s) + \big(\lambda^s - \sum_{s'\in\mathcal{S}} \pi^{s'} \lambda^{s'}\big) x^s \Big) \\
&x^s \in \mathcal{X} \;, \quad \forall s \in \mathcal{S} \\
&y^s \in \mathcal{Y}^s(x^s) \;, \quad \forall s \in \mathcal{S}
\end{aligned}
$$

- is equivalent to

$$
\max_{\{\lambda^s\}_{s\in\mathcal{S}}} \sum_{s\in\mathcal{S}} \pi^s \quad
\begin{aligned}
&\min_{(x^s,y^s)} \Big(j^s(x^s, y^s) + \big(\lambda^s - \textstyle\sum_{s'\in\mathcal{S}} \pi^{s'} \lambda^{s'}\big) x^s \Big) \\
&x^s \in \mathcal{X} \\
&y^s \in \mathcal{Y}^s(x^s)
\end{aligned}
$$

Under proper assumptions
— to be seen later, as they require recalls in duality theory —
the dual problem can be solved by an algorithm "à la Uzawa"
yielding the following
scenario decomposition algorithm

Scheme of the scenario decomposition algorithm

Data: step $\rho > 0$, initial multipliers $\left\{\lambda^s_{(0)}\right\}_{s\in\mathcal{S}}$ and first decision $\bar{\mathbf{x}}_{(0)}$;
Result: optimal first decision $\mathbf{x}$;
repeat
  forall *scenarios* $s \in \mathcal{S}$ **do**
    Solve the deterministic minimization problem for scenario s, with a penalization $+\lambda^s_{(k)}\left(\mathbf{x}^s_{(k+1)} - \bar{\mathbf{x}}_{(k)}\right)$, and obtain optimal first decision $\mathbf{x}^s_{(k+1)}$;
  Update the mean first decisions

$$\bar{\mathbf{x}}_{(k+1)} = \sum_{s\in\mathcal{S}} \pi^s \mathbf{x}^s_{(k+1)} \;;$$

  Update the multipliers by

$$\lambda^s_{(k+1)} = \lambda^s_{(k)} + \rho\left(\mathbf{x}^s_{(k+1)} - \bar{\mathbf{x}}_{(k+1)}\right) \;, \;\; \forall s \in \mathcal{S} \;;$$

until $\mathbf{x}^s_{(k+1)} - \sum_{s'\in\mathcal{S}} \pi^{s'} \mathbf{x}^{s'}_{(k+1)} = 0 \;, \;\; \forall s \in \mathcal{S}$;

Outline of the presentation

Recalls and exercises on continuous optimization

http://cermics.enpc.fr/~delara/TEACHING/slides_optimization.pdf

Progressive Hedging

Rockafellar, R.T., Wets R. J-B.
Scenario and policy aggregation in optimization under uncertainty,
Mathematics of Operations Research, 16, pp. 119-147, 1991

http://cermics.enpc.fr/~delara/TEACHING/

CEA-EDF-INRIA_2012/Roger_Wets4.pdf

The "plus" of Progressive Hedging

- In addition to the variables x^s, we introduce a new variable $\bar{x}$, so that the non-anticipativity constraint becomes $x^s = \bar{x}$
- We dualize this constraint with an augmented Lagrangian term, yielding to an optimization problem with variables $x^{\cdot}, \bar{x}, \lambda$
- When the multiplier λ is fixed, we minimize the primal problem which, unfortunately, is not separable with respect to scenarios s
- Luckily, we recover separability by solving sequentially "à la Gauss-Seidel"

$$\min_{x^{\cdot}} \mathcal{L}(x^{\cdot}, \bar{x}_{(k)}, \lambda_{(k)})$$
$$\min_{\bar{x}} \mathcal{L}(x^{\cdot}_{(k+1)}, \bar{x}, \lambda_{(k)})$$

because the first problem is separable with respect to scenarios s

Scheme of the Progressive Hedging algorithm

Data: penalty $r > 0$, initial multipliers $\left\{\lambda^s_{(0)}\right\}_{s\in\mathcal{S}}$ and first decision $\bar{\mathbf{x}}_{(0)}$;

Result: optimal first decision $\mathbf{x}$;

repeat

forall *scenarios* $s \in \mathcal{S}$ **do**

Solve the deterministic minimization problem for scenario s, with penalization $+\lambda^s_{(k)}\left(\mathbf{x}^s_{(k+1)} - \bar{\mathbf{x}}_{(k)}\right) + \frac{r}{2}\left\|\mathbf{x}^s_{(k+1)} - \bar{\mathbf{x}}_{(k)}\right\|^2$, and obtain optimal first decision $\mathbf{x}^s_{(k+1)}$;

Update the mean first decisions

$$\bar{\mathbf{x}}_{(k+1)} = \sum_{s\in\mathcal{S}} \pi^s \mathbf{x}^s_{(k+1)} \ ;$$

Update the multipliers by

$$\lambda^s_{(k+1)} = \lambda^s_{(k)} + r\left(\mathbf{x}^s_{(k+1)} - \bar{\mathbf{x}}_{(k+1)}\right) \ , \ \ \forall s \in \mathcal{S} \ ;$$

until $\mathbf{x}^s_{(k+1)} - \sum_{s'\in\mathcal{S}} \pi^{s'} \mathbf{x}^{s'}_{(k+1)} = 0 \ , \ \ \forall s \in \mathcal{S}$;

Outline of the presentation

What awaits us

- We show how we can also obtain two-stage risk-averse programs, when we handle risk by means of the Tail Value at Risk

Outline of the presentation

What happens if we want to minimize risk, not mathematical expectation?

- Instead of minimizing the mathematical expectation

$$\mathbb{E}[\mathbf{C}] \quad (= \sum_{s \in \mathcal{S}} \pi^s \mathbf{C}^s)$$

- we want to minimize the Tail Value at Risk (at level $\lambda \in [0,1[$), given by the Rockafellar-Uryasev formula

$$TVaR_\lambda[\mathbf{C}] = \inf_{r \in \mathbb{R}} \left\{ \frac{\mathbb{E}[(\mathbf{C} - r)_+]}{1-\lambda} + r \right\}$$

- whose limit cases are mean and worst case

$$TVaR_0[\mathbf{C}] = \mathbb{E}[\mathbf{C}]$$
$$TVaR_1[\mathbf{C}] = \lim_{\lambda \to 1} TVaR_\lambda[\mathbf{C}] = \sup_{\omega \in \Omega} \mathbf{C}(\omega)$$

Minimizing the Tail Value at Risk of costs: convex piecewise linear programming formulation

- The risk-averse stochastic convex piecewise linear program

$$\min_{x\in\mathbb{R}^n}\min_{r\in\mathbb{R}}\left\{r+\frac{1}{1-\lambda}\sum_{s\in\mathcal{S}}\pi^s\Big(\max_{i=1,\ldots,m}\langle c_i^s \mid x\rangle + b_i^s - r\Big)_+\right\}$$

- can be written as the convex piecewise linear program

$$\begin{aligned}\min_{x\in\mathbb{R}^n}\min_{r\in\mathbb{R}}\min_{(u^s)_{s\in\mathcal{S}}\in\mathbb{R}^{\mathcal{S}}}\quad & r+\frac{1}{1-\lambda}\sum_{s\in\mathcal{S}}\pi^s(u^s-r)_+\\ & u^s\geq\langle c_1^s \mid x\rangle + b_1^s\ ,\ \ \forall s\in\mathcal{S}\\ & \vdots\\ & u^s\geq\langle c_m^s \mid x\rangle + b_m^s\ ,\ \ \forall s\in\mathcal{S}\end{aligned}$$

Minimizing the Tail Value at Risk of costs: linear programming formulation

- The risk-averse stochastic convex piecewise linear program

$$\min_{x \in \mathbb{R}^n} \min_{r \in \mathbb{R}} \left\{ r + \frac{1}{1-\lambda} \sum_{s \in \mathcal{S}} \pi^s \Big(\max_{i=1,\ldots,m} \langle c_i^s \mid x \rangle + b_i^s - r \Big)_+ \right\}$$

- can be written as the linear program

$$\begin{aligned}
\min_{x \in \mathbb{R}^n} \min_{r \in \mathbb{R}} \min_{(v^s)_{s \in \mathcal{S}} \in \mathbb{R}^{\mathcal{S}}} \quad & r + \frac{1}{1-\lambda} \sum_{s \in \mathcal{S}} \pi^s v^s \\
& v^s \geq \langle c_1^s \mid x \rangle + b_1^s - r \,, \quad \forall s \in \mathcal{S} \\
& \vdots \\
& v^s \geq \langle c_m^s \mid x \rangle + b_m^s - r \,, \quad \forall s \in \mathcal{S} \\
& v^s \geq 0 \,, \quad \forall s \in \mathcal{S}
\end{aligned}$$

How to use risk-averse stochastic programming in practice?

- Denote by x_λ^* the (supposed unique) solution
- As $1-\lambda$ measures the upper probability of risky events, start with $\lambda = 0$ and display, to the decision-maker, the risk-neutral solution x_0^* and the probability distribution (histogram) of the random costs

$$s \mapsto \max_{i=1,\dots,m} \langle c_i^s \mid x_0^* \rangle + b_i^s$$

- Then move to the confidence level $\lambda = 0.99$ (only events with probability less than 1% are considered), and do the same
- For a range of possible values for λ, display, to the decision-maker, the solution x_λ^* and the histogram of the random costs

$$s \mapsto \max_{i=1,\dots,m} \langle c_i^s \mid x_\lambda^* \rangle + b_i^s$$

- The decision-maker should choose his confidence level λ

We can also minimize the mean costs, while controlling for large costs

- Instead of only minimizing the mathematical expectation

$$\mathbb{E}[\mathbf{C}] \quad (= \sum_{s \in \mathcal{S}} \pi^s \mathbf{C}^s)$$

- we add the constraint that the Tail Value at Risk (at level $\lambda \in [0, 1[$) is not too large

$$TVaR_\lambda[\mathbf{C}] = \inf_{r \in \mathbb{R}} \left\{ \frac{\mathbb{E}[(\mathbf{C} - r)_+]}{1 - \lambda} + r \right\} \leq C^\sharp$$

- We can also choose to minimize a mixture

$$\theta \mathbb{E}[\mathbf{C}] + (1 - \theta) TVaR_\lambda[\mathbf{C}] = \inf_{r \in \mathbb{R}} \left\{ \theta \mathbb{E}[\mathbf{C}] + (1 - \theta) \frac{\mathbb{E}[(\mathbf{C} - r)_+]}{1 - \lambda} + (1 - \theta) r \right\}$$

Minimizing a mixture: convex piecewise linear programming formulation

- The risk-averse stochastic convex piecewise linear program

$$\min_{x\in\mathbb{R}^n}\min_{r\in\mathbb{R}}\left\{\theta\sum_{s\in\mathcal{S}}\pi^s\max_{i=1,\ldots,m}\langle c_i^s \mid x\rangle + b_i^s \right.$$
$$\left. +(1-\theta)r+\frac{1-\theta}{1-\lambda}\sum_{s\in\mathcal{S}}\pi^s\Big(\max_{i=1,\ldots,m}\langle c_i^s \mid x\rangle + b_i^s - r\Big)_+\right\}$$

- can be written as the convex piecewise linear program

$$\min_{x\in\mathbb{R}^n}\min_{r\in\mathbb{R}}\min_{(u^s)_{s\in\mathcal{S}}\in\mathbb{R}^{\mathcal{S}}}\sum_{s\in\mathcal{S}}\pi^s\left\{\theta u^s+(1-\theta)r+\frac{1-\theta}{1-\lambda}(u^s-r)_+\right\}$$
$$u^s\geq\langle c_1^s \mid x\rangle + b_1^s\,,\quad\forall s\in\mathcal{S}$$
$$\vdots$$
$$u^s\geq\langle c_m^s \mid x\rangle + b_m^s\,,\quad\forall s\in\mathcal{S}$$

Minimizing a mixture: linear programming formulation

- The risk-averse stochastic convex piecewise linear program

$$\min_{x \in \mathbb{R}^n} \min_{r \in \mathbb{R}} \Big\{ \theta \sum_{s \in \mathcal{S}} \pi^s \max_{i=1,\ldots,m} \langle c_i^s \mid x \rangle + b_i^s \\ +(1-\theta) r + \frac{1-\theta}{1-\lambda} \sum_{s \in \mathcal{S}} \pi^s \Big(\max_{i=1,\ldots,m} \langle c_i^s \mid x \rangle + b_i^s - r \Big)_+ \Big\}$$

- can be written as the linear program

$$\min_{x \in \mathbb{R}^n} \min_{r \in \mathbb{R}} \min_{(u^s)_{s \in \mathcal{S}} \in \mathbb{R}^{\mathcal{S}}} \min_{(v^s)_{s \in \mathcal{S}} \in \mathbb{R}^{\mathcal{S}}} \sum_{s \in \mathcal{S}} \pi^s \Big\{ \theta u^s + (1-\theta) r + \frac{1-\theta}{1-\lambda} v^s \Big\}$$

$$\begin{aligned} & u^s \geq \langle c_1^s \mid x \rangle + b_1^s \ , \ \ \forall s \in \mathcal{S} \\ & \vdots \\ & u^s \geq \langle c_m^s \mid x \rangle + b_m^s \ , \ \ \forall s \in \mathcal{S} \\ & v^s \geq u^s - r \ , \ \ \forall s \in \mathcal{S} \\ & v^s \geq 0 \ , \ \ \forall s \in \mathcal{S} \end{aligned}$$

How to use risk-averse stochastic programming in practice?

- Denote by $x^*_{\lambda,\theta}$ the (supposed unique) solution
- As $1-\lambda$ measures the upper probability of risky events, let the decision-maker choose a confidence level λ — $\lambda = 0.99$ (only events with probability less than 1% are considered), $\lambda = 0.95$, $\lambda = 0.90$, for instance
- Start with $\theta = 0$ and display, to the decision-maker, the risk-neutral solution $x^*_{\lambda,0}$ (which does not depend on λ) and the probability distribution (histogram) of the random costs

$$s \mapsto \max_{i=1,\dots,m} \left\langle c_i^s \mid x^*_{\lambda,0} \right\rangle + b_i^s$$

- Increase θ from 0 to 1, and display, to the decision-maker, the solution $x^*_{\lambda,\theta}$ and the histogram of the random costs

$$s \mapsto \max_{i=1,\dots,m} \left\langle c_i^s \mid x^*_{\lambda,\theta} \right\rangle + b_i^s$$

- The decision-maker reveals his confidence level λ and his mixture $(\theta, 1-\theta)$ as he selects his prefered histogram

Outline of the presentation

Minimizing the Tail Value at Risk of costs: linear programming formulation

- The risk-averse stochastic linear program with recourse

$$\min_{x,\{y^s\}_{s\in\mathcal{S}}} \min_{r\in\mathbb{R}} \left\{ r + \frac{1}{1-\lambda} \sum_{s\in\mathcal{S}} \pi^s \Big(\langle c^s \mid x \rangle + \langle p^s \mid y^s \rangle \Big)_+ \right\}$$

- can be written as the linear program

$$\begin{aligned} \min_{x,\{y^s\}_{s\in\mathcal{S}}} \min_{r} \min_{(v^s)_{s\in\mathcal{S}}} \quad & r + \frac{1}{1-\lambda} \sum_{s\in\mathcal{S}} \pi^s v^s \\ v^s - \langle c^s \mid x \rangle - \langle p^s \mid y^s \rangle & \geq 0 \,, \quad \forall s \in \mathcal{S} \\ v^s & \geq 0 \,, \quad \forall s \in \mathcal{S} \\ y^s & \geq 0 \,, \quad \forall s \in \mathcal{S} \\ A^s x + b^s + y^s & \geq 0 \,, \quad \forall s \in \mathcal{S} \end{aligned}$$

Minimizing a mixture: linear programming formulation

- The risk-averse stochastic linear program with recourse

$$\min_{x,\{y^s\}_{s\in\mathcal{S}}} \min_{r\in\mathbb{R}} \Bigg\{ \theta \sum_{s\in\mathcal{S}} \pi^s \Big(\langle c^s \mid x \rangle + \langle p^s \mid y^s \rangle \Big) + (1-\theta) r + \frac{1-\theta}{1-\lambda} \sum_{s\in\mathcal{S}} \pi^s \Big(\langle c^s \mid x \rangle + \langle p^s \mid y^s \rangle \Big)_+ \Bigg\}$$

- can be written as the linear program

$$\min_{x,\{y^s\}_{s\in\mathcal{S}}} \min_{r} \min_{(u^s,v^s)_{s\in\mathcal{S}}} \sum_{s\in\mathcal{S}} \pi^s \left\{ \theta u^s + (1-\theta) r + \frac{1-\theta}{1-\lambda} v^s \right\}$$

$$\begin{aligned}
u^s - \langle c^s \mid x \rangle - \langle p^s \mid y^s \rangle &\geq 0 \,, \quad \forall s \in \mathcal{S} \\
v^s - u^s + r &\geq 0 \,, \quad \forall s \in \mathcal{S} \\
v^s &\geq 0 \,, \quad \forall s \in \mathcal{S} \\
y^s &\geq 0 \,, \quad \forall s \in \mathcal{S} \\
A^s x + b^s + y^s &\geq 0 \,, \quad \forall s \in \mathcal{S}
\end{aligned}$$

What land have we covered?

- We have introduced one and two-stage optimization problems under uncertainty
- Thanks to a general framework, using risk measures, stochastic and robust optimization appear as (important) special cases
- We have presented resolution methods by scenario decomposition for two-stage optimization problems
- Dealing with multi-stage optimization problems requires specific tools, as is the notion of state

“Self-promotion, nobody will do it for you” ;-)