Optimization under uncertainty

V. Leclère

November 4, 2026





Presentation Outline

1. Optimization under uncertainty
   - Why uncertainty changes the problem
   - Three approaches to uncertainty
   - Scoring an uncertain cost
   - Enforcing an uncertain constraint
   - Evaluating a decision

Colour coding

Best effort:

- in the text, violet for definitions and red for key results and assumptions;
- in the maths, y_1, $\boldsymbol{y}_2$ decisions, ξ, $\boldsymbol{\xi}$ the uncertainty, λ dual variables, and y_1 a quantity held fixed.

After this session, you should be able to

- show that optimizing scenario by scenario, or at the average, does not necessarily produce a good decision;
- write the robust, stochastic and distributionally robust versions of a problem with uncertain data;
- score a random cost, or a constraint, by its expectation, value at risk, average value at risk or worst case;
- evaluate a candidate decision out of sample, on scenarios the optimizer never saw.

Presentation Outline

1. Optimization under uncertainty
   - Why uncertainty changes the problem
   - Three approaches to uncertainty
   - Scoring an uncertain cost
   - Enforcing an uncertain constraint
   - Evaluating a decision

Why is there uncertainty?

An optimization model needs data (costs, capacities, yields...), some of which is uncertain when the decision is taken. It may be

- measured imperfectly: a reaction yield, a pipe's resistance, a demand elasticity;
- in the future: tomorrow's wind, next week's demand, next year's electricity price;
- missing or corrupted, and the decision cannot wait;
- ...

A standard optimization problem

$$\begin{aligned} \min_{y_1 \in \mathbb{R}^n} \quad & \ell(y_1) \\ \text{s.t.} \quad & g(y_1) \leq 0 \end{aligned}$$

where

- y_1 is the decision (the subscript: first decision of the model);
- ℓ is the cost, or objective;
- $g(y_1) \leq 0$ are the constraints.

♥ An optimization problem with uncertainty

Adding an uncertainty ξ:

$$\begin{aligned} \min_{y_1} \quad & \ell(y_1, \xi) \\ \text{s.t.} \quad & g(y_1, \xi) \leq 0 \end{aligned}$$

♥ An optimization problem with uncertainty

Adding an uncertainty ξ:

$$\begin{aligned} \min_{y_1} \quad & \ell(y_1, \xi) \\ \text{s.t.} \quad & g(y_1, \xi) \leq 0 \end{aligned}$$

- ξ is unknown, and exogenous: the decision does not influence it.
- The cost is not well defined: for a fixed y_1, $\ell(y_1, \xi)$ is a function of ξ, not a number.
- The constraints are not well defined: $g(y_1, \xi) \leq 0$ has an unknown element ξ.

Optimizing at the average

Keep a single scenario, the mean: the expected value problem

$$v^{\mathrm{EV}} := \min_{y_1} \ \ell(y_1, \mathbb{E}_{\mathbb{P}}[\boldsymbol{\xi}]) \ ,$$

whose solution y_1^{EV} has expected cost $v^{\mathrm{EEV}} := \mathbb{E}_{\mathbb{P}}[\ell(y_1^{\mathrm{EV}}, \boldsymbol{\xi})]$, not v^{EV}.

A cyclist rides home on a road with traffic both ways; the pavement is safe but slow. He has had a drink: wherever he aims, he drifts left or right with equal probability. What does the expected value problem recommend?

Optimizing at the average

Keep a single scenario, the mean: the expected value problem

$$v^{\mathrm{EV}} := \min_{y_1} \ \ell(y_1, \mathbb{E}_{\mathbb{P}}[\boldsymbol{\xi}]) \ ,$$

whose solution y_1^{EV} has expected cost $v^{\mathrm{EEV}} := \mathbb{E}_{\mathbb{P}}[\ell(y_1^{\mathrm{EV}}, \boldsymbol{\xi})]$, not v^{EV}.

A cyclist rides home on a road with traffic both ways; the pavement is safe but slow. He has had a drink: wherever he aims, he drifts left or right with equal probability. What does the expected value problem recommend?
Aim at the painted line in the middle, where nothing drives. He is never on it: v^{EV} is achieved with probability 0, and the pavement is the answer.

Can't we just look at different values of the parameters?

- Natural idea: fix the uncertain parameters at a few values, the scenarios, and optimize for each: scenario-by-scenario optimization.
- Not enough, and can be misleading.
- Example:
  - an investor may buy, or sell short, 1000€ of a stock for a flat fee of 50€, or stand aside;
  - the stock gains or loses 10%, with equal probability;
  - ➥ what does optimizing each scenario recommend?

Can't we just look at different values of the parameters?

- Natural idea: fix the uncertain parameters at a few values, the scenarios, and optimize for each: scenario-by-scenario optimization.
- Not enough, and can be misleading.
- Example:
  - an investor may buy, or sell short, 1000€ of a stock for a flat fee of 50€, or stand aside;
  - the stock gains or loses 10%, with equal probability;
  - ➥ what does optimizing each scenario recommend?

Can't we just look at different values of the parameters?

- Natural idea: fix the uncertain parameters at a few values, the scenarios, and optimize for each: scenario-by-scenario optimization.
- Not enough, and can be misleading.
- Example:
  - an investor may buy, or sell short, 1000€ of a stock for a flat fee of 50€, or stand aside;
  - the stock gains or loses 10%, with equal probability;
  - ➥ what does optimizing each scenario recommend?

Answer. Up: buy; down: sell short. Both scenarios promise a cost of -50, that is, a gain of 50, with opposite trades. In expectation either trade gains 0 and costs the fee, 50: stand aside, which no scenario recommends.

Common solution might not be relevant

A plant makes products A, B and C:

- shared capacity of 100 units, production cost 10 per unit;
- selling prices 11, 12 and 13;
- demand for A unlimited; B and C are substitutes, of total demand 100: ξ for B, $100 - \xi$ for C;
- production is planned before ξ is known; unsold units are lost.

1. For a given ξ, write the optimization problem;
2. What is, for each ξ, the optimal production plan?
3. Is there a plan that is optimal for every ξ?
4. If ξ is 0 or 100 with equal probability, what is the optimal plan? Comments?

Common solution might not be relevant

A plant makes products A, B and C:

- shared capacity of 100 units, production cost 10 per unit;
- selling prices 11, 12 and 13;
- demand for A unlimited; B and C are substitutes, of total demand 100: ξ for B, $100 - \xi$ for C;
- production is planned before ξ is known; unsold units are lost.

1. For a given ξ, write the optimization problem;
2. What is, for each ξ, the optimal production plan?
3. Is there a plan that is optimal for every ξ?
4. If ξ is 0 or 100 with equal probability, what is the optimal plan? Comments?

Common solution might not be relevant

A plant makes products A, B and C:

- shared capacity of 100 units, production cost 10 per unit;
- selling prices 11, 12 and 13;
- demand for A unlimited; B and C are substitutes, of total demand 100: ξ for B, $100 - \xi$ for C;
- production is planned before ξ is known; unsold units are lost.

1. For a given ξ, write the optimization problem;
2. What is, for each ξ, the optimal production plan?
3. Is there a plan that is optimal for every ξ?
4. If ξ is 0 or 100 with equal probability, what is the optimal plan? Comments?

Answer. For a given ξ: C up to $100 - \xi$, B up to ξ, no A, at a cost $\xi - 300$; no plan is optimal for two values of ξ. The expected cost is $-y_A + 4\,y_B + 3.5\,y_C$: make only A, at -100, which appears in no scenario optimum. The average of the scenario plans, 50 of B and 50 of C, costs 375 on average: it is also the expected value solution, $v^{\mathrm{EV}} = -250$ and $v^{\mathrm{EEV}} = 375$.

Flexibility is not valued

- A buyer needs one unit of a good. Buying today costs 100.
- Waiting means paying an unknown price $p \in [50, 200]$.
- For 10, a trader offers the option (not the obligation) to buy later at 100.

1. What is the best decision when p is known in advance?
2. What is it when p is 50 or 200 with equal probability?

Flexibility is not valued

- A buyer needs one unit of a good. Buying today costs 100.
- Waiting means paying an unknown price $p \in [50, 200]$.
- For 10, a trader offers the option (not the obligation) to buy later at 100.

1. What is the best decision when p is known in advance?
2. What is it when p is 50 or 200 with equal probability?

Answer.

1. Buy today if $p > 100$, wait otherwise: the option never helps.
2. Buying today costs 100, waiting 125 on average, and the option $10 + \frac{1}{2}(50) + \frac{1}{2}(100) = 85$: buy the option.

A scenario fixes the future, so it never values the right to change your mind.

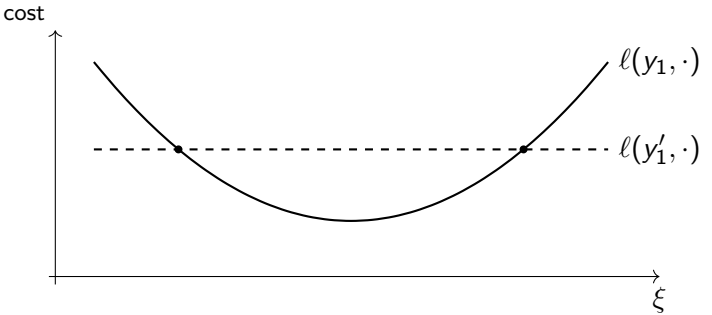
Neither the average scenario, nor the scenario solutions, nor their average recovers the optimal decision, and none values flexibility.

Presentation Outline

1. Optimization under uncertainty
   - Why uncertainty changes the problem
   - Three approaches to uncertainty
   - Scoring an uncertain cost
   - Enforcing an uncertain constraint
   - Evaluating a decision

Comparing two decisions

The cost $\ell(y_1, \xi)$ is a function of ξ, not a number: comparing two decisions means comparing two functions, which have no natural total order.



A model must collapse $\ell(y_1, \cdot)$ into one number, and say what $g(y_1, \xi) \leq 0$ requires.

Three approaches: a set, a law, a set of laws

- Robust Optimization (RO): $\xi \in \mathcal{R}$, a known uncertainty set, taken adversarially,

$$\min_{y_1} \sup_{\xi \in \mathcal{R}} \ell(y_1, \xi) \quad \text{s.t.} \quad g(y_1, \xi) \leq 0, \quad \forall \xi \in \mathcal{R} \, .$$

- Stochastic Programming (SP): $\xi \sim \mathbb{P}$, a random variable of known law,

$$\min_{y_1} \mathbb{E}_{\mathbb{P}}\big[\ell(y_1, \xi)\big] \quad \text{s.t.} \quad g(y_1, \xi) \leq 0, \quad \mathbb{P}\text{-a.s.}$$

- Distributionally Robust Optimization (DRO): the law is only assumed to lie in an ambiguity set $\mathcal{Q}$ of probability laws,

$$\min_{y_1} \sup_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}_{\mathbb{Q}}\big[\ell(y_1, \xi)\big] \, ,$$

the constraints holding $\mathbb{Q}$-a.s. for every $\mathbb{Q} \in \mathcal{Q}$.

Three approaches: a set, a law, a set of laws

- Robust Optimization (RO): $\xi \in \mathcal{R}$, a known uncertainty set, taken adversarially,

$$\min_{y_1} \sup_{\xi \in \mathcal{R}} \ell(y_1, \xi) \quad \text{s.t.} \quad g(y_1, \xi) \leq 0, \quad \forall \xi \in \mathcal{R} \, .$$

- Stochastic Programming (SP): $\boldsymbol{\xi} \sim \mathbb{P}$, a random variable of known law,

$$\min_{y_1} \mathbb{E}_{\mathbb{P}}\big[\ell(y_1, \boldsymbol{\xi})\big] \quad \text{s.t.} \quad g(y_1, \boldsymbol{\xi}) \leq 0, \quad \mathbb{P}\text{-a.s.}$$

- Distributionally Robust Optimization (DRO): the law is only assumed to lie in an ambiguity set $\mathcal{Q}$ of probability laws,

$$\min_{y_1} \sup_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}_{\mathbb{Q}}\big[\ell(y_1, \xi)\big] \, ,$$

the constraints holding $\mathbb{Q}$-a.s. for every $\mathbb{Q} \in \mathcal{Q}$.

Three approaches: a set, a law, a set of laws

- Robust Optimization (RO): $\xi \in \mathcal{R}$, a known uncertainty set, taken adversarially,

$$\min_{y_1} \sup_{\xi \in \mathcal{R}} \ell(y_1, \xi) \quad \text{s.t.} \quad g(y_1, \xi) \leq 0, \quad \forall \xi \in \mathcal{R} \, .$$

- Stochastic Programming (SP): $\boldsymbol{\xi} \sim \mathbb{P}$, a random variable of known law,

$$\min_{y_1} \mathbb{E}_{\mathbb{P}}\big[\ell(y_1, \boldsymbol{\xi})\big] \quad \text{s.t.} \quad g(y_1, \boldsymbol{\xi}) \leq 0, \quad \mathbb{P}\text{-a.s.}$$

- Distributionally Robust Optimization (DRO): the law is only assumed to lie in an ambiguity set $\mathcal{Q}$ of probability laws,

$$\min_{y_1} \sup_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}_{\mathbb{Q}}\big[\ell(y_1, \boldsymbol{\xi})\big] \, ,$$

  the constraints holding $\mathbb{Q}$-a.s. for every $\mathbb{Q} \in \mathcal{Q}$.

Three approaches: a set, a law, a set of laws

- Robust Optimization (RO): $\xi \in \mathcal{R}$, a known uncertainty set, taken adversarially,

$$\min_{y_1} \sup_{\xi \in \mathcal{R}} \ell(y_1, \xi) \quad \text{s.t.} \quad g(y_1, \xi) \leq 0, \quad \forall \xi \in \mathcal{R} \, .$$

- Stochastic Programming (SP): $\boldsymbol{\xi} \sim \mathbb{P}$, a random variable of known law,

$$\min_{y_1} \mathbb{E}_{\mathbb{P}}\big[\ell(y_1, \boldsymbol{\xi})\big] \quad \text{s.t.} \quad g(y_1, \boldsymbol{\xi}) \leq 0, \quad \mathbb{P}\text{-a.s.}$$

- Distributionally Robust Optimization (DRO): the law is only assumed to lie in an ambiguity set $\mathcal{Q}$ of probability laws,

$$\min_{y_1} \sup_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}_{\mathbb{Q}}\big[\ell(y_1, \boldsymbol{\xi})\big] \, ,$$

  the constraints holding $\mathbb{Q}$-a.s. for every $\mathbb{Q} \in \mathcal{Q}$.

DRO extends both: $\mathcal{Q} = \{\mathbb{P}\}$ gives SP, and $\mathcal{Q} = \{\delta_\xi \mid \xi \in \mathcal{R}\}$, the Dirac masses on $\mathcal{R}$, gives RO.

Requirements and limits

- Stochastic programming:
  - needs a law $\mathbb{P}$ of $\boldsymbol{\xi}$, rarely known, estimated from data;
  - can be hard to solve: the expectation, an integral, is replaced by finitely many values, which blows up the dimension;
  - has dedicated methods (e.g. Benders decomposition).
- Robust optimization:
  - needs an uncertainty set $\mathcal{R}$;
  - can be overly conservative: the size of $\mathcal{R}$ largely decides how much;
  - has a complexity that depends strongly on $\mathcal{R}$.
- Distributionally robust optimization:
  - mixes the two: a stochastic problem whose law is chosen robustly;
  - is a fast-growing field;
  - is still harder to implement;
  - typically takes $\mathcal{Q}$ a ball around the empirical law $\widehat{\mathbb{P}}_N$ of a sample, large enough to contain the true law $\mathbb{P}$ with high confidence.

Requirements and limits

- Stochastic programming:
  - needs a law $\mathbb{P}$ of $\boldsymbol{\xi}$, rarely known, estimated from data;
  - can be hard to solve: the expectation, an integral, is replaced by finitely many values, which blows up the dimension;
  - has dedicated methods (e.g. Benders decomposition).
- Robust optimization:
  - needs an uncertainty set $\mathcal{R}$;
  - can be overly conservative: the size of $\mathcal{R}$ largely decides how much;
  - has a complexity that depends strongly on $\mathcal{R}$.
- Distributionally robust optimization:
  - mixes the two: a stochastic problem whose law is chosen robustly;
  - is a fast-growing field;
  - is still harder to implement;
  - typically takes $\mathcal{Q}$ a ball around the empirical law $\widehat{\mathbb{P}}_N$ of a sample, large enough to contain the true law $\mathbb{P}$ with high confidence.

Requirements and limits

- Stochastic programming:
  - needs a law $\mathbb{P}$ of $\boldsymbol{\xi}$, rarely known, estimated from data;
  - can be hard to solve: the expectation, an integral, is replaced by finitely many values, which blows up the dimension;
  - has dedicated methods (e.g. Benders decomposition).
- Robust optimization:
  - needs an uncertainty set $\mathcal{R}$;
  - can be overly conservative: the size of $\mathcal{R}$ largely decides how much;
  - has a complexity that depends strongly on $\mathcal{R}$.
- Distributionally robust optimization:
  - mixes the two: a stochastic problem whose law is chosen robustly;
  - is a fast-growing field;
  - is still harder to implement;
  - typically takes $\mathcal{Q}$ a ball around the empirical law $\widehat{\mathbb{P}}_N$ of a sample, large enough to contain the true law $\mathbb{P}$ with high confidence.

The (deterministic) newsvendor problem

Each morning a newsboy buys papers at unit cost c and sells them during the day at unit price p, with $0 < c < p$; unsold papers are worthless. The day's demand is d papers.

How shall we model this?

The (deterministic) newsvendor problem

Each morning a newsboy buys papers at unit cost c and sells them during the day at unit price p, with $0 < c < p$; unsold papers are worthless. The day's demand is d papers.

How shall we model this?

- Decision $y_1 \geq 0$, the papers bought, of which $\min(y_1, d)$ are sold
- Cost $\ell(y_1) = c\, y_1 - p \min(y_1, d)$

$$\min_{y_1 \geq 0} \quad c\, y_1 - p \min(y_1, d)\ ,$$

solved by $y_1^\star = d$: the one number nobody has in the morning.

The (stochastic) newsvendor problem

The demand is unknown when buying: a random variable $\boldsymbol{d}$ of known law.

- The stock $y_1 \geq 0$ stays a number, fixed in the morning for every outcome;
- the cost becomes random; we score it by its expectation.

The (stochastic) newsvendor problem

The demand is unknown when buying: a random variable $\boldsymbol{d}$ of known law.

- The stock $y_1 \geq 0$ stays a number, fixed in the morning for every outcome;
- the cost becomes random; we score it by its expectation.

$$\min_{y_1 \geq 0} \quad \mathbb{E}_{\mathbb{P}}\big[c\, y_1 - p \min(y_1, \boldsymbol{d})\big]$$

Why the expectation?

The (stochastic) newsvendor problem

The demand is unknown when buying: a random variable $\boldsymbol{d}$ of known law.

- The stock $y_1 \geq 0$ stays a number, fixed in the morning for every outcome;
- the cost becomes random; we score it by its expectation.

$$\min_{y_1 \geq 0} \quad \mathbb{E}_{\mathbb{P}}\big[c\, y_1 - p \min(y_1, \boldsymbol{d})\big]$$

Why the expectation?
The law of large numbers: the newsvendor sells every day, and his average daily cost approaches the expected cost.

Solving the stochastic newsvendor problem

Proposition (Critical fractile)

Let $0 < c < p$ and $\boldsymbol{d} \geq 0$ have cdf F, and $F^{-1}(\alpha) := \inf\{t \mid F(t) \geq \alpha\}$. A stock $y_1^\star \geq 0$ is optimal if and only if

$$\mathbb{P}(\boldsymbol{d} < y_1^\star) \;\leq\; \frac{p-c}{p} \;\leq\; \mathbb{P}(\boldsymbol{d} \leq y_1^\star)\,,$$

and $F^{-1}\big((p-c)/p\big)$ is one. If F is continuous: $F(y_1^\star) = (p-c)/p$.

Solving the stochastic newsvendor problem

Proposition (Critical fractile)

Let $0 < c < p$ and $\boldsymbol{d} \geq 0$ have cdf F, and $F^{-1}(\alpha) := \inf\{t \mid F(t) \geq \alpha\}$. A stock $y_1^\star \geq 0$ is optimal if and only if

$$\mathbb{P}(\boldsymbol{d} < y_1^\star) \;\leq\; \frac{p-c}{p} \;\leq\; \mathbb{P}(\boldsymbol{d} \leq y_1^\star)\,,$$

and $F^{-1}\big((p-c)/p\big)$ is one. If F is continuous: $F(y_1^\star) = (p-c)/p$.

Proof. $J(y_1) := \mathbb{E}_{\mathbb{P}}\big[c\,y_1 - p\min(y_1, \boldsymbol{d})\big]$ is finite, since $\min(y_1, \boldsymbol{d}) \in [0, y_1]$, and convex; by dominated convergence

$$J'_+(y_1) = c - p\,\mathbb{P}\big(\boldsymbol{d} > y_1\big), \qquad J'_-(y_1) = c - p\,\mathbb{P}\big(\boldsymbol{d} \geq y_1\big).$$

Optimality is $J'_-(y_1^\star) \leq 0 \leq J'_+(y_1^\star)$, and

$$J'_-(y_1^\star) \leq 0 \iff \mathbb{P}\big(\boldsymbol{d} < y_1^\star\big) \leq \tfrac{p-c}{p}, \qquad J'_+(y_1^\star) \geq 0 \iff \mathbb{P}\big(\boldsymbol{d} \leq y_1^\star\big) \geq \tfrac{p-c}{p}$$

(at 0, only $J'_+ \geq 0$ is needed, and $\mathbb{P}(\boldsymbol{d} < 0) = 0$). $F^{-1}\big((p-c)/p\big)$ satisfies it by right continuity of F. □

The robust newsvendor problem

The demand d is unknown, only known to lie in $[\underline{d}, \overline{d}]$, and chosen adversarially.

The robust newsvendor problem

The demand d is unknown, only known to lie in $[\underline{d}, \overline{d}]$, and chosen adversarially.

$$\min_{y_1 \geq 0} \quad \max_{d \in [\underline{d}, \overline{d}]} c\, y_1 - p \min(y_1, d)$$

The robust newsvendor problem

The demand d is unknown, only known to lie in $[\underline{d}, \overline{d}]$, and chosen adversarially.

$$\min_{y_1 \geq 0} \quad \max_{d \in [\underline{d}, \overline{d}]} c\, y_1 - p \min(y_1, d)$$

The inner cost is nonincreasing in d, so the maximum is at $d = \underline{d}$ for every y_1:

$$\min_{y_1 \geq 0} \quad c\, y_1 - p \min(y_1, \underline{d}) \ ,$$

the deterministic problem at the lower bound: $y_1^\star = \underline{d}$.

Presentation Outline

1. Optimization under uncertainty
   - Why uncertainty changes the problem
   - Three approaches to uncertainty
   - Scoring an uncertain cost
   - Enforcing an uncertain constraint
   - Evaluating a decision

Alternative cost functions I

- For a random cost $\ell(y_1, \xi)$, it is natural to minimize $\mathbb{E}_{\mathbb{P}}[\ell(y_1, \xi)]$.
- Justified by the law of large numbers when the decision is repeated many times and only the total matters: the average cost paid approaches the expectation.
  - ➥ The newsvendor sells 250 days a year: is the expectation his score? And if he pays rent every week?

- But the expectation may not reflect your attitude to risk:
  - Pay 1€ for one chance in ten of winning 10€? And 1000€ for one chance in ten of winning 10 000€?
  - To catch a flight within the hour: one route takes 50 minutes for sure; another 40 minutes with probability 4/5, 70 minutes with probability 1/5. Which one?

Alternative cost functions I

- For a random cost $\ell(y_1, \xi)$, it is natural to minimize $\mathbb{E}_{\mathbb{P}}[\ell(y_1, \xi)]$.
- Justified by the law of large numbers when the decision is repeated many times and only the total matters: the average cost paid approaches the expectation.
  - ➥ The newsvendor sells 250 days a year: is the expectation his score? And if he pays rent every week? Living at home, yes: the year's cost is close to 250 times the expected daily cost. Owing rent on Friday, no: a bad week is not made up by later good ones, so score the tail.
- But the expectation may not reflect your attitude to risk:
  - Pay 1€ for one chance in ten of winning 10€? And 1000€ for one chance in ten of winning 10 000€?
  - To catch a flight within the hour: one route takes 50 minutes for sure; another 40 minutes with probability 4/5, 70 minutes with probability 1/5. Which one?

Alternative cost functions II

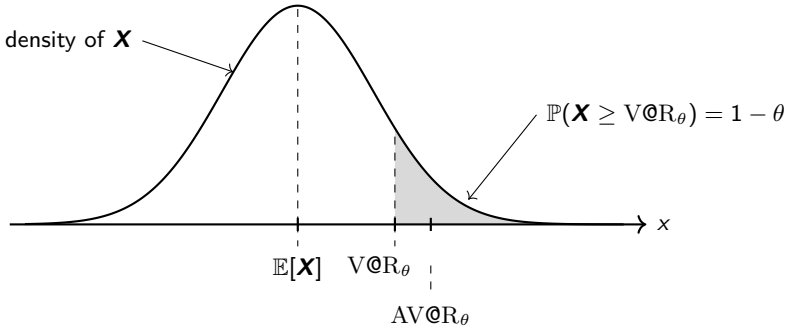
A risk measure ρ collapses the random cost into one number: $\min_{y_1} \rho(\ell(y_1, \boldsymbol{\xi}))$. Besides $\rho = \mathbb{E}_{\mathbb{P}}$:

- threshold probability, $\mathbb{P}(\ell(y_1, \boldsymbol{\xi}) > 0)$: how often, not by how much;
- Value-at-Risk at level $\theta \in [0, 1]$ of a random cost $\boldsymbol{X}$, $\mathrm{V@R}_\theta(\boldsymbol{X}) := \inf\{t \mid \mathbb{P}(\boldsymbol{X} \leq t) \geq \theta\}$, the smallest cost exceeded with probability at most $1 - \theta$;
- Average Value-at-Risk, the mean cost over the worst fraction $1 - \theta$ of outcomes; for $\theta < 1$ and $\boldsymbol{X}$ with a density and a finite expectation, $\mathrm{AV@R}_\theta(\boldsymbol{X}) := \mathbb{E}_{\mathbb{P}}[\boldsymbol{X} \mid \boldsymbol{X} > \mathrm{V@R}_\theta(\boldsymbol{X})]$.[1] From the expectation ($\theta = 0$) to the worst case ($\theta \to 1$).[2]
- Not recommended: the expectation plus a multiple of the variance (Markowitz), which is not monotone: a cost that is never positive can score above a cost that is always zero.

[1] For $0 < \theta < 1$ and a cost of finite expectation, of any law, scenarios included: $\mathrm{AV@R}_\theta(\boldsymbol{X}) = \min_{\eta \in \mathbb{R}} \left\{\eta + \frac{1}{1-\theta} \mathbb{E}_{\mathbb{P}}\left[\left(\boldsymbol{X} - \eta\right)^+\right]\right\}$, attained at $\eta = \mathrm{V@R}_\theta(\boldsymbol{X})$ (Rockafellar and Uryasev).

[2] Some authors use the complementary level: $\mathrm{AV@R}_\alpha$ averages the worst fraction α of the outcomes, $\alpha = 1 - \theta$.

Value at risk and average value at risk



A Gaussian cost, $\theta = 0.9$. The value at risk is exceeded on the shaded tail, of probability $1 - \theta$; the average value at risk is the tail's mean, hence to its right.

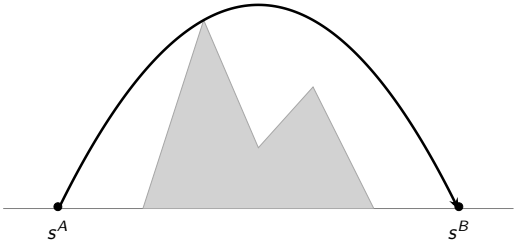
Presentation Outline

1. Optimization under uncertainty
   - Why uncertainty changes the problem
   - Three approaches to uncertainty
   - Scoring an uncertain cost
   - Enforcing an uncertain constraint
   - Evaluating a decision

A drone in a corridor

A drone flies from s^A at date 0 to s^B at date T; its position $s_t \in \mathbb{R}^2$ must stay in the authorized airspace $\mathcal{X}$, closed and not the whole plane. The decision is the displacement y_t over period t, of norm at most Δ, and energy grows with its square:

$$\begin{aligned} \min_{(y_t)_{t\in[T]}} \quad & \sum_{t=1}^{T} \|y_t\|^2 \\ \text{s.t.} \quad & s_t = s_{t-1} + y_t, && \forall t \in [T] \\ & s_t \in \mathcal{X}, && \forall t \in [T] \\ & \|y_t\| \leq \Delta, && \forall t \in [T] \\ & s_0 = s^A, \quad s_T = s^B\ . \end{aligned}$$



Alternative constraints

- Extending $g(y_1, \xi) \leq 0$ to $g(y_1, \boldsymbol{\xi}) \leq 0$ $\mathbb{P}$-a.s. can be extremely conservative, often infeasible.
- Example. Imprecise propulsion: over period t the drone covers $y_t + \xi_t$, with $\xi_t \sim \mathcal{N}(0, \sigma^2 I_2)$ independent over t, and the commands are fixed in advance. Required almost surely, the problem is infeasible, twice over:
  - s_T has a law with a density, so $\mathbb{P}(s_T = s^B) = 0$;
  - even with a landing zone: a Gaussian charges the whole plane and the complement of $\mathcal{X}$ is open and nonempty, so every plan leaves $\mathcal{X}$ with positive probability, at every period.
- The deviation from the plan at period t has standard deviation $\sigma\sqrt{t}$: a corridor of $10\sigma\sqrt{T}$ on either side of the planned path is left with probability below 10^{-21} at every period. Incredibly small, but not zero.

Alternative constraints

- Extending $g(y_1, \xi) \leq 0$ to $g(y_1, \boldsymbol{\xi}) \leq 0$ $\mathbb{P}$-a.s. can be extremely conservative, often infeasible.
- Example. Imprecise propulsion: over period t the drone covers $y_t + \boldsymbol{\xi}_t$, with $\boldsymbol{\xi}_t \sim \mathcal{N}(0, \sigma^2 I_2)$ independent over t, and the commands are fixed in advance. Required almost surely, the problem is infeasible, twice over:
  - s_T has a law with a density, so $\mathbb{P}(s_T = s^B) = 0$;
  - even with a landing zone: a Gaussian charges the whole plane and the complement of $\mathcal{X}$ is open and nonempty, so every plan leaves $\mathcal{X}$ with positive probability, at every period.
- The deviation from the plan at period t has standard deviation $\sigma\sqrt{t}$: a corridor of $10\sigma\sqrt{T}$ on either side of the planned path is left with probability below 10^{-21} at every period. Incredibly small, but not zero.

Alternative constraints II

Five constraints from one requirement, each of the form $\rho(g(y_1, \xi)) \leq 0$:

- $\mathbb{E}_{\mathbb{P}}[g(y_1, \xi)] \leq 0$, in expectation: quality-of-service constraints, where a violation in one outcome is offset by slack in another;
- $\mathbb{P}(g(y_1, \xi) \leq 0) \geq \theta$, a chance constraint, conservative for θ close to 1: easy to state, but silent on what happens when it is violated;
- $\mathrm{AV@R}_\theta(g(y_1, \xi)) \leq 0$, by a risk measure;
- $g(y_1, \xi) \leq 0$, $\mathbb{P}$-a.s., almost surely: right when the support is bounded and the requirement physical;
- $g(y_1, \xi) \leq 0$, $\forall \xi \in \mathcal{R}$, robustly: the almost-sure constraint with the support replaced by a set you choose (for $g(y_1, \cdot)$ continuous, the two coincide when $\mathcal{R} = \mathrm{supp}(\xi)$).

- The chance constraint is $\mathrm{V@R}_\theta(g(y_1, \xi)) \leq 0$, whatever the law.
- The $\mathrm{AV@R}_\theta$ constraint implies it, since $\mathrm{AV@R}_\theta \geq \mathrm{V@R}_\theta$, and is convex whenever $g(\cdot, \xi)$ is.

Alternative constraints II

Five constraints from one requirement, each of the form $\rho(g(y_1, \xi)) \leq 0$:

- $\mathbb{E}_{\mathbb{P}}[g(y_1, \xi)] \leq 0$, in expectation: quality-of-service constraints, where a violation in one outcome is offset by slack in another;
- $\mathbb{P}(g(y_1, \xi) \leq 0) \geq \theta$, a chance constraint, conservative for θ close to 1: easy to state, but silent on what happens when it is violated;
- $\mathrm{AV@R}_\theta(g(y_1, \xi)) \leq 0$, by a risk measure;
- $g(y_1, \xi) \leq 0$, $\mathbb{P}$-a.s., almost surely: right when the support is bounded and the requirement physical;
- $g(y_1, \xi) \leq 0$, $\forall \xi \in \mathcal{R}$, robustly: the almost-sure constraint with the support replaced by a set you choose (for $g(y_1, \cdot)$ continuous, the two coincide when $\mathcal{R} = \mathrm{supp}(\xi)$).

- The chance constraint is $\mathrm{V@R}_\theta(g(y_1, \xi)) \leq 0$, whatever the law.
- The $\mathrm{AV@R}_\theta$ constraint implies it, since $\mathrm{AV@R}_\theta \geq \mathrm{V@R}_\theta$, and is convex whenever $g(\cdot, \xi)$ is.

Presentation Outline

1. Optimization under uncertainty
   - Why uncertainty changes the problem
   - Three approaches to uncertainty
   - Scoring an uncertain cost
   - Enforcing an uncertain constraint
   - Evaluating a decision

Computing expectation

- Computing $\mathbb{E}_{\mathbb{P}}\left[\ell(y_1, \boldsymbol{\xi})\right]$, even for a given y_1, is costly:
  - with a density, a multidimensional integral;
  - discrete, a sum over all realizations, possibly huge;
  - given by code returning samples, no exact value at all.

Computing expectation

- Computing $\mathbb{E}_{\mathbb{P}}\big[\ell(y_1, \boldsymbol{\xi})\big]$, even for a given y_1, is costly:
  - with a density, a multidimensional integral;
  - discrete, a sum over all realizations, possibly huge;
  - given by code returning samples, no exact value at all.

Solution: sampling, justified by the law of large numbers (LLN) and the central limit theorem (CLT).

- Draw N independent realizations $(\xi^n)_{n\in[N]}$ of $\boldsymbol{\xi}$;
- compute the sample average $\dfrac{1}{N}\sum_{n=1}^{N} \ell(y_1, \xi^n)$;
- the CLT gives an asymptotic confidence interval: at 95%, the sample average $\pm\, 1.96\, \hat{\sigma}/\sqrt{N}$, with $\hat{\sigma}$ the empirical standard deviation.

Computing expectation

- Computing $\mathbb{E}_{\mathbb{P}}\left[\ell(y_1, \boldsymbol{\xi})\right]$, even for a given y_1, is costly:
  - with a density, a multidimensional integral;
  - discrete, a sum over all realizations, possibly huge;
  - given by code returning samples, no exact value at all.

Solution: sampling, justified by the law of large numbers (LLN) and the central limit theorem (CLT).

- Draw N independent realizations $(\xi^n)_{n\in[N]}$ of $\boldsymbol{\xi}$;
- compute the sample average $\frac{1}{N}\sum_{n=1}^{N} \ell(y_1, \xi^n)$;
- the CLT gives an asymptotic confidence interval: at 95%, the sample average $\pm 1.96\,\hat{\sigma}/\sqrt{N}$, with $\hat{\sigma}$ the empirical standard deviation.

This is the Monte Carlo method (details with the sample average approximation).

Consequence: evaluating a decision is difficult

- Even evaluating a solution can require approximation.
- So can checking admissibility: a chance constraint is the expectation of an indicator, an almost-sure constraint a statement about every outcome.
- First-order information (gradients) is harder still.

Standard remedy: sample, and solve the sampled problem (Sample Average Approximation).

Consequence: evaluating a decision is difficult

- Even evaluating a solution can require approximation.
- So can checking admissibility: a chance constraint is the expectation of an indicator, an almost-sure constraint a statement about every outcome.
- First-order information (gradients) is harder still.

Standard remedy: sample, and solve the sampled problem (Sample Average Approximation).

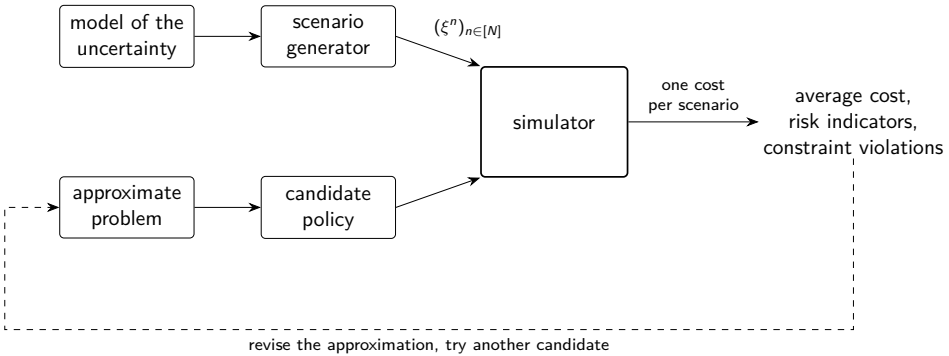
♥ Optimization problem and simulator

- Stochastic optimization problems are often not well posed, and must be approximated before being solved.
- Good practice: write a simulator, code that takes realizations of the uncertainty and a policy (a rule turning the available information into a decision), and returns the costs and whatever else is worth watching.

Three steps:

1. define the simulator first, before solving any model: it defines what a good decision is;
2. obtain a candidate policy by solving an approximate problem;
3. evaluate it on the simulator, on more than the average cost (spread, constraint violations...), and if it is not good enough, go back to step 2 with a different approximation.

Evaluating a candidate on the simulator



The scenario generator produces realizations, the approximate problem a policy; the simulator runs one against the other and returns one cost per scenario, from which indicators are estimated. The dashed arrow goes back to the approximate problem, not to the uncertainty model.

♥ Judge a candidate out of sample

- The sample that produced a candidate cannot judge it: a decision minimizing the average cost over $(\xi^n)_{n\in[N]}$ is fitted to those draws, and its average on them is optimistic on average.
- Out-of-sample: evaluate on a fresh sample, never shown to the optimizer; an unbiased estimate of its expected cost under the sampled law.
- Backtest: evaluate on held-out data, such as historical records; this tests the law itself, misspecified or shifted since the fit.
- Compare two methods by running both policies on one common out-of-sample set, never by the optimal values they report.

Example. Newsvendor, $c = 1$, $p = 2$, demand uniform on $[0, 100]$: the optimal stock is 50, of expected cost -25. Fitted to three days of demand 30, 60, 80, the stock is 60: it costs -40 on those days, and -24 in expectation.

♥ Judge a candidate out of sample

- The sample that produced a candidate cannot judge it: a decision minimizing the average cost over $(\xi^n)_{n\in[N]}$ is fitted to those draws, and its average on them is optimistic on average.
- Out-of-sample: evaluate on a fresh sample, never shown to the optimizer; an unbiased estimate of its expected cost under the sampled law.
- Backtest: evaluate on held-out data, such as historical records; this tests the law itself, misspecified or shifted since the fit.
- Compare two methods by running both policies on one common out-of-sample set, never by the optimal values they report.

Example. Newsvendor, $c = 1$, $p = 2$, demand uniform on $[0, 100]$: the optimal stock is 50, of expected cost -25. Fitted to three days of demand 30, 60, 80, the stock is 60: it costs -40 on those days, and -24 in expectation.

Conclusion

Facing an optimization problem under uncertainty, decide

- how to model the uncertainty (a set, a random variable, a set of laws);
- how to score a random cost, i.e. your risk preference (expectation, value at risk. . .);
- what a constraint requires when it cannot hold in every outcome;
- what information is available for each decision: the information structure (later);
- and set up a simulator to evaluate your solutions.

Course plan

1. Stochastic programming formulation
2. Robust optimization: introduction, approaches and examples
3. Decomposition methods for two-stage stochastic programming
4. Methods for multistage stochastic programming
5. Advanced methods for robust optimization
6. Exam

What you must remember

- Before solving, decide how a random cost is scored, what a constraint requires when it cannot hold in every outcome, and what information is available for each decision.
- Do not replace the uncertainty by a single forecast, nor choose among scenario solutions: a deterministic problem does not value flexibility.
- Score by the expectation only if the decision is repeated many times and only the total matters.
- Beware of requiring a constraint almost surely.
- Build the simulator first, and judge every candidate on scenarios the optimizer never saw.

What you should remember

- Three approaches to uncertainty: a set, scored by its worst case; a law, scored by its expectation; a set of laws, scored by its worst expectation.
- The value at risk considers the number of bad outcomes, not how bad they are; the average value at risk weighs them, running from the expectation to the worst case.
- A chance constraint is a value-at-risk constraint, and the average-value-at-risk constraint at the same level implies it.

Questions?



You can contact me at: vincent.leclere@enpc.fr