Optimization under uncertainty: stochastic programming

V. Leclère

November, 17st 2023

Presentation Outline

Presentation Outline

Why is there uncertainty?

- Parameters might not be known exactly (there might be some error in the measurement);
- We want to take into account event that have yet to come (*e.g.*, weather), such event could be predicted but not with certainty;
- Data might be missing or corrupted, and we have to take a decision based on available data nonetheless;
- ...

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

- A natural idea, when confronted with uncertainty, is to just look at various possible values of the uncertain parameters and optimize for each of them (sometimes called "scenario optimization").
- Unfortunately, this is not enough and can even be quite misleading.
- Let's take a first example:
  - You have 1000€ to trade on the stock market;
  - Consider a stock that has a 50% chance of going up by 10% and a 50% chance of going down by 10%;
  - Assume that, for a flat fee of 50€, you can buy or sell the stock;
  - ➥ What happens if you just look at the two values and optimize for each?

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

- A natural idea, when confronted with uncertainty, is to just look at various possible values of the uncertain parameters and optimize for each of them (sometimes called "scenario optimization").
- Unfortunately, this is not enough and can even be quite misleading.
- Let's take a first example:
  - You have 1000€ to trade on the stock market;
  - Consider a stock that has a 50% chance of going up by 10% and a 50% chance of going down by 10%;
  - Assume that, for a flat fee of 50€, you can buy or sell the stock;
  - ➥ What happens if you just look at the two values and optimize for each?

Common solution might not be relevant

Assume that you can produce 3 products A, B and C.

- Assume total units produced is limited to 100, with a production cost of 10 per unit;
- Selling price of A is 11 per unit, B is 12 per unit and C is 13 per unit;
- The demand for A is unlimited, for B and C the total demand is 100, but the repartition is unknown – let denote ξ the demand for B.

1. For a given ξ, write the optimization problem;
2. What is, for each ξ, the optimal production plan?
3. Is there a decision common to all ξ?
4. Assume now that ξ take value 0 or 100 with equal probability, what is the optimal production plan? Comments?

Common solution might not be relevant

Assume that you can produce 3 products A, B and C.

- Assume total units produced is limited to 100, with a production cost of 10 per unit;
- Selling price of A is 11 per unit, B is 12 per unit and C is 13 per unit;
- The demand for A is unlimited, for B and C the total demand is 100, but the repartition is unknown – let denote ξ the demand for B.

1. For a given ξ, write the optimization problem;
2. What is, for each ξ, the optimal production plan?
3. Is there a decision common to all ξ?
4. Assume now that ξ take value 0 or 100 with equal probability, what is the optimal production plan? Comments?

Flexibility is not valued

- Assume that you need to buy a certain amount of a given product. The current price is 100.
- You can buy now, or wait and buy at an unknown price $p \in [50, 200]$.
- A trader offers you, for 10, to have the option to buy, later, at the current price 100.

1. What is the optimal decision if you know the value of p?
2. Assume that p can only take value 50 and 200 with equal probability, what is the optimal decision?

Presentation Outline

A standard optimization problem

$$\begin{aligned} \min_{u_0} \quad & L(u_0) \\ s.t. \quad & g(u_0) \leq 0 \end{aligned}$$

where

- u_0 is the control, or decision.
- L is the cost or objective function.
- $g(u_0) \leq 0$ represent the constraint(s).

The (deterministic) newsboy problem

In the 50's a boy would buy a stock u of newspapers each morning at a cost c, and sell them all day long for a price p. The number of people interested in buying a paper during the day is d. We assume that $0 < c < p$.

How shall we model this ?

The (deterministic) newsboy problem

In the 50's a boy would buy a stock u of newspapers each morning at a cost c, and sell them all day long for a price p. The number of people interested in buying a paper during the day is d. We assume that $0 < c < p$.

How shall we model this ?

- Control $u \in \mathbb{R}^+$
- Cost $L(u) = cu - p\min(u, d)$

Leading to

$$\begin{aligned} \min_u \quad & cu - p\min(u, d) \\ s.t. \quad & u \geq 0 \end{aligned}$$

An optimization problem with uncertainty

Adding uncertainty ξ in the mix

$$\min_{u} \quad L(u,\xi)$$
$$s.t. \quad g(u,\xi) \leq 0$$

An optimization problem with uncertainty

Adding uncertainty ξ in the mix

$$\begin{aligned} \min_u \quad & L(u,\xi) \\ s.t. \quad & g(u,\xi) \leq 0 \end{aligned}$$

Remarks:

- ξ is unknown. Two main ways of modeling it:
  - $\xi \in R$ with a known uncertainty set R, and a pessimistic approach. This is the Robust Optimization approach (RO).
  - $\boldsymbol{\xi} \sim \mathbb{P}$ is a random variable with known probability law. This is the Stochastic Programming approach (SP).

An optimization problem with uncertainty

Adding uncertainty ξ in the mix

$$\begin{aligned} \min_u \quad & L(u,\xi) \\ s.t. \quad & g(u,\xi) \leq 0 \end{aligned}$$

Remarks:

- ξ is unknown. Two main ways of modeling it:
  - $\xi \in R$ with a known uncertainty set R, and a pessimistic approach. This is the Robust Optimization approach (RO).
  - $\boldsymbol{\xi} \sim \mathbb{P}$ is a random variable with known probability law. This is the Stochastic Programming approach (SP).
- Cost is not well defined.
  - RO : $\max_{\xi \in R} L(u,\xi)$.
  - SP : $\mathbb{E}\big[L(u,\boldsymbol{\xi})\big]$.

An optimization problem with uncertainty

Adding uncertainty ξ in the mix

$$\min_u \quad L(u,\xi)$$
$$s.t. \quad g(u,\xi) \leq 0$$

Remarks:

- ξ is unknown. Two main ways of modeling it:
  - $\xi \in R$ with a known uncertainty set R, and a pessimistic approach. This is the Robust Optimization approach (RO).
  - $\boldsymbol{\xi} \sim \mathbb{P}$ is a random variable with known probability law. This is the Stochastic Programming approach (SP).
- Cost is not well defined.
  - RO : $\max_{\xi \in R} L(u,\xi)$.
  - SP : $\mathbb{E}\big[L(u,\boldsymbol{\xi})\big]$.
- Constraints are not well defined.
  - RO : $g(u,\xi) \leq 0, \qquad \forall \xi \in R$.
  - SP : $g(u,\boldsymbol{\xi}) \leq 0, \qquad \mathbb{P}-a.s.$

Requirements and limits

- Stochastic optimization :
  - requires a law $\mathbb{P}$ of the uncertainty ξ
  - can be hard to solve (generally require discretizing the support and blowing up the dimension of the problem)
  - there exists specific methods (like Bender's decomposition)
- Robust optimization :
  - requires an uncertainty set R
  - can be overly conservative, even for reasonable R
  - complexity strongly depend on the choice of R
- Distributionally robust optimization :
  - is a mix between robust and stochastic optimization
  - consists in solving a stochastic optimization problem where the law is chosen in a robust way
  - is a fast growing fields with multiple recent results
  - but is still harder to implement than other approaches

Requirements and limits

- Stochastic optimization :
  - requires a law $\mathbb{P}$ of the uncertainty ξ
  - can be hard to solve (generally require discretizing the support and blowing up the dimension of the problem)
  - there exists specific methods (like Bender's decomposition)
- Robust optimization :
  - requires an uncertainty set R
  - can be overly conservative, even for reasonable R
  - complexity strongly depend on the choice of R
- Distributionally robust optimization :
  - is a mix between robust and stochastic optimization
  - consists in solving a stochastic optimization problem where the law is chosen in a robust way
  - is a fast growing fields with multiple recent results
  - but is still harder to implement than other approaches

Requirements and limits

- Stochastic optimization :
  - requires a law $\mathbb{P}$ of the uncertainty ξ
  - can be hard to solve (generally require discretizing the support and blowing up the dimension of the problem)
  - there exists specific methods (like Bender's decomposition)
- Robust optimization :
  - requires an uncertainty set R
  - can be overly conservative, even for reasonable R
  - complexity strongly depend on the choice of R
- Distributionally robust optimization :
  - is a mix between robust and stochastic optimization
  - consists in solving a stochastic optimization problem where the law is chosen in a robust way
  - is a fast growing fields with multiple recent results
  - but is still harder to implement than other approaches

Distributionally Robust Optimization

- SO: $\min_u \quad \mathbb{E}_{\mathbb{P}}\big[L(u,\boldsymbol{\xi})\big]$
- RO: $\min_u \quad \max_{\xi \in R} L(u,\xi)$
- DRO: $\min_u \quad \max_{\mathbb{Q} \in \mathcal{Q}} \quad \mathbb{E}_{\mathbb{Q}}\big[L(u,\boldsymbol{\xi})\big]$

DRO bridge stochastic and robust optimization:

- If $\mathcal{Q} = \{\mathbb{P}\}$, then DRO reduce to SO;
- if $\{\delta_\xi \mid \xi \in R\} \subset \mathcal{Q}$, then DRO is equivalent to RO.

DRO is a recent and very active field. One of the main idea is to choose $\mathcal{Q}$ as a ball, in some sense, around an empirical probability measure $\hat{\mathbb{P}}_N$, such that the true, unknown, probability $\mathbb{P}$ is contained in $\mathcal{Q}$ with high confidence.

➥ We won't discuss it further in this course.

Distributionally Robust Optimization

- SO: $\min_{u} \quad \mathbb{E}_{\mathbb{P}}\big[L(u,\xi)\big]$
- RO: $\min_{u} \quad \max_{\xi \in R} L(u,\xi)$
- DRO: $\min_{u} \quad \max_{\mathbb{Q} \in \mathcal{Q}} \quad \mathbb{E}_{\mathbb{Q}}\big[L(u,\xi)\big]$

DRO bridge stochastic and robust optimization:

- If $\mathcal{Q} = \{\mathbb{P}\}$, then DRO reduce to SO;
- if $\{\delta_\xi \mid \xi \in R\} \subset \mathcal{Q}$, then DRO is equivalent to RO.

DRO is a recent and very active field. One of the main idea is to choose $\mathcal{Q}$ as a ball, in some sense, around an empirical probability measure $\hat{\mathbb{P}}_N$, such that the true, unknown, probability $\mathbb{P}$ is contained in $\mathcal{Q}$ with high confidence.

➥ We won't discuss it further in this course.

Distributionally Robust Optimization

- SO: $\min_{u} \quad \mathbb{E}_{\mathbb{P}}\big[L(u, \boldsymbol{\xi})\big]$
- RO: $\min_{u} \quad \max_{\xi \in R} L(u, \xi)$
- DRO: $\min_{u} \quad \max_{\mathbb{Q} \in \mathcal{Q}} \quad \mathbb{E}_{\mathbb{Q}}\big[L(u, \boldsymbol{\xi})\big]$

DRO bridge stochastic and robust optimization:

- If $\mathcal{Q} = \{\mathbb{P}\}$, then DRO reduce to SO;
- if $\{\delta_\xi \mid \xi \in R\} \subset \mathcal{Q}$, then DRO is equivalent to RO.

DRO is a recent and very active field. One of the main idea is to choose $\mathcal{Q}$ as a ball, in some sense, around an empirical probability measure $\hat{\mathbb{P}}_N$, such that the true, unknown, probability $\mathbb{P}$ is contained in $\mathcal{Q}$ with high confidence.

➥ We won't discuss it further in this course.

The (stochastic) newsboy problem

Demand d is unknown at time of purchasing. We model it as a random variable $\boldsymbol{d}$ with known law. Note that

- the control $u \in \mathbb{R}^+$ is deterministic
- the cost is a random variable (depending of $\boldsymbol{d}$). We choose to minimize its expectation.

The (stochastic) newsboy problem

Demand d is unknown at time of purchasing. We model it as a random variable $\boldsymbol{d}$ with known law. Note that

- the control $u \in \mathbb{R}^+$ is deterministic
- the cost is a random variable (depending of $\boldsymbol{d}$). We choose to minimize its expectation.

We consider the following problem

$$\begin{aligned} \min_{u} \quad & \mathbb{E}\big[cu - p\min(u, \boldsymbol{d})\big] \\ s.t. \quad & u \geq 0 \end{aligned}$$

How can we justify the expectation ?

The (stochastic) newsboy problem

Demand d is unknown at time of purchasing. We model it as a random variable $\boldsymbol{d}$ with known law. Note that

- the control $u \in \mathbb{R}^+$ is deterministic
- the cost is a random variable (depending of $\boldsymbol{d}$). We choose to minimize its expectation.

We consider the following problem

$$\begin{aligned} \min_u \quad & \mathbb{E}\big[cu - p\min(u, \boldsymbol{d})\big] \\ s.t. \quad & u \geq 0 \end{aligned}$$

How can we justify the expectation ?
By law of large number: the Newsboy is going to sell newspaper again and again. Then optimizing the sum over time of its gains is closely related to optimizing the expected gains.

Solving the stochastic newsboy problem

For simplicity assume that the demand $\boldsymbol{d}$ has a continuous density f. Define $J(u)$ the expected "loss" of the newsboy if he bought u newspaper. We have

$$\begin{aligned} J(u) &= \mathbb{E}\Big[cu - p\min(u, \boldsymbol{d})\Big] \\ &= (c-p)u - p\mathbb{E}\Big[\min(0, \boldsymbol{d} - u)\Big] \\ &= (c-p)u - p\int_{-\infty}^{u} (x-u)f(x)dx \\ &= (c-p)u - p\Big(\int_{-\infty}^{u} xf(x)dx - u\int_{-\infty}^{u} f(x)dx\Big) \end{aligned}$$

Solving the stochastic newsboy problem

For simplicity assume that the demand $\boldsymbol{d}$ has a continuous density f. Define $J(u)$ the expected "loss" of the newsboy if he bought u newspaper. We have

$$J(u) = \mathbb{E}\Big[cu - p\min(u, \boldsymbol{d})\Big]$$

$$= (c-p)u - p\Big(\int_{-\infty}^{u} xf(x)dx - u\int_{-\infty}^{u} f(x)dx\Big)$$

Thus,

$$J'(u) = (c-p) - p\Big(uf(u) - \int_{-\infty}^{u} f(x)dx - uf(u)\Big) = c - p + pF(u)$$

where F is the cumulative distribution function (cdf) of $\boldsymbol{d}$. F being non decreasing, the optimum

control u^* is such that $J'(u^*) = 0$, which is

$$u^* \in F^{-1}\Big(\frac{p-c}{p}\Big)$$

The robust newsboy problem

Demand d is unknown at time of purchasing. We assume that it will be in the set $[\underline{d}, \overline{d}]$.

The robust newsboy problem

Demand d is unknown at time of purchasing. We assume that it will be in the set $[\underline{d}, \overline{d}]$.

The robust problem consist in solving

$$\min_{u} \quad \max_{d\in[\underline{d},\overline{d}]} cu - p\min(u, d)$$
$$s.t. \quad u \geq 0$$

The robust newsboy problem

Demand d is unknown at time of purchasing. We assume that it will be in the set $[\underline{d}, \overline{d}]$.
The robust problem consist in solving

$$\begin{aligned} \min_{u} \quad & \max_{d \in [\underline{d}, \overline{d}]} \; cu - p\min(u, d) \\ s.t. \quad & u \geq 0 \end{aligned}$$

By monotonicity it is equivalent to

$$\begin{aligned} \min_{u} \quad & cu - p\min(u, \underline{d}) \\ s.t. \quad & u \geq 0 \end{aligned}$$

Alternative cost functions I

- When the cost $L(u,\xi)$ is random it might be natural to want to minimize its expectation $\mathbb{E}\big[L(u,\xi)\big]$.
- This is even justified if the same problem is solved a large number of time (Law of Large Number).
- In some cases the expectation is not really representative of your risk attitude. Lets consider two examples:
  - Are you ready to pay \$1000 to have one chance over ten to win \$10000 ?
  - You need to be at the airport in 1 hour or you miss your flight, you have the choice between two mean of transport, one of them take surely 50', the other take 40' four times out of five, and 70' one time out of five.

Alternative cost functions II

Here are some cost functions you might consider

- Probability of reaching a given level of cost : $\mathbb{P}(L(u,\xi) \leq 0)$
- Value-at-Risk of costs $V@R_\alpha(L(u,\xi))$, where for any real valued random variable $\boldsymbol{X}$,

$$V@R_\alpha(\boldsymbol{X}) := \inf_{t\in\mathbb{R}} \Big\{\mathbb{P}(\boldsymbol{X} \geq t) \leq \alpha\Big\}.$$

  In other word there is only a probability of α of obtaining a cost worse than $V@R_\alpha(\boldsymbol{X})$.
- Average Value-at-Risk of costs $AV@R_\alpha(L(u,\xi))$, which is the expected cost over the α worst outcomes.

Alternative constraints

- The natural extension of the deterministic constraint $g(u,\xi) \leq 0$ to $g(u,\xi) \leq 0 \; \mathbb{P}-as$ can be extremely conservative, and even often without any admissible solutions.
- For example, if u is a level of production that need to be greater than the demand. In a deterministic setting the realized demand is equal to the forecast. In a stochastic setting we add an error to the forecast. If the error is unbounded (e.g. Gaussian) no control u is admissible.

Alternative constraints II

Here are a few possible constraints

- $\mathbb{E}\big[g(u,\xi)\big] \leq 0$, for quality of service like constraint.
- $\mathbb{P}(g(u,\xi) \leq 0) \geq 1-\alpha$ for chance constraint. Chance constraint is easy to present, but might lead to misconception as nothing is said on the event where the constraint is not satisfied.
- $AV@R_\alpha(g(u,\xi)) \leq 0$

Presentation Outline

Computing expectation

- Computing an expectation $\mathbb{E}\big[L(u,\xi)\big]$ for a given u is costly.
- If ξ is a r.v. with known law admitting a density, $\mathbb{E}\big[L(u,\xi)\big]$ is a (multidimensional) integral.
- If ξ is a r.v. with known discrete law, $\mathbb{E}\big[L(u,\xi)\big]$ is a sum over all possible realizations of ξ, which can be huge.
- If ξ is a r.v. that can be simulated but with unknown law, $\mathbb{E}\big[L(u,\xi)\big]$ cannot be computed exactly.

Computing expectation

- Computing an expectation $\mathbb{E}\big[L(u,\xi)\big]$ for a given u is costly.
- If ξ is a r.v. with known law admitting a density, $\mathbb{E}\big[L(u,\xi)\big]$ is a (multidimensional) integral.
- If ξ is a r.v. with known discrete law, $\mathbb{E}\big[L(u,\xi)\big]$ is a sum over all possible realizations of ξ, which can be huge.
- If ξ is a r.v. that can be simulated but with unknown law, $\mathbb{E}\big[L(u,\xi)\big]$ cannot be computed exactly.

Solution : use Law of Large Number (LLN) and Central Limit Theorem (CLT).

- Draw $N \simeq 1000$ realization of ξ.
- Compute the sample average $\frac{1}{N}\sum_{s=1}^{N} L(u,\xi_s)$.
- Use CLT to give an asymptotic confidence interval of the expectation.

Computing expectation

- Computing an expectation $\mathbb{E}\big[L(u,\xi)\big]$ for a given u is costly.
- If ξ is a r.v. with known law admitting a density, $\mathbb{E}\big[L(u,\xi)\big]$ is a (multidimensional) integral.
- If ξ is a r.v. with known discrete law, $\mathbb{E}\big[L(u,\xi)\big]$ is a sum over all possible realizations of ξ, which can be huge.
- If ξ is a r.v. that can be simulated but with unknown law, $\mathbb{E}\big[L(u,\xi)\big]$ cannot be computed exactly.

Solution : use Law of Large Number (LLN) and Central Limit Theorem (CLT).

- Draw $N \simeq 1000$ realization of ξ.
- Compute the sample average $\frac{1}{N}\sum_{s=1}^{N} L(u,\xi_s)$.
- Use CLT to give an asymptotic confidence interval of the expectation.

This is known as the Monte-Carlo method.

Consequence : evaluating a solution is difficult

- In stochastic optimization even evaluating the value of a solution can be difficult an require approximate methods.
- The same holds true for checking admissibility of a candidate solution.
- It is even more difficult to obtain first order informations (gradient).

Standard solution : sampling and solving the sampled problem (Sample Average Approximation).

Consequence : evaluating a solution is difficult

- In stochastic optimization even evaluating the value of a solution can be difficult an require approximate methods.
- The same holds true for checking admissibility of a candidate solution.
- It is even more difficult to obtain first order informations (gradient).

Standard solution : sampling and solving the sampled problem (Sample Average Approximation).

Recall on CLT

- Let $\{C_i\}_{i\in\mathbb{N}}$ be a sequence of identically distributed random variables with finite variance.
- Then the Central Limit Theorem ensures that

$$\sqrt{N}\Big(\frac{\sum_{i=1}^N \boldsymbol{C}_i}{N} - \mathbb{E}[\boldsymbol{C}_1]\Big) \Longrightarrow G \sim \mathcal{N}(0, Var[\boldsymbol{C}_1]) ,$$

  where the convergence is in law.
- In practice, it is often used in the following way. Asymptotically,

$$\mathbb{P}\Big(\mathbb{E}[C_1] \in \Big[\bar{\boldsymbol{C}}_N - \frac{1.96\boldsymbol{\sigma}_N}{\sqrt{N}}, \bar{\boldsymbol{C}}_N + \frac{1.96\boldsymbol{\sigma}_N}{\sqrt{N}}\Big]\Big) \simeq 95\% ,$$

  where $\bar{\boldsymbol{C}}_N = \frac{\sum_{i=1}^N \boldsymbol{C}_i}{N}$ is the empirical mean and $\boldsymbol{\sigma}_N = \sqrt{\frac{\sum_{i=1}^N (\boldsymbol{C}_i - \bar{\boldsymbol{C}}_N)^2}{N-1}}$ the empirical standard deviation.

Optimization problem and simulator

- Generally speaking stochastic optimization problem are not well posed and often need to be approximated before solving them.
- Good practice consists in defining a simulator, i.e. a representation of the "real problem" on which solution can be tested.
- Then find a candidate solution by solving an (or multiple) approximated problem.
- Finally evaluate the candidate solutions on the simulator. The comparison can be done on more than one dimension (e.g. constraints, risk...)

Conclusion

When addressing an optimization problem under uncertain one has to consider carefully

- How to model uncertainty? (random variable or uncertainty set)
- How to represent your attitude toward risk? (expectation, probability level,...)
- How to include constraints?
- What is your information stucture? (More on that later)
- Set up a simulator and 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

Presentation Outline

One-Stage Problems

Assume that ξ has a discrete distribution [1], with $\mathbb{P}(\xi = \xi^s) = \pi^s > 0$ for $s \in [\![1, S]\!]$. Then, the one-stage problem

$$\begin{aligned} \min_{u_0} \quad & \mathbb{E}\Big[L(u_0, \xi)\Big] \\ s.t. \quad & g(u_0, \xi) \leq 0, \qquad \mathbb{P} - a.s \end{aligned}$$

can be written

$$\begin{aligned} \min_{u_0} \quad & \sum_{s=1}^{S} \pi^s L(u_0, \xi^s) \\ s.t \quad & g(u_0, \xi^s) \leq 0, \qquad \forall s \in [\![1, S]\!]. \end{aligned}$$

[1] If the distribution is continuous we can sample and work on the sampled distribution, this is called the Sample Average Approximation approach with lots of guarantee and results

Newsvendor problem (continued)

We assume that the demand can take value $\{d^s\}_{s\in[\![1,S]\!]}$ with probabilities $\{\pi^s\}_{s\in[\![1,S]\!]}$.

Newsvendor problem (continued)

We assume that the demand can take value $\{d^s\}_{s\in[\![1,S]\!]}$ with probabilities $\{\pi^s\}_{s\in[\![1,S]\!]}$.
In this case the stochastic newsvendor problem reads

$$\begin{aligned} \min_{u} \quad & \sum_{s=1}^{S} \pi^s \Big(cu - p \min(u, d^s) \Big) \\ s.t. \quad & u \geq 0 \end{aligned}$$

Presentation Outline

Recourse Variable

In most problem we can make a correction u_1 once the uncertainty is known:

$$u_0 \rightsquigarrow \xi_1 \rightsquigarrow u_1.$$

As recourse control $\boldsymbol{u}_1$ is a function of $\boldsymbol{\xi}$ it is a random variable, the two-stage optimization problem then reads

$$\begin{aligned} \min_{u_0, \boldsymbol{u}_1} \quad & \mathbb{E}\Big[L(u_0, \boldsymbol{\xi}, \boldsymbol{u}_1)\Big] \\ s.t. \quad & g(u_0, \boldsymbol{\xi}, \boldsymbol{u}_1) \leq 0, \qquad \mathbb{P} - a.s \\ & \boldsymbol{u}_1 \preceq \boldsymbol{\xi} \end{aligned}$$

- u_0 is called a first stage control
- $\boldsymbol{u}_1$ is called a second stage (or recourse) control

Recourse Variable

In most problem we can make a correction u_1 once the uncertainty is known:

$$u_0 \rightsquigarrow \xi_1 \rightsquigarrow u_1.$$

As recourse control $\boldsymbol{u}_1$ is a function of $\boldsymbol{\xi}$ it is a random variable, the two-stage optimization problem then reads

$$\begin{aligned} \min_{u_0,\boldsymbol{u}_1} \quad & \mathbb{E}\Big[L(u_0,\boldsymbol{\xi},\boldsymbol{u}_1)\Big] \\ s.t. \quad & g(u_0,\boldsymbol{\xi},\boldsymbol{u}_1) \leq 0, \qquad \mathbb{P}-a.s \\ & \boldsymbol{u}_1 \preceq \boldsymbol{\xi} \end{aligned}$$

- u_0 is called a first stage control
- $\boldsymbol{u}_1$ is called a second stage (or recourse) control

Two-stage Problem

The extensive formulation of

$$\begin{aligned} \min_{u_0, \boldsymbol{u}_1} \quad & \mathbb{E}\Big[L(u_0, \boldsymbol{\xi}, \boldsymbol{u}_1)\Big] \\ s.t. \quad & g(u_0, \boldsymbol{\xi}, \boldsymbol{u}_1) \leq 0, \qquad \mathbb{P} - a.s \\ & \boldsymbol{u}_1 \preceq \boldsymbol{\xi} \end{aligned}$$

is

$$\begin{aligned} \min_{u_0, \{u_1^s\}_{s \in [\![1,S]\!]}} \quad & \sum_{s=1}^{S} \pi^s L(u_0, \xi^s, u_1^s) \\ s.t \quad & g(u_0, \xi^s, u_1^s) \leq 0, \qquad \forall s \in [\![1, S]\!]. \end{aligned}$$

It is a deterministic problem that can be solved with standard tools or specific methods.

Two-stage newsvendor problem

We can represent the newsvendor problem in a 2-stage framework.

- Let u_0 be the number of newspaper bought in the morning.
- let u_1 be the number of newspaper sold during the day.

Two-stage newsvendor problem

We can represent the newsvendor problem in a 2-stage framework.

- Let u_0 be the number of newspaper bought in the morning.
  $\rightsquigarrow$ first stage control
- let u_1 be the number of newspaper sold during the day.
  $\rightsquigarrow$ second stage control

Two-stage newsvendor problem

We can represent the newsvendor problem in a 2-stage framework.

- Let u_0 be the number of newspaper bought in the morning.
  $\rightsquigarrow$ first stage control
- let $\boldsymbol{u}_1$ be the number of newspaper sold during the day.
  $\rightsquigarrow$ second stage control

The problem reads

$$
\begin{aligned}
\min_{u_0, \boldsymbol{u}_1} \quad & \mathbb{E}\Big[c u_0 - p \boldsymbol{u}_1 \Big] & \\
s.t. \quad & u_0 \geq 0 & \\
& \boldsymbol{u}_1 \leq u_0 & \mathbb{P} - as \\
& \boldsymbol{u}_1 \leq \boldsymbol{d} & \mathbb{P} - as \\
& \sigma(\boldsymbol{u}_1) \subset \sigma(\boldsymbol{d}) &
\end{aligned}
$$

Two-stage newsvendor problem II

In extensive formulation the problem reads

$$\begin{aligned}
\min_{u_0,\{u_1^s\}_{s\in[\![1,S]\!]}} \quad & \sum_{s=1}^{S} \pi^s (c u_0 - p u_1^s) \\
s.t. \quad & u_0 \geq 0 \\
& u_1^s \leq u_0 && \forall s \in [\![1,S]\!] \\
& u_1^s \leq d^s && \forall s \in [\![1,S]\!]
\end{aligned}$$

Note that there are as many second-stage control u_1^s as there are possible realization of the demand $\boldsymbol{d}$, but only one first-stage control u_0.

Two-stage newsvendor problem II

In extensive formulation the problem reads

$$
\begin{aligned}
\min_{u_0, \{u_1^s\}_{s \in [\![1,S]\!]}} \quad & \sum_{s=1}^{S} \pi^s \left(c u_0 - p u_1^s \right) \\
s.t. \quad & u_0 \geq 0 \\
& u_1^s \leq u_0 && \forall s \in [\![1, S]\!] \\
& u_1^s \leq d^s && \forall s \in [\![1, S]\!]
\end{aligned}
$$

Note that there are as many second-stage control u_1^s as there are possible realization of the demand $\boldsymbol{d}$, but only one first-stage control u_0.

Presentation Outline

Time decomposition of the problem

We presented the generic two-stage problem as

$$\begin{aligned} \min_{u_0, \boldsymbol{u}_1} \quad & \mathbb{E}\Big[L(u_0, \boldsymbol{\xi}, \boldsymbol{u}_1)\Big] \\ s.t. \quad & g(u_0, \boldsymbol{\xi}, \boldsymbol{u}_1) \leq 0, \qquad \mathbb{P} - a.s \\ & \boldsymbol{u}_1 \preceq \boldsymbol{\xi} \end{aligned}$$

With $L(u_0, \xi, u_1) = L_0(u_0) + L_1(u_0, \xi, u_1)$, it can also be written as

$$\begin{aligned} \min_{u_0} \quad & L_0(u_0) + \mathbb{E}\Big[\tilde{Q}(u_0, \boldsymbol{\xi})\Big] \\ s.t. \quad & g_0(u_0) \leq 0 \end{aligned}$$

where

$$\begin{aligned} \tilde{Q}(u_0, \xi) := \min_{u_1} \quad & L_1(u_0, \xi, u_1) \\ s.t. \quad & g_1(u_0, \xi, u_1) \leq 0 \end{aligned}$$

Time decomposition of the problem

We presented the generic two-stage problem as

$$\begin{aligned} \min_{u_0, \boldsymbol{u}_1} \quad & \mathbb{E}\Big[L(u_0, \boldsymbol{\xi}, \boldsymbol{u}_1)\Big] \\ s.t. \quad & g(u_0, \boldsymbol{\xi}, \boldsymbol{u}_1) \leq 0, \qquad \mathbb{P} - a.s \\ & \boldsymbol{u}_1 \preceq \boldsymbol{\xi} \end{aligned}$$

With $L(u_0, \xi, u_1) = L_0(u_0) + L_1(u_0, \xi, u_1)$, it can also be written as

$$\begin{aligned} \min_{u_0} \quad & L_0(u_0) + \mathbb{E}\Big[\tilde{Q}(u_0, \boldsymbol{\xi})\Big] \qquad \text{first stage problem} \\ s.t. \quad & g_0(u_0) \leq 0 \end{aligned}$$

where

$$\begin{aligned} \tilde{Q}(u_0, \xi) := \min_{u_1} \quad & L_1(u_0, \xi, u_1) \qquad \text{second stage problem} \\ s.t. \quad & g_1(u_0, \xi, u_1) \leq 0 \end{aligned}$$

Time decomposition of the problem

We presented the generic two-stage problem as

$$\begin{aligned} \min_{u_0, \boldsymbol{u}_1} \quad & \mathbb{E}\Big[L(u_0, \boldsymbol{\xi}, \boldsymbol{u}_1)\Big] \\ s.t. \quad & g(u_0, \boldsymbol{\xi}, \boldsymbol{u}_1) \leq 0, \qquad \mathbb{P} - a.s \\ & \boldsymbol{u}_1 \preceq \boldsymbol{\xi} \end{aligned}$$

With $L(u_0, \xi, u_1) = L_0(u_0) + L_1(u_0, \xi, u_1)$, it can also be written as

$$\begin{aligned} \min_{u_0} \quad & L_0(u_0) + \mathbb{E}\Big[\tilde{Q}(u_0, \boldsymbol{\xi})\Big] \qquad \text{first stage problem} \\ s.t. \quad & g_0(u_0) \leq 0 \end{aligned}$$

where

$$\begin{aligned} \tilde{Q}(u_0, \xi) := \min_{u_1} \quad & L_1(u_0, \xi, u_1) \qquad \text{second stage problem} \\ s.t. \quad & g_1(u_0, \xi, u_1) \leq 0 \end{aligned}$$

Admissible set

For a given decomposition, we set

$$U_0 := \{u_0 \in \mathbb{R}^{n_0} \quad | \quad g_0(u_0) \leq 0\}$$
$$\widetilde{U}_1(u_0,\xi) := \{u_1 \in \mathbb{R}^{n_1} \quad | \quad g_1(u_0,\xi,u_1) \leq 0\}$$

Note that

- $\widetilde{U}_1(u_0,\xi)$ is the set of admissible solutions of the second stage problem
- U_0 contains the set of admissible solutions of the first stage problem

Admissible set

For a given decomposition, we set

$$\begin{aligned} U_0 &:= \{ u_0 \in \mathbb{R}^{n_0} \quad | \quad g_0(u_0) \leq 0 \} \\ \widetilde{U}_1(u_0, \xi) &:= \{ u_1 \in \mathbb{R}^{n_1} \quad | \quad g_1(u_0, \xi, u_1) \leq 0 \} \end{aligned}$$

Note that

- $\widetilde{U}_1(u_0, \xi)$ is the set of admissible solutions of the second stage problem
- U_0 contains the set of admissible solutions of the first stage problem

- We say that we are in a complete recourse framework, if for all $u_0 \in \mathbb{R}^n$, and almost-all possible outcome $\boldsymbol{\xi}$, there exists a control $\boldsymbol{u}_1$ that is admissible, i.e.,

$$\mathbb{P}\big(\widetilde{U}_1(u_0, \boldsymbol{\xi}) \neq \emptyset\big) = 1, \quad \forall u_0 \in \mathbb{R}^{n_0}.$$

- We say that we are in a relatively complete recourse framework, if for all $u_0 \in U_0$, and almost-all possible outcome ξ, there exists a control $\boldsymbol{u}_1$ that is admissible, i.e.,

$$\mathbb{P}\big(\widetilde{U}_1(u_0, \boldsymbol{\xi}) \neq \emptyset\big) = 1, \quad \forall u_0 \in U_0.$$

Obtaining relatively complete recourse

Assume that the two-stage program is given by

$$\min_{u_0 \in U_0} \Big\{ L_0(u_0) + \mathbb{E}\big[\tilde{Q}(u_0, \xi)\big] \Big\} \quad \text{and} \quad \tilde{Q}(u_0, \xi) := \min_{u_1 \in \widetilde{U}_1(u_0, \xi)} L_1(u_0, \xi, u_1)$$

with finite value, but not necessarily relatively complete recourse.

Then the program is equivalent to

$$\min_{u_0 \in U_0 \cap U_0^{ind}} \Big\{ L_0(u_0) + \mathbb{E}\big[\tilde{Q}(u_0, \xi)\big] \Big\} \quad \text{and} \quad \tilde{Q}(u_0, \xi) := \min_{u_1 \in \widetilde{U}_1(u_0, \xi)} L_1(u_0, \xi, u_1)$$

where U_0^{ind} is the set of induced constraints given by

$$U_0^{ind} = \Big\{ u_0 \in \mathbb{R}^{n_0} \quad | \quad \mathbb{P}\big(\widetilde{U}_1(u_0, \xi) \neq \emptyset\big) = 1 \Big\},$$

and with this formulation, we are in a relatively complete recourse framework.

Obtaining relatively complete recourse

Assume that the two-stage program is given by

$$\min_{u_0 \in U_0} \Big\{ L_0(u_0) + \mathbb{E}\big[\tilde{Q}(u_0,\xi)\big] \Big\} \qquad \text{and} \qquad \tilde{Q}(u_0,\xi) := \min_{u_1 \in \widetilde{U}_1(u_0,\xi)} L_1(u_0,\xi,u_1)$$

with finite value, but not necessarily relatively complete recourse.
Then the program is equivalent to

$$\min_{u_0 \in U_0 \cap U_0^{ind}} \Big\{ L_0(u_0) + \mathbb{E}\big[\tilde{Q}(u_0,\xi)\big] \Big\} \qquad \text{and} \qquad \tilde{Q}(u_0,\xi) := \min_{u_1 \in \widetilde{U}_1(u_0,\xi)} L_1(u_0,\xi,u_1)$$

where U_0^{ind} is the set of induced constraints given by

$$U_0^{ind} = \Big\{ u_0 \in \mathbb{R}^{n_0} \quad | \quad \mathbb{P}\big(\widetilde{U}_1(u_0,\xi) \neq \emptyset\big) = 1 \Big\},$$

and with this formulation, we are in a relatively complete recourse framework.

Obtaining relatively complete recourse

Assume that the two-stage program is given by

$$\min_{u_0 \in U_0} \Big\{ L_0(u_0) + \mathbb{E}\big[\tilde{Q}(u_0, \boldsymbol{\xi})\big] \Big\} \qquad \text{and} \qquad \tilde{Q}(u_0, \boldsymbol{\xi}) := \min_{u_1 \in \widetilde{U}_1(u_0, \xi)} L_1(u_0, \xi, u_1)$$

with finite value, but not necessarily relatively complete recourse.
Then the program is equivalent to

$$\min_{u_0 \in U_0 \cap U_0^{ind}} \Big\{ L_0(u_0) + \mathbb{E}\big[\tilde{Q}(u_0, \boldsymbol{\xi})\big] \Big\} \qquad \text{and} \qquad \tilde{Q}(u_0, \boldsymbol{\xi}) := \min_{u_1 \in \widetilde{U}_1(u_0, \xi)} L_1(u_0, \xi, u_1)$$

where U_0^{ind} is the set of induced constraints given by

$$U_0^{ind} = \Big\{ u_0 \in \mathbb{R}^{n_0} \quad | \quad \mathbb{P}\big(\widetilde{U}_1(u_0, \boldsymbol{\xi}) \neq \emptyset\big) = 1 \Big\},$$

and with this formulation, we are in a relatively complete recourse framework.

Presentation Outline

1. Optimization under uncertainty
   - The need of considering uncertainty while making a decision
   - Some considerations on dealing with uncertainty
   - Evaluating a solution
2. Stochastic Programming Approach
   - One-stage Problems
   - Two-stage Problems
   - Recourse assumptions
3. Information and discretization
   - Information Frameworks
   - Sample Average Approximation

Two-stage framework : three information models

Consider the problem

$$\min_{\boldsymbol{u}_0, \boldsymbol{u}_1} \mathbb{E}\left[L(\boldsymbol{u}_0, \boldsymbol{\xi}, \boldsymbol{u}_1)\right]$$

- Open-Loop case : $\boldsymbol{u}_0$ and $\boldsymbol{u}_1$ are deterministic. In this case both controls are choosen without any knowledge of the alea $\boldsymbol{\xi}$. The set of control is small, and an optimal control can be found through specific method (e.g. Stochastic Gradient).
- Two-Stage case : $\boldsymbol{u}_0$ is deterministic and $\boldsymbol{u}_1$ is measurable with respect to $\boldsymbol{\xi}$. This is the problem tackled by the Stochastic Programming case.
- Anticipative case : $\boldsymbol{u}_0$ and $\boldsymbol{u}_1$ are measurable with respect to $\boldsymbol{\xi}$. This case consists in solving one deterministic problem per possible outcome of the alea, and taking the expectation of the value of this problems.

Two-stage framework : three information models

Consider the problem

$$\min_{\boldsymbol{u}_0, \boldsymbol{u}_1} \mathbb{E}\big[L(\boldsymbol{u}_0, \boldsymbol{\xi}, \boldsymbol{u}_1)\big]$$

- Open-Loop case : $\boldsymbol{u}_0$ and $\boldsymbol{u}_1$ are deterministic. In this case both controls are choosen without any knowledge of the alea $\boldsymbol{\xi}$. The set of control is small, and an optimal control can be found through specific method (e.g. Stochastic Gradient).
- Two-Stage case : $\boldsymbol{u}_0$ is deterministic and $\boldsymbol{u}_1$ is measurable with respect to $\boldsymbol{\xi}$. This is the problem tackled by the Stochastic Programming case.
- Anticipative case : $\boldsymbol{u}_0$ and $\boldsymbol{u}_1$ are measurable with respect to $\boldsymbol{\xi}$. This case consists in solving one deterministic problem per possible outcome of the alea, and taking the expectation of the value of this problems.

Two-stage framework : three information models

Consider the problem

$$\min_{\boldsymbol{u}_0,\boldsymbol{u}_1} \mathbb{E}\big[L(\boldsymbol{u}_0,\boldsymbol{\xi},\boldsymbol{u}_1)\big]$$

- Open-Loop case : $\boldsymbol{u}_0$ and $\boldsymbol{u}_1$ are deterministic. In this case both controls are choosen without any knowledge of the alea $\boldsymbol{\xi}$. The set of control is small, and an optimal control can be found through specific method (e.g. Stochastic Gradient).
- Two-Stage case : $\boldsymbol{u}_0$ is deterministic and $\boldsymbol{u}_1$ is measurable with respect to $\boldsymbol{\xi}$. This is the problem tackled by the Stochastic Programming case.
- Anticipative case : $\boldsymbol{u}_0$ and $\boldsymbol{u}_1$ are measurable with respect to $\boldsymbol{\xi}$. This case consists in solving one deterministic problem per possible outcome of the alea, and taking the expectation of the value of this problems.

Splitted formulation

The extended formulation (in a compact way)

$$\min_{u_0, \{u_1^s\}_{s \in [\![1,S]\!]}} \quad \sum_{s=1}^{S} \pi^s L(u_0, \xi^s, u_1^s)$$
$$s.t \quad g(u_0, \xi^s, u_1^s) \leq 0, \qquad \forall s \in [\![1, S]\!].$$

Can be written in a splitted formulation

$$\min_{u_0^s, \{u_1^s\}_{s \in [\![1,S]\!]}} \quad \sum_{s=1}^{S} \pi^s L(u_0^s, \xi^s, u_1^s)$$
$$s.t \quad g(u_0^s, \xi^s, u_1^s) \leq 0, \qquad \forall s \in [\![1, S]\!]$$
$$u_0^s = u_0^{s'} \qquad \forall s, s'$$

Splitted formulation

The extended formulation (in a compact way)

$$\min_{u_0, \{u_1^s\}_{s\in[\![1,S]\!]}} \quad \sum_{s=1}^{S} \pi^s L(u_0, \xi^s, u_1^s)$$
$$s.t \quad g(u_0, \xi^s, u_1^s) \leq 0, \qquad \forall s \in [\![1,S]\!].$$

Can be written in a splitted formulation

$$\min_{\bar{u}_0, u_0^s, \{u_1^s\}_{s\in[\![1,S]\!]}} \quad \sum_{s=1}^{S} \pi^s L(u_0^s, \xi^s, u_1^s)$$
$$s.t \quad g(u_0^s, \xi^s, u_1^s) \leq 0, \qquad \forall s \in [\![1,S]\!]$$
$$u_0^s = \bar{u}_0 \qquad \forall s$$

Splitted formulation

The extended formulation (in a compact way)

$$\min_{u_0, \{u_1^s\}_{s \in [\![1,S]\!]}} \quad \sum_{s=1}^{S} \pi^s L(u_0, \xi^s, u_1^s)$$
$$s.t \quad g(u_0, \xi^s, u_1^s) \leq 0, \qquad \forall s \in [\![1, S]\!].$$

Can be written in a splitted formulation

$$\min_{u_0^s, \{u_1^s\}_{s \in [\![1,S]\!]}} \quad \sum_{s=1}^{S} \pi^s L(u_0^s, \xi^s, u_1^s)$$
$$s.t \quad g(u_0^s, \xi^s, u_1^s) \leq 0, \qquad \forall s \in [\![1, S]\!]$$
$$u_0^s = \sum_{s'} \pi^{s'} u_0^{s'} \qquad \forall s$$

Information models for the Newsvendor I

Open-loop :

$$\begin{aligned}
\min_{u_0,u_1} \quad & \sum_{s=1}^{S} \pi^s (c u_0 - p u_1) \\
s.t. \quad & u_0 \geq 0 \\
& u_1 \leq u_0 \\
& u_1 \leq d^s & \forall s \in [\![1, S]\!]
\end{aligned}$$

Information models for the Newsvendor II

Two-stage :

$$
\begin{aligned}
\min_{u_0,\{u_1^s\}_{s\in[\![1,S]\!]}} \quad & \sum_{s=1}^{S} \pi^s \left(cu_0 - pu_1^s\right) \\
s.t. \quad & u_0 \geq 0 \\
& u_1^s \leq u_0 & \forall s \in [\![1, S]\!] \\
& u_1^s \leq d_s & \forall s \in [\![1, S]\!]
\end{aligned}
$$

Information models for the Newsvendor III

Anticipative :

$$\begin{aligned} \min_{\{u_0^s, u_1^s\}_{s \in [\![1,S]\!]}} \quad & \sum_{s=1}^{S} \pi^s \left(c u_0^s - p u_1^s \right) & \\ s.t. \quad & u_0^s \geq 0 & \forall s \in [\![1, S]\!] \\ & u_1^s \leq u_0 & \forall s \in [\![1, S]\!] \\ & u_1^s \leq d^s & \forall s \in [\![1, S]\!] \end{aligned}$$

Comparing the information models

The three information models can be written this way :

$$\begin{aligned} \min_{\{u_0^s, u_1^s\}_{s \in [\![1,S]\!]}} \quad & \sum_{s=1}^{S} \pi^s \left(c u_0^s - p u_1^s \right) \\ s.t. \quad & u_0^s \geq 0 && \forall s \in [\![1, S]\!] \\ & u_1^s \leq u_0 && \forall i \in [\![1, S]\!] \\ & u_1^s \leq d^s && \forall i \in [\![1, S]\!] \end{aligned}$$

Hence, by simple comparison of constraints we have

$$V^{anticipative}$$

Comparing the information models

The three information models can be written this way :

$$\begin{aligned}
\min_{\{u_0^s, u_1^s\}_{s\in[\![1,S]\!]}} \quad & \sum_{s=1}^{S} \pi^s \left(cu_0^s - pu_1^s\right) & \\
s.t. \quad & u_0^s \geq 0 & \forall s \in [\![1, S]\!] \\
& u_1^s \leq u_0 & \forall i \in [\![1, S]\!] \\
& u_1^s \leq d^s & \forall i \in [\![1, S]\!] \\
& u_0^s = u_0^{s'} & \forall s, s'
\end{aligned}$$

Hence, by simple comparison of constraints we have

$$V^{anticipative} \leq V^{2-stage}$$

Comparing the information models

The three information models can be written this way :

$$\begin{aligned}
\min_{\{u_0^s, u_1^s\}_{s\in[\![1,S]\!]}} \quad & \sum_{s=1}^{S} \pi^s \left(c u_0^s - p u_1^s\right) & \\
s.t. \quad & u_0^s \geq 0 & \forall s \in [\![1, S]\!] \\
& u_1^s \leq u_0 & \forall i \in [\![1, S]\!] \\
& u_1^s \leq d^s & \forall i \in [\![1, S]\!] \\
& u_0^s = u_0^{s'} & \forall s, s' \\
& u_1^s = u_1^{s'} & \forall s, s'
\end{aligned}$$

Hence, by simple comparison of constraints we have

$$V^{anticipative} \leq V^{2-stage} \leq V^{OL}.$$

Value of information

- The Expected Value of Perfect Information (EVPI) is defined as

$$EVPI = v^{2-stage} - v^{anticipative} \geq 0.$$

- Its the maximum amount of money you can gain by getting more information (e.g. incorporating better statistical model in your problem)
- The Value of Stochastic Solution is defined as

$$VSS = v^{OL} - v^{2-stage} \geq 0.$$

- The expected value problem is the value of the deterministic problem where the randomness is replaced by its expectation

$$v^{EV} = \min_{u_0, u_1} L(u_0, \mathbb{E}[\xi], u_1).$$

- If (u_0^{EV}, u_1^{EV}) is the solution of the EV problem, then $\mathbb{E}\big[L(u_0^{EV}, \xi, u_1^{EV})\big]$, is known as Expected Value of Expected Value problem v^{EEV}.

Value of information

- The Expected Value of Perfect Information (EVPI) is defined as

$$EVPI = v^{2-stage} - v^{anticipative} \geq 0.$$

- Its the maximum amount of money you can gain by getting more information (e.g. incorporating better statistical model in your problem)
- The Value of Stochastic Solution is defined as

$$VSS = v^{OL} - v^{2-stage} \geq 0.$$

- The expected value problem is the value of the deterministic problem where the randomness is replaced by its expectation

$$v^{EV} = \min_{u_0, u_1} L(u_0, \mathbb{E}[\xi], u_1).$$

- If (u_0^{EV}, u_1^{EV}) is the solution of the EV problem, then $\mathbb{E}\big[L(u_0^{EV}, \xi, u_1^{EV})\big]$, is known as Expected Value of Expected Value problem v^{EEV}.

Value of information

- The Expected Value of Perfect Information (EVPI) is defined as

$$EVPI = v^{2-stage} - v^{anticipative} \geq 0.$$

- Its the maximum amount of money you can gain by getting more information (e.g. incorporating better statistical model in your problem)
- The Value of Stochastic Solution is defined as

$$VSS = v^{OL} - v^{2-stage} \geq 0.$$

- The expected value problem is the value of the deterministic problem where the randomness is replaced by its expectation

$$v^{EV} = \min_{u_0, u_1} L(u_0, \mathbb{E}[\xi], u_1).$$

- If (u_0^{EV}, u_1^{EV}) is the solution of the EV problem, then $\mathbb{E}\left[L(u_0^{EV}, \xi, u_1^{EV})\right]$, is known as Expected Value of Expected Value problem v^{EEV}.

Value of information

- The Expected Value of Perfect Information (EVPI) is defined as

$$EVPI = v^{2-stage} - v^{anticipative} \geq 0.$$

- Its the maximum amount of money you can gain by getting more information (e.g. incorporating better statistical model in your problem)
- The Value of Stochastic Solution is defined as

$$VSS = v^{OL} - v^{2-stage} \geq 0.$$

- The expected value problem is the value of the deterministic problem where the randomness is replaced by its expectation

$$v^{EV} = \min_{u_0, u_1} L(u_0, \mathbb{E}[\xi], u_1).$$

- If (u_0^{EV}, u_1^{EV}) is the solution of the EV problem, then $\mathbb{E}\left[L(u_0^{EV}, \xi, u_1^{EV})\right]$, is known as Expected Value of Expected Value problem v^{EEV}.

Value of information

- The Expected Value of Perfect Information (EVPI) is defined as

$$EVPI = v^{2-stage} - v^{anticipative} \geq 0.$$

- Its the maximum amount of money you can gain by getting more information (e.g. incorporating better statistical model in your problem)
- The Value of Stochastic Solution is defined as

$$VSS = v^{OL} - v^{2-stage} \geq 0.$$

- The expected value problem is the value of the deterministic problem where the randomness is replaced by its expectation

$$v^{EV} = \min_{u_0, u_1} L(u_0, \mathbb{E}[\xi], u_1).$$

- If (u_0^{EV}, u_1^{EV}) is the solution of the EV problem, then $\mathbb{E}\big[L(u_0^{EV}, \xi, u_1^{EV})\big]$, is known as Expected Value of Expected Value problem v^{EEV}.

Comparison and convexity

- Without assumption we have

$$v^{EEV} \geq v^{OL} \geq v^{2-stage} \geq v^{anticipative}$$

- If additionally L is jointly convex we have

$$\begin{aligned} v^{anticipative} &= \mathbb{E}\Big[L(u_0^{\xi}, \xi, u_1^{\xi})\Big] \\ &\geq L(\mathbb{E}\big[u_0^{\xi}\big], \mathbb{E}\big[\xi\big], \mathbb{E}\big[u_1^{\xi})\big] \\ &\geq L(u_0^{EV}, \mathbb{E}\big[\xi\big], u_1^{EV}) = v^{EV} \end{aligned}$$

- Hence, under convexity we have,

$$v^{EEV} \geq v^{OL} \geq v^{2-stage} \geq v^{anticipative} \geq v^{EV}$$

Solving the problems

- The solution of v^{EEV} is easy to find (one deterministic problem), and its value is obtained by Monte-Carlo.
- v^{OL} can be approximated through specific methods (e.g. SG).
- $v^{2-stage}$ is obtained through Stochastic Programming specific methods. There are two main approaches:
  - Lagrangian decomposition methods (like Progressive-Hedging algorithm).
  - Benders decomposition methods (like L-shaped or nested-decomposition methods).
- $v^{anticipative}$ is difficult to compute exactly but can be estimated through Monte-Carlo approach by drawing a reasonable number of realizations of ξ, solving the deterministic problem for each realization ξ_i and taking the means of the value of the deterministic problem.
- v^{EV} is easy to compute, but is usefull only in the convex case.

Solving the problems

- The solution of v^{EEV} is easy to find (one deterministic problem), and its value is obtained by Monte-Carlo.
- v^{OL} can be approximated through specific methods (e.g. SG).
- $v^{2-stage}$ is obtained through Stochastic Programming specific methods. There are two main approaches:
  - Lagrangian decomposition methods (like Progressive-Hedging algorithm).
  - Benders decomposition methods (like L-shaped or nested-decomposition methods).
- $v^{anticipative}$ is difficult to compute exactly but can be estimated through Monte-Carlo approach by drawing a reasonable number of realizations of ξ, solving the deterministic problem for each realization ξ_i and taking the means of the value of the deterministic problem.
- v^{EV} is easy to compute, but is usefull only in the convex case.

Solving the problems

- The solution of v^{EEV} is easy to find (one deterministic problem), and its value is obtained by Monte-Carlo.
- v^{OL} can be approximated through specific methods (e.g. SG).
- $v^{2-stage}$ is obtained through Stochastic Programming specific methods. There are two main approaches:
  - Lagrangian decomposition methods (like Progressive-Hedging algorithm).
  - Benders decomposition methods (like L-shaped or nested-decomposition methods).
- $v^{anticipative}$ is difficult to compute exactly but can be estimated through Monte-Carlo approach by drawing a reasonable number of realizations of ξ, solving the deterministic problem for each realization ξ_i and taking the means of the value of the deterministic problem.
- v^{EV} is easy to compute, but is usefull only in the convex case.

Solving the problems

- The solution of v^{EEV} is easy to find (one deterministic problem), and its value is obtained by Monte-Carlo.
- v^{OL} can be approximated through specific methods (e.g. SG).
- $v^{2-stage}$ is obtained through Stochastic Programming specific methods. There are two main approaches:
  - Lagrangian decomposition methods (like Progressive-Hedging algorithm).
  - Benders decomposition methods (like L-shaped or nested-decomposition methods).
- $v^{anticipative}$ is difficult to compute exactly but can be estimated through Monte-Carlo approach by drawing a reasonable number of realizations of ξ, solving the deterministic problem for each realization ξ_i and taking the means of the value of the deterministic problem.
- v^{EV} is easy to compute, but is usefull only in the convex case.

Solving the problems

- The solution of v^{EEV} is easy to find (one deterministic problem), and its value is obtained by Monte-Carlo.
- v^{OL} can be approximated through specific methods (e.g. SG).
- $v^{2-stage}$ is obtained through Stochastic Programming specific methods. There are two main approaches:
  - Lagrangian decomposition methods (like Progressive-Hedging algorithm).
  - Benders decomposition methods (like L-shaped or nested-decomposition methods).
- $v^{anticipative}$ is difficult to compute exactly but can be estimated through Monte-Carlo approach by drawing a reasonable number of realizations of ξ, solving the deterministic problem for each realization ξ_i and taking the means of the value of the deterministic problem.
- v^{EV} is easy to compute, but is usefull only in the convex case.

Solving the problems

- The solution of v^{EEV} is easy to find (one deterministic problem), and its value is obtained by Monte-Carlo.
- v^{OL} can be approximated through specific methods (e.g. SG).
- $v^{2-stage}$ is obtained through Stochastic Programming specific methods. There are two main approaches:
  - Lagrangian decomposition methods (like Progressive-Hedging algorithm).
  - Benders decomposition methods (like L-shaped or nested-decomposition methods).
- $v^{anticipative}$ is difficult to compute exactly but can be estimated through Monte-Carlo approach by drawing a reasonable number of realizations of ξ, solving the deterministic problem for each realization ξ_i and taking the means of the value of the deterministic problem.
- v^{EV} is easy to compute, but is usefull only in the convex case.

Solving the problems

- The solution of v^{EEV} is easy to find (one deterministic problem), and its value is obtained by Monte-Carlo.
- v^{OL} can be approximated through specific methods (e.g. SG).
- $v^{2-stage}$ is obtained through Stochastic Programming specific methods. There are two main approaches:
  - Lagrangian decomposition methods (like Progressive-Hedging algorithm).
  - Benders decomposition methods (like L-shaped or nested-decomposition methods).
- $v^{anticipative}$ is difficult to compute exactly but can be estimated through Monte-Carlo approach by drawing a reasonable number of realizations of ξ, solving the deterministic problem for each realization ξ_i and taking the means of the value of the deterministic problem.
- v^{EV} is easy to compute, but is usefull only in the convex case.

Presentation Outline

1. Optimization under uncertainty
   - The need of considering uncertainty while making a decision
   - Some considerations on dealing with uncertainty
   - Evaluating a solution
2. Stochastic Programming Approach
   - One-stage Problems
   - Two-stage Problems
   - Recourse assumptions
3. Information and discretization
   - Information Frameworks
   - Sample Average Approximation

How to deal with continuous distributions ?

Recall that if ξ as finite support we rewrite the 2-stage problem

$$\min_{u_0, \boldsymbol{u}_1} \quad \mathbb{E}\Big[L(u_0, \boldsymbol{\xi}, \boldsymbol{u}_1)\Big]$$
$$s.t. \quad g(u_0, \boldsymbol{\xi}, \boldsymbol{u}_1) \leq 0, \qquad \mathbb{P} - a.s$$

as

$$\min_{u_0, \{u_1^s\}_{s\in[\![1,S]\!]}} \quad \sum_{s=1}^{S} \pi^s L(u_0, \xi^s, u_1^s)$$
$$s.t \quad g(u_0, \xi^s, u_1^s) \leq 0, \qquad \forall s \in [\![1, S]\!].$$

If we consider a continuous distribution (e.g. a Gaussian), we would need aedinfinite number of recourse variables to obtain an extensive formulation.

Simplest idea: sample ξ

First consider the one-stage problem

$$\min_{u \in U} \mathbb{E}\big[L(u,\boldsymbol{\xi})\big] \qquad (\mathcal{P})$$

- Draw a sample $(\xi^1, \dots, \xi^N)$ (in a i.i.d setting with law $\boldsymbol{\xi}$).
- Consider the empirical probability $\hat{\mathbb{P}}_N = \frac{1}{N}\sum_{i=1}^N \delta_{\xi^i}$.
- Replace $\mathbb{P}$ by $\hat{\mathbb{P}}_N$ to obtain a finite-dimensional problem that can be solved.
- This means solving

$$\min_{u \in U} \frac{1}{N}\sum_{i=1}^N L(u,\xi^i) \qquad (\mathcal{P}_N)$$

- We denote by $\hat{\boldsymbol{v}}_N$ (resp. v^*) the value of $(\mathcal{P}_N)$ (resp. $(\mathcal{P})$), and $\boldsymbol{S}_n$ the set of optimal solutions (resp. S^*).

Biased estimator

Generically speaking the estimators of the minimum are biased

$$\mathbb{E}\big[\hat{\boldsymbol{v}}_N\big] \leq \mathbb{E}\big[\hat{\boldsymbol{v}}_{N+1}\big] \leq v^*$$

proof :

- Let $(\xi_i)_{i\in\mathbb{N}}$ be a sequence of iid copies of ξ
- Set $J(u) := \mathbb{E}\big[L(u,\xi)\big]$, $\boldsymbol{J}_N(u) := \frac{1}{N}\sum_{i=1}^N L(u,\xi_i)$
- We have, for every $u' \in U$, $\boldsymbol{J}_N(u') \geq \inf_{u\in U} \boldsymbol{J}_N(u)$.
- Taking the expectation yields,

$$J(u') = \mathbb{E}\Big[\boldsymbol{J}_N(u')\Big] \geq \mathbb{E}\Big[\inf_{u\in U} \boldsymbol{J}_N(u)\Big] = \mathbb{E}\big[\hat{\boldsymbol{v}}_N\big].$$

- We now take the infimum over $u' \in U$, to obtain $v^* = \inf_{u'\in U} J(u') \geq \mathbb{E}\big[\hat{\boldsymbol{v}}_N\big]$.

Biased estimator

Generically speaking the estimators of the minimum are biased

$$\mathbb{E}\big[\hat{\boldsymbol{v}}_N\big] \leq \mathbb{E}\big[\hat{\boldsymbol{v}}_{N+1}\big] \leq v^*$$

proof :

- Let $(\boldsymbol{\xi}_i)_{i\in\mathbb{N}}$ be a sequence of iid copies of $\boldsymbol{\xi}$
- Set $J(u) := \mathbb{E}\big[L(u,\boldsymbol{\xi})\big]$, $\boldsymbol{J}_N(u) := \frac{1}{N}\sum_{i=1}^N L(u,\boldsymbol{\xi}_i)$
- We have, for every $u' \in U$, $\boldsymbol{J}_N(u') \geq \inf_{u\in U} \boldsymbol{J}_N(u)$.
- Taking the expectation yields,

$$J(u') = \mathbb{E}\Big[\boldsymbol{J}_N(u')\Big] \geq \mathbb{E}\Big[\inf_{u\in U} \boldsymbol{J}_N(u)\Big] = \mathbb{E}\big[\hat{\boldsymbol{v}}_N\big].$$

- We now take the infimum over $u' \in U$, to obtain $v^* = \inf_{u'\in U} J(u') \geq \mathbb{E}\big[\hat{\boldsymbol{v}}_N\big]$.

Decreasing bias

We now show that the bias is monotonically decreasing. Notice that

$$J_{N+1}(u) = \frac{1}{N+1}\sum_{i=1}^{N+1}\Big[\frac{1}{N}\sum_{k\neq i} L(u,\xi_k)\Big].$$

Hence,

$$\begin{aligned}
\mathbb{E}\big[\hat{v}_{N+1}\big] &= \mathbb{E}\Big[\inf_{u\in U} J_{N+1}(u)\Big] = \mathbb{E}\Big[\inf_{u\in U}\frac{1}{N+1}\sum_{i=1}^{N+1}\Big[\frac{1}{N}\sum_{k\neq i} L(u,\xi_k)\Big]\Big] \\
&\geq \mathbb{E}\Big[\frac{1}{N+1}\sum_{i=1}^{N+1}\inf_{u_i\in U}\Big[\frac{1}{N}\sum_{k\neq i} L(u_i,\xi_k)\Big]\Big] \\
&= \frac{1}{N+1}\sum_{i=1}^{N+1}\mathbb{E}\Big[\inf_{u_i\in U}\Big[\frac{1}{N}\sum_{k\neq i} L(u_i,\xi_k)\Big]\Big] = \frac{1}{N+1}\sum_{i=1}^{N+1}\mathbb{E}\big[\hat{v}_N\big] = \mathbb{E}\big[\hat{v}_N\big]
\end{aligned}$$

which ends the proof.

Questions?

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