An Introduction to Stochastic Programming

V. Leclère (CERMICS, ENPC)
GT COSMOS

November 25, 2016

Contents

Contents

An optimization problem

A standard optimization problem

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

An optimization problem with uncertainty

Adding uncertainty ξ in the mix

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

Remarks:

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

An optimization problem with uncertainty

Adding uncertainty ξ in the mix

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

Remarks:

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

An optimization problem with uncertainty

Adding uncertainty ξ in the mix

$$\min_{u_0} \quad L(u_0, \xi)$$
$$s.t. \quad g(u_0, \xi) \leq 0$$

Remarks:

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

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 : $\mathbb{P}(L(u, \boldsymbol{\xi}) \leq 0)$
- Value-at-Risk of costs $V@R_\alpha(L(u, \boldsymbol{\xi}))$, 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 $AV@R_\alpha(L(u, \boldsymbol{\xi}))$, 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 $g(u, \boldsymbol{\xi}) \leq 0 \; \mathbb{P} - as$ can be extremely conservative, and even often without any admissible solutions.
- For example, if u is a level of production that need to be greated 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 unbouded (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.
- $AV@R_\alpha(g(u,\xi)) \leq 0$

Contents

One-Stage Problem

Assume that $\boldsymbol{\Xi}$ as a discrete distribution, with $\mathbb{P}(\boldsymbol{\xi} = \xi_i) = p_i > 0$ for $i \in [\![1, n]\!]$. Then, the one-stage problem

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

can be written

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

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 the recourse control u_1 is a function of ξ it is a random variable. The two-stage optimization problem then reads

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

Two-stage Problem

The extensive formulation of

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

is

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

Recourse assumptions

- We say that we are in a *complete recourse* framework, if for all u_0, and all possible outcome $\boldsymbol{\xi}$, every control $\boldsymbol{u}_1$ is admissible.
- We say that we are in a *relatively complete recourse* framework, if for all u_0, and all possible outcome ξ, there exists a control $\boldsymbol{u}_1$ that is admissible.
- For a lot of algorithm relatively complete recourse is a condition of convergence. It means that there is no *induced* constraints.

Multi-stage Problem

- We can consider a multi-stage problems with successive decisions and aleas

$$u_0 \rightsquigarrow \xi_1 \rightsquigarrow u_1 \rightsquigarrow \xi_2 \rightsquigarrow \cdots u_T.$$

- If each each alea ξ_i has 10 possible realizations, then there are
  - 1 control u_0
  - 10 control u_1^i
  - 100 control u_2^i
  - ...
- In practice only two or three-stage problem can be solved by Stochastic Programming approaches.
- Remark : a stage is not necessarily a time-step.

Contents

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 approach : $\boldsymbol{u}_0$ 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 approach : $\boldsymbol{u}_0$ is deterministic and $\boldsymbol{u}_1$ is measurable with respect to $\boldsymbol{\xi}$. This is the problem tackled by Stochastic Programming method.
- Anticipative approach : $\boldsymbol{u}_0$ and $\boldsymbol{u}_1$ are measurable with respect to $\boldsymbol{\xi}$. This approach consists in solving one deterministic problem per possible outcome of the alea, and taking the expectation of the value of this problems.

Comparing the models

- By simple comparison of constraints we have

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

- V^{OL} can be approximated through specific methods (e.g. Stochastic Gradient).
- $V^{2-stage}$ 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 ξ and taking the means of the values.

Information structures in the multistage setting

Open-Loop Every decision $(u_t)_{t\in[\![0,T-1]\!]}$ is taken before any noises $(\xi_t)_{t\in[\![0,T-1]\!]}$ is known. We decide a planning, and stick to it.

Decision Hazard Decision $\boldsymbol{u}_t$ is taken knowing all past noises $\xi_0,\dots,\xi_t$, but not knowing $\boldsymbol{\xi}_{t+1},\dots,\boldsymbol{\xi}_T$.

Hazard Decision Decision $\boldsymbol{u}_t$ is taken knowing all past noises $\xi_0,\dots,\xi_t$, and the next noise $\boldsymbol{\xi}_{t+1}$ but not knowing $\boldsymbol{\xi}_{t+2},\dots,\boldsymbol{\xi}_T$.

Anticipative Every decision $(\boldsymbol{u}_t)_{t\in[\![0,T-1]\!]}$ is taken knowing the whole scenario $(\boldsymbol{\xi}_t)_{t\in[\![0,T-1]\!]}$. There is one deterministic optimization problem by scenario.

With the same objective function this gives better and better value as the solution use more and more information.

Information structures: comments

Open-Loop This case can happen in practice (e.g. fixed planning). There are specific methods to solve this type of optimization problem (e.g. stochastic gradient methods).

Decision Hazard The decision $\boldsymbol{u}_t$ is taken at the beginning of period $[t, t+1[$. The decision is always implementable, and might be conservative as it doesnot leverage any prediction over the noise in $[t, t+1[$.

Hazard Decision The decision $\boldsymbol{u}_t$ is taken at the end of period $[t, t+1[$. The modelization is optimistic as it assumes perfect knowledge that might not be available in practice.

Anticipative This problem is never realistic. However it is a lower bound of the real problem that can be estimated through Monte-Carlo and deterministic optimization.

Contents

Expected value of information

We call *Expected value of information* the difference of value between the real information framework and an anticipative solution (if you had a crystal ball)

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

We are now going to give a price to knowing on which scenario we are.

Expected value of information

We call *Expected value of information* the difference of value between the real information framework and an anticipative solution (if you had a crystal ball)

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

We are now going to give a price to knowing on which scenario we are.

Two-stage Problem

Recall that the extensive formulation of

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

is

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

Rewriting non-anticipativity constraint

Which we can equivalently write

$$\begin{aligned} \min_{\{u_0^i, u_1^i\}_{i\in[\![1,n]\!]}} \quad & \sum_{i=1}^{n} p_i L(u_0^i, \xi_i, u_1^i) \\ s.t \quad & g(u_0^i, \xi_i, u_1^i) \leq 0, \qquad \forall i \in [\![1, n]\!] \\ & u_i = u_j, \qquad \forall i,\ \forall j, \end{aligned}$$

Or again

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

Dualizing non-anticipativity constraint

Dualizing the non-anticipativity constraint we obtain

$$\begin{aligned} \min_{\{u_0^i, u_1^i\}_{i \in [\![1,n]\!]}} \quad & \max_{\lambda : \sum_{i=1}^n p_i \lambda_i = 0} \sum_{i=1}^n p_i \Big[L(u_0, \xi_i, u_1^i) + \lambda_i u_0^i \Big] \\ s.t \quad & g(u_0, \xi_i, u_1^i) \leq 0, \qquad \forall i \in [\![1, n]\!] \end{aligned}$$

With dual

$$\begin{aligned} \max_{\lambda : \sum_{i=1}^n p_i \lambda_i = 0} \sum_{i=1}^n p_i \min_{\{u_0^i, u_1^i\}_{i \in [\![1,n]\!]}} \quad & L(u_0, \xi_i, u_1^i) + \lambda_i u_0^i \\ s.t \quad & g(u_0, \xi_i, u_1^i) \leq 0 \end{aligned}$$

References

- Lectures on Stochastic Programming (Dentcheva, Ruszczynski, Shapiro) chap. 2
- Scenarios and policy aggregation in optimization under uncertainty, (Rockafellar, Wets) 1991
- keyword : progressive-hedging

Contents

Open-Loop Feedback approach

- Another road toward the multistage, that doesnot really rely on a probabilistic vision of the world consists in:
  - consider a forecast of the futur noises
  - solve the deterministic problem
  - apply the first controls until you have more information on the scenario (either because the forecast is not exact or because you can refine the forecast)
  - resolve the deterministic problem with the new information
  - repeat
- Open-Loop Feedback is easy to implement, but it is hard to give theoretical guarantees.
- Open-Loop Feedback doesnot take into account the fact that we might need to modify the solution later on.

Repeated Two-stage Stochastic Programming

- Multi-stage stochastic program are numerically extremely difficult to solve (without Markovian assumption).
- Open-Loop Feedback control doesnot integrate the stochasticity of the problem when designing a solution.
- A mid-way approach consists at any step t to design a two-stage program in order to determine the first stage control u_t to apply to the system. The recourse controls will not be used.
- Questions arise as to what should be first stage control and what should be recourse variable.
- In any cases this type of approach are only heuristics solution to multi-stage approach.

Contents

How to deal with continuous distributions ?

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

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

as

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

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

Simplest idea: sample $\boldsymbol{\xi}$

First consider the one-stage problem

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

- 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

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

- 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^*).

Consistence of estimators and convergence results

- Generically speaking the estimators of the minimum are biased

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

- Under technical assumptions (compacity of admissible solution, lower semicontinuity of costs, ...) we obtain:
  - Law of Large Number extension: $\hat{\boldsymbol{v}}_N \to v^*$ almost surely (according to sampling probability).
  - Convergence of controls: $\mathbb{D}(\boldsymbol{S}_N, S^*) \to 0$ almost surely.
  - Central Limit Theorem ($S = \{u^*\}$): $(\sqrt{N}\hat{\boldsymbol{v}}_N - v^*) \to \boldsymbol{Y}_{u^*}$ where $Y_{u^*} \sim \mathcal{N}(0, \sigma(L(u^*, \boldsymbol{\xi})))$.
  - Central Limit Theorem extension: $\sqrt{N}(\hat{\boldsymbol{v}}_N - v^*) \to \inf_u \boldsymbol{Y}_u$ where $Y_u \sim \mathcal{N}(0, \sigma(L(u, \boldsymbol{\xi})))$.
- Good reference for precise results : Lectures on Stochastic Programming (Dentcheva, Ruszczynski, Shapiro) chap. 5.

Multi-stage SAA

- $2-stage$ problem are special cases of one-stage problem.
- If there is relativaly complete recourse, above results apply directly.
- In the multi-stage case we have to generate a tree and not simply scenario realizations.
- Above results are still available, but the number N should be the number of children at each time-step, thus the total number of scenario is N^T.

Contents

A convex problem

We consider the following robust problem

$$\min_{x \in \mathcal{X}} \quad c^T x$$
$$s.t. \quad g(x,\boldsymbol{\xi}) \leq 0 \qquad \mathbb{P}-a.s.$$

assuming that $g(\cdot,\xi)$ is convex.

The approach consists in drawing N independent sample of ξ, denoted $\{\xi_i\}_{i \in [\![1,n]\!]}$, and solving the following relaxation

$$\min_{x \in \mathcal{X}} \quad c^T x$$
$$s.t. \quad g(x,\xi_i) \leq 0 \qquad \forall i \in [\![1,n]\!]$$

A convex problem

We consider the following robust problem

$$\begin{aligned} \min_{x\in\mathcal{X}} \quad & c^T x \\ s.t. \quad & g(x,\boldsymbol{\xi}) \leq 0 \qquad\qquad \mathbb{P}-a.s. \end{aligned}$$

assuming that $g(\cdot,\xi)$ is convex.
The approach consists in drawing N independent sample of $\boldsymbol{\xi}$, denoted $\{\xi_i\}_{i\in[\![1,n]\!]}$, and solving the following relaxation

$$\begin{aligned} \min_{x\in\mathcal{X}} \quad & c^T x \\ s.t. \quad & g(x,\xi_i) \leq 0 \qquad\qquad \forall i \in [\![1,n]\!] \end{aligned}$$

Solution confidence

We define the probability of violation for decision $x \in \mathbb{R}^n$,

$$G(x) := \mathbb{P}\Big(g(x,\xi) > 0\Big).$$

For an independently drawn (from ξ) sample of size N, we construct the SAA problem and denote $\hat{x}_N$ the (assumed unique) optimal solution.

Then, we have

$$\mathbb{E}\Big[G(\hat{x}_N)\Big] \leq \frac{n}{N+1}.$$

Consequently, by Markov, if we want with probability at least

$1-\beta$, a solution $\hat{x}_N$ with $G(\hat{x}_N) \leq \varepsilon$, we need to choose $N \geq \frac{n}{\varepsilon\beta} - 1$. A subtler bound can be determined :

$$N \geq \frac{2}{\varepsilon}\ln(\frac{1}{\beta}) + 2n + \frac{2n}{\varepsilon}\ln(\frac{2}{\varepsilon}).$$

Solution confidence

We define the probability of violation for decision $x \in \mathbb{R}^n$,

$$G(x) := \mathbb{P}\Big(g(x,\xi) > 0\Big).$$

For an independently drawn (from ξ) sample of size N, we construct the SAA problem and denote $\hat{x}_N$ the (assumed unique) optimal solution.

Then, we have

$$\mathbb{E}\Big[G(\hat{x}_N)\Big] \leq \frac{n}{N+1}.$$

Consequently, by Markov, if we want with probability at least

$1-\beta$, a solution $\hat{x}_N$ with $G(\hat{x}_N) \leq \varepsilon$, we need to choose $N \geq \frac{n}{\varepsilon\beta} - 1$. A subtler bound can be determined :

$$N \geq \frac{2}{\varepsilon}\ln(\frac{1}{\beta}) + 2n + \frac{2n}{\varepsilon}\ln(\frac{2}{\varepsilon}).$$

Solution confidence

We define the probability of violation for decision $x \in \mathbb{R}^n$,

$$G(x) := \mathbb{P}\Big(g(x, \boldsymbol{\xi}) > 0\Big).$$

For an independently drawn (from $\boldsymbol{\xi}$) sample of size N, we construct the SAA problem and denote $\hat{\boldsymbol{x}}_N$ the (assumed unique) optimal solution.

Then, we have

$$\mathbb{E}\Big[G(\hat{\boldsymbol{x}}_N)\Big] \leq \frac{n}{N+1}.$$

Consequently, by Markov, if we want with probability at least

$1-\beta$, a solution $\hat{x}_N$ with $G(\hat{x}_N) \leq \varepsilon$, we need to choose $N \geq \frac{n}{\varepsilon\beta} - 1$. A subtler bound can be determined :

$$N \geq \frac{2}{\varepsilon}\ln\left(\frac{1}{\beta}\right) + 2n + \frac{2n}{\varepsilon}\ln\left(\frac{2}{\varepsilon}\right).$$

Extensions

- Non-unique optimum of SAA can be dealt with through a deterministic optimal solution selection.
- Convex cost function can be dealt with through a re-writing of constraints: minimizing $c(x)$ is equivalent to minimizing z under the constraint $c(x) \leq z$.
- We can also deal with the case where an SAA problem is unbounded.

Problem formulation

The problem we want to solve reads

$$\min_{x\in\mathcal{X}} \quad c(x)$$
$$s.t. \quad \mathbb{P}\Big(g(x,\boldsymbol{\xi}) \leq 0\Big) \geq 1-\varepsilon$$

The approach consists in drawing N independent sample of $\boldsymbol{\xi}$, denoted $\{\xi_i\}_{i\in[\![1,n]\!]}$ and approximating the law of the random variable $\boldsymbol{\xi}$ by a uniform law over the samples denoted $\mathbb{P}_N$.
The SAA problem consists in solving

$$\min_{x\in\mathcal{X}} \quad c(x)$$
$$s.t. \quad \mathbb{P}_N\Big(g(x,\boldsymbol{\xi}) \leq 0\Big) \geq 1-\varepsilon$$

SAA description

Let $\{\xi_i\}_{i\in[\![1,N]\!]}$ be a sequence of i.i.d random samples of ξ. We define

$$\hat{G}_N(x) := \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{\{g(x,\xi_i)>0\}} \rightarrow_N G(x) := \mathbb{P}(g(x,\xi) > 0).$$

The SAA problem of level γ is defined as

$$\begin{aligned} \min_{x\in\mathcal{X}} \quad & c(x) \\ s.t. \quad & \hat{G}_N(x) \leq \gamma \end{aligned}$$

Intuitively,

- if $\gamma \leq \varepsilon$ then a feasible solution of the SAA is likely to be feasible for the original problem;
- if $\gamma \geq \varepsilon$ then the optimal value of the SAA is likely to be a lower bound for the original problem.

Notations and assumptions

We assume that

- $g(x,\cdot)$ is measurable, $g(\cdot,\xi)$ is continuous;
- c is continuous, $\mathcal{X}$ is compact.

Then G and G_N are lower-semicontinuous, and both problem have optimal solution if feasible.

We denote by

- $X^\sharp(\varepsilon)$ (resp. $\hat{X}_N(\gamma)$) the set of optimal solution for the original problem (resp. the SAA approximation).
- $v(\varepsilon)$ the value of the original problem, and $\hat{v}_N(\gamma)$ the optimal value of the SAA.

Notations and assumptions

We assume that

- $g(x, \cdot)$ is measurable, $g(\cdot, \xi)$ is continuous;
- c is continuous, $\mathcal{X}$ is compact.

Then G and G_N are lower-semicontinuous, and both problem have optimal solution if feasible.

We denote by

- $X^\sharp(\varepsilon)$ (resp. $\hat{X}_N(\gamma)$) the set of optimal solution for the original problem (resp. the SAA approximation).
- $v(\varepsilon)$ the value of the original problem, and $\hat{v}_N(\gamma)$ the optimal value of the SAA.

Convergence theory

- If $\gamma = \varepsilon$, we have $\hat{v}_N(\varepsilon) \to_N v(\varepsilon)$ and $\hat{X}_N(\varepsilon) \to_N X^\sharp(\varepsilon)$ with probability one.
- If $\gamma > \varepsilon$, $\mathbb{P}\big(\hat{v}_n(\gamma) \leq v(\varepsilon)\big) \to 1$ exponentially fast ($e^{-\kappa N}$, with $\kappa := (\gamma - \varepsilon)^2/(2\varepsilon)$).
- Similarly, if $\gamma < \varepsilon$, the probability of an SAA-admissible solution to be admissible for the true problem converges to 1 exponentially fast. We can deduce the a-priori sample size required to obtain a feasible solution with high probability.
- Under reasonable regularity assumption, $\gamma = \varepsilon$ is the optimal choice.

Convergence theory

- If $\gamma = \varepsilon$, we have $\hat{v}_N(\varepsilon) \to_N v(\varepsilon)$ and $\hat{X}_N(\varepsilon) \to_N X^\sharp(\varepsilon)$ with probability one.
- If $\gamma > \varepsilon$, $\mathbb{P}\big(\hat{v}_n(\gamma) \leq v(\varepsilon)\big) \to 1$ exponentially fast ($e^{-\kappa N}$, with $\kappa := (\gamma - \varepsilon)^2/(2\varepsilon)$).
- Similarly, if $\gamma < \varepsilon$, the probability of an SAA-admissible solution to be admissible for the true problem converges to 1 exponentially fast. We can deduce the a-priori sample size required to obtain a feasible solution with high probability.
- Under reasonable regularity assumption, $\gamma = \varepsilon$ is the optimal choice.

Convergence theory

- If $\gamma = \varepsilon$, we have $\hat{v}_N(\varepsilon) \to_N v(\varepsilon)$ and $\hat{X}_N(\varepsilon) \to_N X^\sharp(\varepsilon)$ with probability one.
- If $\gamma > \varepsilon$, $\mathbb{P}\big(\hat{v}_n(\gamma) \leq v(\varepsilon)\big) \to 1$ exponentially fast ($e^{-\kappa N}$, with $\kappa := (\gamma - \varepsilon)^2/(2\varepsilon)$).
- Similarly, if $\gamma < \varepsilon$, the probability of an SAA-admissible solution to be admissible for the true problem converges to 1 exponentially fast. We can deduce the a-priori sample size required to obtain a feasible solution with high probability.
- Under reasonable regularity assumption, $\gamma = \varepsilon$ is the optimal choice.

Convergence theory

- If $\gamma = \varepsilon$, we have $\hat{v}_N(\varepsilon) \to_N v(\varepsilon)$ and $\hat{X}_N(\varepsilon) \to_N X^\sharp(\varepsilon)$ with probability one.
- If $\gamma > \varepsilon$, $\mathbb{P}\big(\hat{v}_n(\gamma) \leq v(\varepsilon)\big) \to 1$ exponentially fast ($e^{-\kappa N}$, with $\kappa := (\gamma - \varepsilon)^2/(2\varepsilon)$).
- Similarly, if $\gamma < \varepsilon$, the probability of an SAA-admissible solution to be admissible for the true problem converges to 1 exponentially fast. We can deduce the a-priori sample size required to obtain a feasible solution with high probability.
- Under reasonable regularity assumption, $\gamma = \varepsilon$ is the optimal choice.

Solution validation

Consider that we have a candidate solution x for the true problem.

- To check the feasability, we consider $\hat{G}_N(x)$ as an unbiased estimator of G. It is then easy to obtain an asymptotic upper bound (confidence β) on G:

$$\hat{G}_N(x) + \Phi^{-1}(\beta)\sqrt{\hat{G}_N(x)(1-\hat{G}_N(x))/N},$$

  to compare to ε.

- To obtain a lower bound for the optimal cost it is enough to solve a number of independent SAA approximation, and taking the minimum of the SAA value. In fact depending on the confidence β required of the lower bound, we can determine the number and sizes of SAA problems, and take the k-th smaller SAA-value instead of the smallest value see [Ahmed2008] for more information.

Solution validation

Consider that we have a candidate solution x for the true problem.

- To check the feasability, we consider $\hat{G}_N(x)$ as an unbiased estimator of G. It is then easy to obtain an asymptotic upper bound (confidence β) on G:

$$\hat{G}_N(x) + \Phi^{-1}(\beta)\sqrt{\hat{G}_N(x)(1-\hat{G}_N(x))/N},$$

  to compare to ε.

- To obtain a lower bound for the optimal cost it is enough to solve a number of independent SAA approximation, and taking the minimum of the SAA value. In fact depending on the confidence β required of the lower bound, we can determine the number and sizes of SAA problems, and take the k-th smaller SAA-value instead of the smallest value see [Ahmed2008] for more information.

Solution validation

Consider that we have a candidate solution x for the true problem.

- To check the feasability, we consider $\hat{G}_N(x)$ as an unbiased estimator of G. It is then easy to obtain an asymptotic upper bound (confidence β) on G:

$$\hat{G}_N(x) + \Phi^{-1}(\beta)\sqrt{\hat{G}_N(x)(1 - \hat{G}_N(x))/N},$$

  to compare to ε.

- To obtain a lower bound for the optimal cost it is enough to solve a number of independent SAA approximation, and taking the minimum of the SAA value. In fact depending on the confidence β required of the lower bound, we can determine the number and sizes of SAA problems, and take the k-th smaller SAA-value instead of the smallest value see [Ahmed2008] for more information.

MIP formulation

A MIP SAA formulation consists in drawing N sample of $\boldsymbol{\xi}$ and solving

$$
\begin{aligned}
\min_{x\in\mathcal{X}} \quad & c(x) \\
s.t. \quad & g(x,\xi_j) \leq M_j z_j && \forall j \in [\![1,N]\!] \\
& \sum_{j=1}^{N} z_j \leq \gamma N \\
& z_j \in \{0,1\} && \forall j \in [\![1,N]\!]
\end{aligned}
$$

Where M_j is a large positive number such that $M_j \geq \max_{x\in\mathcal{X}} g(x,\xi_j)$. Hence, $\sum_{j=1}^{N} z_j \leq \gamma N$ imply that the constraint is satisfied on $(1-\gamma)N$ sample.

Applications ?

- Generally speaking a MIP formulation is hard to solve.
- If c is convex, $\mathcal{X}$ convex it is slightly better.
- If c is linear, $\mathcal{X}$ is conic there exists academic algorithms.
- If c is linear, $\mathcal{X}$ polyhhedral the problem is MILP and good off-the-shelf solver are pretty efficient and allow for reasonable problem.

In any case, knowing tight bounds greatly increase the solver efficiency, and specific bounds can be obtained especially in the separable case.

Conclusion

- Ignoring uncertainty in modelization can be really misleading.
- Costs, constraints, optimal solution are more difficult to represent in an uncertain framework.
- Risk attitude is key and not easy to modelize.
- Multistage stochastic optimization problem are really challenging numerically.
- Two main approaches (for exact solution):
  - Simplify the information structure to fit a 2 or 3-stage Stochastic Programming framework;
  - Make Markovian assumption to use Dynamic Programming approaches.