Stochastic Optimization
One-stage problem

V. Leclère

September 28 2017

École des Ponts ParisTech

UNIVERSITÉ PARIS-EST

Déroulement du cours

1. Problèmes d'optimisation stochastique à une étape
2. Problèmes d'optimisation stochastique à deux étapes
3. Méthodes de résolution et extension au cas multi-stage
4. Programmation Dynamique, cas discret
5. Programmation Dynamique, cas linéaire - quadratique
6. Programmation Dynamique, cas linéaire - convexe

Presentation Outline

A few notions

- set of aleas Ω
- sigma-algebra $\mathcal{F} \subset 2^{\Omega}$
- probability $\mathbb{P} : \mathcal{F} \to [0, 1]$
- real valued random variable $\boldsymbol{X} : \Omega \to \mathbb{R}$
- discrete random variable
- continuous random variable with density function
- expectation : $L^1(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}) \to \mathbb{R}$
- variance $var(\boldsymbol{X}) := \mathbb{E}\Big[(\boldsymbol{X} - \mathbb{E}[X])^2\Big] = \mathbb{E}[\boldsymbol{X}^2] - \mathbb{E}[\boldsymbol{X}]^2$
- standard-deviation $\sigma(\boldsymbol{X}) := \sqrt{var(\boldsymbol{X})}$
- independance

Some calculus

- $\mathbb{E}[\lambda \boldsymbol{X} + \mu \boldsymbol{Y}] = \lambda \mathbb{E}[\boldsymbol{X}] + \mu \mathbb{E}[\boldsymbol{Y}]$
- $var(\boldsymbol{X} + c) = (X)$
- $var(\lambda \boldsymbol{X}) = \lambda^2 var(\boldsymbol{X})$
- $var(\boldsymbol{X} + \boldsymbol{Y}) = var(\boldsymbol{X}) + 2cov(\boldsymbol{X}, \boldsymbol{Y}) + var(\boldsymbol{Y})$
- $\sigma(\boldsymbol{X} + c) = \sigma(\boldsymbol{X})$
- $\sigma(\lambda \boldsymbol{X}) = |\lambda| \sigma(\boldsymbol{X})$

Two results

Theorem (Law of large numbers)

Let $(\boldsymbol{X}_i)_{i\in\mathbb{N}}$ be a sequence of independent and identically distributed random variables (iid r.v.), that are real valued and integrable. Then we have

$$M_n := \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i \to \mathbb{E}\big[\boldsymbol{X}_1\big] \qquad a.s.$$

Theorem (Central Limit Theorem)

Let $(\boldsymbol{X}_i)_{i\in\mathbb{N}}$ be a sequence real valued iid r.v. with finite variance. Then we have

$$\sqrt{n}\frac{M_n - \mathbb{E}\big[\boldsymbol{X}_1\big]}{\sigma(\boldsymbol{X}_1)} \Longrightarrow \mathcal{N}(0,1)$$

Presentation Outline

1. Some probability recalls
2. Dealing with Uncertainty
   - Objective and constraints
   - Evaluating a solution
3. Numerical methods
   - Sample Average Approximation
   - Stochastic Approximation
4. Devoir

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

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

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 ways 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).

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 ways 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 : $\mathbb{E}\big[L(u, \boldsymbol{\xi})\big]$.

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 ways 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 : $g(u, \boldsymbol{\xi}) \leq 0, \qquad \mathbb{P} - a.s.$.

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 $u^\sharp$ is such that $J'(u^\sharp) = 0$, which is

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

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

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

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 : $\mathbb{P}(L(u,\boldsymbol{\xi}) \leq 0)$
- 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 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 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.
- $AV@R_\alpha(g(u,\xi)) \leq 0$

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 $\frac{1}{N}\sum_{i=1}^N L(u,\xi_i)$.
- 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 $\frac{1}{N}\sum_{i=1}^{N} L(u,\xi_i)$.
- 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).

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.

Presentation Outline

How to deal with continuous distributions ?

Recall that if $\boldsymbol{\xi}$ 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

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

- 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 $Y_{u^*} \sim \mathcal{N}(0, \sigma(L(u^*, \xi)))$.
  - Central Limit Theorem extension: $\sqrt{N}(\hat{\boldsymbol{v}}_N - v^*) \to \inf_u \boldsymbol{Y}_u$ where $Y_u \sim \mathcal{N}(0, \sigma(L(u, \xi)))$.
- Good reference for precise results : Lectures on Stochastic Programming (Dentcheva, Ruszczynski, Shapiro) chap. 5.

Presentation Outline

Stochastic Gradient

If $u \mapsto j(u,\xi)$ is convex $\mathbb{P}$-a.s., and J is finite in a neighboorhood of u, then we have

$$\partial J(u) = \mathbb{E}\big[\partial_u j(u,\xi)\big].$$

If moreover $u \mapsto j(u,\xi)$ is differentiable at point u, then so is J, and we have

$$\nabla J(u) = \mathbb{E}\big[\nabla_u j(u,\xi)\big].$$

Stochastic Gradient II

The stochastic (projected) gradient have the following steps

1. Choose u_0. Set $k = 0$.
2. Draw a random realization of the noise ξ_k according to the law of $\boldsymbol{\xi}$.
3. Set direction $d_k := -\nabla_u j(u, \xi_k)$.
4. Choose a step $\tau_k = c/k$.
5. Compute the new point $u_{k+1} := P_U\big(u_k + \tau_k d_k\big)$.
6. Set $k := k + 1$.
7. Test convergence and go to step 2.

Newsvendor problem

1. Solve the (stochastic) newsvendor problem for $c = 1$, $p = 2$ and a uniform demand over $[5, 15]$.
2. Write a Julia function evaluating the empirical cost of a given control on N realisation.
3. Find the optimal value of the problem. Check that the optimal control indeed have this optimal cost up, with precision 95%.
4. Write a function that solve the SAA problem with N scenarios.
5. Show that the SAA value (resp. control) converges almost surely to the optimal value resp. control.
6. (optional) Verify that the expectation of SAA value is a negatively biased estimator.
7. (optional) Find the optimal control through a Stochastic Gradient algorithm.