Stochastic Optimization
Two-stage stochastic programming

V. Leclère

November 4, 2026





Presentation Outline

1. Stochastic Programming Approach
   - One-stage Problems
   - Two-stage Problems
   - The recourse value function
2. Information and discretization
   - Choosing the information structure
   - The value of information
   - Sample Average Approximation

Colour coding

Best effort:

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

After this session, you should be able to

- write a problem as a two-stage problem from its timing, its extensive formulation under finite support, and its recourse value function; say whether it has complete or relatively complete recourse;
- order the values of the nonadaptive, two-stage and perfect-information problems, and compute EVPI and VSS;
- say which way the value of a sampled problem is biased.

Presentation Outline

1. Stochastic Programming Approach
   - One-stage Problems
   - Two-stage Problems
   - The recourse value function
2. Information and discretization
   - Choosing the information structure
   - The value of information
   - Sample Average Approximation

Finitely supported noise

Assumption (FSN, finitely supported noise)

$\boldsymbol{\xi}$ has a finite support, enumerated without repetition by $\Omega := [S] = \{1, \dots, S\}$,

$$\mathbb{P}(\boldsymbol{\xi} = \xi^\omega) = \pi^\omega > 0, \qquad \omega \in \Omega\,.$$

We call ξ^ω a scenario and π^ω its probability.[a]

[a] A continuous law can be sampled, and the sample used instead: the Sample Average Approximation, at the end of this lecture.

One-Stage Problems

Proposition

Under (FSN), the one-stage stochastic problem

$$\begin{aligned} \min_{y_1 \in \mathbb{R}^{n_1}} \quad & \mathbb{E}_{\mathbb{P}}\Big[\ell(y_1, \xi)\Big] \\ s.t. \quad & g(y_1, \xi) \leq 0, \qquad \mathbb{P}\text{-}a.s. \end{aligned}$$

is equivalent to the deterministic problem

$$\begin{aligned} \min_{y_1 \in \mathbb{R}^{n_1}} \quad & \sum_{\omega \in \Omega} \pi^{\omega} \ell(y_1, \xi^{\omega}) \\ s.t. \quad & g(y_1, \xi^{\omega}) \leq 0, \qquad \forall \omega \in \Omega \, . \end{aligned}$$

Newsvendor problem (continued)

The demand takes the values $(d^\omega)_{\omega\in\Omega}$ with probabilities $(\pi^\omega)_{\omega\in\Omega}$.

Newsvendor problem (continued)

The demand takes the values $(d^\omega)_{\omega\in\Omega}$ with probabilities $(\pi^\omega)_{\omega\in\Omega}$.
The stochastic newsvendor problem reads

$$\min_{y_1\geq 0} \quad \sum_{\omega\in\Omega} \pi^\omega \Big(c\, y_1 - p \min(y_1, d^\omega) \Big) ,$$

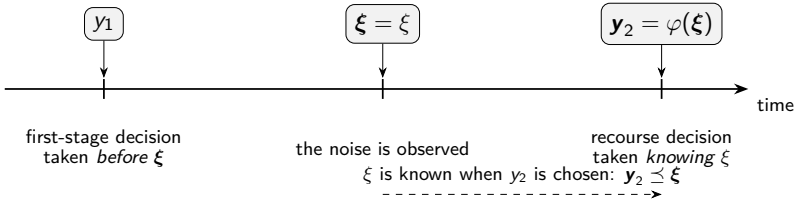
convex and piecewise affine in one variable, with a breakpoint at each demand scenario.

Presentation Outline

1. Stochastic Programming Approach
   - One-stage Problems
   - Two-stage Problems
   - The recourse value function
2. Information and discretization
   - Choosing the information structure
   - The value of information
   - Sample Average Approximation

Recourse

Usually we can act again once the uncertainty is revealed: a second decision y_2 corrects the first, $y_1 \rightsquigarrow \boldsymbol{\xi} \rightsquigarrow y_2$.

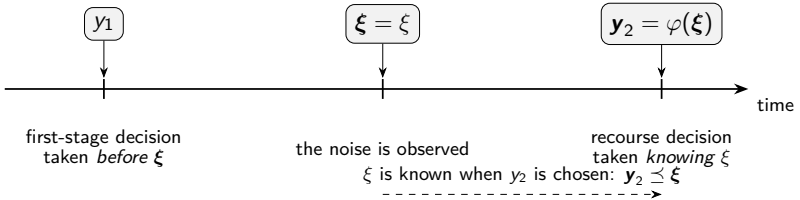


If it may depend on the observation, then $\boldsymbol{y}_2 = \varphi(\boldsymbol{\xi})$ for some φ on $\Xi := \mathrm{supp}(\boldsymbol{\xi})$: the recourse is a random variable, measurable with respect to $\boldsymbol{\xi}$, written

$$\boldsymbol{y}_2 \preceq \boldsymbol{\xi} \,.$$

Recourse

Usually we can act again once the uncertainty is revealed: a second decision y_2 corrects the first, $y_1 \rightsquigarrow \boldsymbol{\xi} \rightsquigarrow y_2$.



If it may depend on the observation, then $\boldsymbol{y}_2 = \varphi(\boldsymbol{\xi})$ for some φ on $\Xi := \mathrm{supp}(\boldsymbol{\xi})$: the recourse is a random variable, measurable with respect to $\boldsymbol{\xi}$, written

$$\boldsymbol{y}_2 \preceq \boldsymbol{\xi} \,.$$

♥ Two-stage Problem

The recourse $\boldsymbol{y}_2$ being a random variable, the two-stage stochastic problem reads, in compact form,

$$\begin{array}{ll} \min\limits_{y_1, \boldsymbol{y}_2} & \mathbb{E}_{\mathbb{P}}\Big[\ell(y_1, \boldsymbol{\xi}, \boldsymbol{y}_2)\Big] \\ \text{s.t.} & g(y_1, \boldsymbol{\xi}, \boldsymbol{y}_2) \leq 0, \qquad \mathbb{P}\text{-a.s.} \\ & \boldsymbol{y}_2 \preceq \boldsymbol{\xi} \end{array}$$

over $y_1 \in \mathbb{R}^{n_1}$ deterministic and $\boldsymbol{y}_2$ random, with values in $\mathbb{R}^{n_2}$.

- y_1 is the first-stage decision;
- $\boldsymbol{y}_2$ is the second-stage (or recourse) decision.

♥ Two-stage Problem

The recourse $\boldsymbol{y}_2$ being a random variable, the two-stage stochastic problem reads, in compact form,

$$\begin{array}{ll} \min\limits_{y_1, \boldsymbol{y}_2} & \mathbb{E}_{\mathbb{P}}\Big[\ell(y_1, \boldsymbol{\xi}, \boldsymbol{y}_2)\Big] \\ \text{s.t.} & g(y_1, \boldsymbol{\xi}, \boldsymbol{y}_2) \leq 0, \qquad \mathbb{P}\text{-a.s.} \\ & \boldsymbol{y}_2 \preceq \boldsymbol{\xi} \end{array}$$

over $y_1 \in \mathbb{R}^{n_1}$ deterministic and $\boldsymbol{y}_2$ random, with values in $\mathbb{R}^{n_2}$.

- y_1 is the first-stage decision;
- $\boldsymbol{y}_2$ is the second-stage (or recourse) decision.

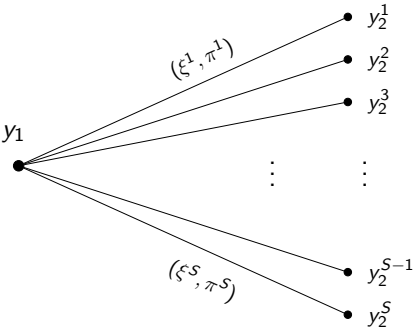
Extensive formulation

Proposition (Extensive formulation)

Under (FSN), the two-stage problem is equivalent to the deterministic problem

$$\min_{y_1,(y_2^\omega)_{\omega\in\Omega}} \quad \sum_{\omega\in\Omega} \pi^\omega \ell(y_1,\xi^\omega,y_2^\omega)$$
$$s.t. \quad g(y_1,\xi^\omega,y_2^\omega)\le 0, \qquad \forall \omega\in\Omega\,,$$

in $n_1 + S\,n_2$ variables, with $y_2^\omega = \boldsymbol{y}_2$ on the event $\{\boldsymbol{\xi} = \xi^\omega\}$.



A deterministic problem, the extensive formulation: give it to a solver first; decomposition methods when it is too large.

Two-stage newsvendor problem I

The newsvendor as a two-stage problem:

- y_1, the papers bought in the morning
- y_2, the papers sold during the day

Two-stage newsvendor problem

The newsvendor as a two-stage problem:

- y_1, the papers bought in the morning $\rightsquigarrow$ first-stage decision;
- y_2, the papers sold during the day $\rightsquigarrow$ second-stage decision.

Two-stage newsvendor problem

The newsvendor as a two-stage problem:

- y_1, the papers bought in the morning $\rightsquigarrow$ first-stage decision;
- y_2, the papers sold during the day $\rightsquigarrow$ second-stage decision.

$$\begin{aligned} \min_{y_1 \geq 0,\, \boldsymbol{y}_2 \geq 0} \quad & \mathbb{E}_{\mathbb{P}}\left[c\, y_1 - p\, \boldsymbol{y}_2\right] \\ \text{s.t.} \quad & \boldsymbol{y}_2 \leq y_1, \quad \boldsymbol{y}_2 \leq \boldsymbol{d}, \qquad \mathbb{P}\text{-a.s.} \\ & \boldsymbol{y}_2 \preceq \boldsymbol{d} \end{aligned}$$

Sales are bounded by the stock and the demand, and depend on the demand; the purchase does not.

Two-stage newsvendor problem II

In extensive formulation:

$$\begin{aligned} \min_{y_1,(y_2^\omega)_{\omega\in\Omega}} \quad & \sum_{\omega\in\Omega} \pi^\omega \left(c\, y_1 - p\, y_2^\omega\right) \\ \text{s.t.} \quad & 0 \leq y_2^\omega \leq y_1 && \forall \omega \in \Omega \\ & y_2^\omega \leq d^\omega && \forall \omega \in \Omega \end{aligned}$$

A linear program in $1 + S$ variables: one second-stage decision y_2^ω per demand scenario, a single first-stage decision y_1.

Two-stage newsvendor problem II

In extensive formulation:

$$\begin{aligned} \min_{y_1,(y_2^\omega)_{\omega\in\Omega}} \quad & \sum_{\omega\in\Omega} \pi^\omega \left(c\, y_1 - p\, y_2^\omega\right) & \\ \text{s.t.} \quad & 0 \leq y_2^\omega \leq y_1 & \forall \omega \in \Omega \\ & y_2^\omega \leq d^\omega & \forall \omega \in \Omega \end{aligned}$$

A linear program in $1 + S$ variables: one second-stage decision y_2^ω per demand scenario, a single first-stage decision y_1.

Presentation Outline

1. Stochastic Programming Approach
   - One-stage Problems
   - Two-stage Problems
   - The recourse value function
2. Information and discretization
   - Choosing the information structure
   - The value of information
   - Sample Average Approximation

Stage decomposition

Suppose the cost and constraints split as

$$\ell(y_1, \xi, y_2) = \ell_1(y_1) + \ell_2(y_1, \xi, y_2),$$
$$g(y_1, \xi, y_2) \leq 0 \iff \Big(g_1(y_1) \leq 0 \quad \text{and} \quad g_2(y_1, \xi, y_2) \leq 0\Big).$$

A modelling choice: a split always exists ($\ell_1 \equiv 0$, $g_1 \equiv 0$), but is not unique.

The recourse value function is the value of the second-stage problem

$$\dot{V}_2(y_1, \xi) := \inf_{y_2 \in \mathbb{R}^{n_2}} \quad \ell_2(y_1, \xi, y_2)$$
$$\text{s.t.} \quad g_2(y_1, \xi, y_2) \leq 0\ ,$$

with $\inf \emptyset = +\infty$; its expectation is $V_2(y_1) := \mathbb{E}_{\mathbb{P}}\big[\dot{V}_2(y_1, \xi)\big]$ (the dot: still a function of the noise).

(BB) Bounded below. For every y_1 with $g_1(y_1) \leq 0$ and every $\xi \in \Xi$, $\dot{V}_2(y_1, \xi) > -\infty$.

Stage decomposition

Suppose the cost and constraints split as

$$\ell(y_1, \xi, y_2) = \ell_1(y_1) + \ell_2(y_1, \xi, y_2),$$
$$g(y_1, \xi, y_2) \leq 0 \iff \Big(g_1(y_1) \leq 0 \quad \text{and} \quad g_2(y_1, \xi, y_2) \leq 0\Big).$$

A modelling choice: a split always exists ($\ell_1 \equiv 0$, $g_1 \equiv 0$), but is not unique.

The recourse value function is the value of the second-stage problem

$$\dot{V}_2(y_1, \xi) := \inf_{y_2 \in \mathbb{R}^{n_2}} \quad \ell_2(y_1, \xi, y_2)$$
$$\text{s.t.} \quad g_2(y_1, \xi, y_2) \leq 0\,,$$

with $\inf \emptyset = +\infty$; its expectation is $V_2(y_1) := \mathbb{E}_{\mathbb{P}}\big[\dot{V}_2(y_1, \xi)\big]$ (the dot: still a function of the noise).

(BB) Bounded below. For every y_1 with $g_1(y_1) \leq 0$ and every $\xi \in \Xi$, $\dot{V}_2(y_1, \xi) > -\infty$.

Stage decomposition

Suppose the cost and constraints split as

$$\ell(y_1, \xi, y_2) = \ell_1(y_1) + \ell_2(y_1, \xi, y_2),$$
$$g(y_1, \xi, y_2) \leq 0 \iff \Big(g_1(y_1) \leq 0 \quad \text{and} \quad g_2(y_1, \xi, y_2) \leq 0\Big).$$

A modelling choice: a split always exists ($\ell_1 \equiv 0$, $g_1 \equiv 0$), but is not unique.

The recourse value function is the value of the second-stage problem

$$\dot{V}_2(y_1, \xi) := \inf_{y_2 \in \mathbb{R}^{n_2}} \quad \ell_2(y_1, \xi, y_2)$$
$$\text{s.t.} \quad g_2(y_1, \xi, y_2) \leq 0\,,$$

with $\inf \emptyset = +\infty$; its expectation is $V_2(y_1) := \mathbb{E}_{\mathbb{P}}\big[\dot{V}_2(y_1, \boldsymbol{\xi})\big]$ (the dot: still a function of the noise).

(BB) Bounded below. For every y_1 with $g_1(y_1) \leq 0$ and every $\xi \in \Xi$, $\dot{V}_2(y_1, \xi) > -\infty$.

♥ Stage decomposition theorem

Theorem (Stage decomposition)

Under (FSN), (BB) and a split, the two-stage problem is equivalent to the first-stage problem

$$\begin{aligned} \min_{y_1 \in \mathbb{R}^{n_1}} \quad & \ell_1(y_1) + V_2(y_1) \\ s.t. \quad & g_1(y_1) \leq 0 \,, \end{aligned}$$

in the sense that

- *the two problems have the same value;*
- *the first stage of a two-stage optimum is optimal for the first-stage problem;*
- *conversely, an optimum $y_1^\star$ of the first-stage problem is the first stage of a two-stage optimum, provided the infima defining $\dot{V}_2(y_1^\star, \xi^\omega)$ are attained.*

Proof idea. In the extensive formulation, minimize first over $(y_2^\omega)_{\omega \in \Omega}$ at fixed y_1: the problem separates, with value $\sum_{\omega \in \Omega} \pi^\omega \dot{V}_2(y_1, \xi^\omega) = V_2(y_1)$.

♥ Stage decomposition theorem

Theorem (Stage decomposition)

Under (FSN), (BB) and a split, the two-stage problem is equivalent to the first-stage problem

$$\begin{aligned} \min_{y_1 \in \mathbb{R}^{n_1}} \quad & \ell_1(y_1) + V_2(y_1) \\ s.t. \quad & g_1(y_1) \leq 0 \,, \end{aligned}$$

in the sense that

- *the two problems have the same value;*
- *the first stage of a two-stage optimum is optimal for the first-stage problem;*
- *conversely, an optimum $y_1^\star$ of the first-stage problem is the first stage of a two-stage optimum, provided the infima defining $\dot{V}_2(y_1^\star, \xi^\omega)$ are attained.*

Proof idea. In the extensive formulation, minimize first over $(y_2^\omega)_{\omega \in \Omega}$ at fixed y_1: the problem separates, with value $\sum_{\omega \in \Omega} \pi^\omega \dot{V}_2(y_1, \xi^\omega) = V_2(y_1)$.

Admissible sets

For a given split, we set

$$\mathcal{Y}_1 := \{y_1 \in \mathbb{R}^{n_1} \mid g_1(y_1) \leq 0\}, \qquad \mathcal{Y}_2(y_1, \xi) := \{y_2 \in \mathbb{R}^{n_2} \mid g_2(y_1, \xi, y_2) \leq 0\} .$$

- $\mathcal{Y}_2(y_1, \xi)$ is the admissible set of the second-stage problem; $\check{V}_2(y_1, \xi) = +\infty$ exactly when it is empty;
- $\mathcal{Y}_1$ only contains the admissible set of the first-stage problem.

Proposition

Under (FSN) and (BB), the first-stage decisions of finite cost are

$$\mathcal{Y}_1^{\text{feas}} := \{y_1 \in \mathcal{Y}_1 \mid \ell_1(y_1) + V_2(y_1) < +\infty\} = \{y_1 \in \mathcal{Y}_1 \mid \mathbb{P}(\mathcal{Y}_2(y_1, \xi) \neq \emptyset) = 1\} \subseteq \mathcal{Y}_1 ,$$

and the inclusion may be strict.

Admissible sets

For a given split, we set

$$\mathcal{Y}_1 := \{y_1 \in \mathbb{R}^{n_1} \mid g_1(y_1) \leq 0\}, \qquad \mathcal{Y}_2(y_1, \xi) := \{y_2 \in \mathbb{R}^{n_2} \mid g_2(y_1, \xi, y_2) \leq 0\} .$$

- $\mathcal{Y}_2(y_1, \xi)$ is the admissible set of the second-stage problem; $\dot{V}_2(y_1, \xi) = +\infty$ exactly when it is empty;
- $\mathcal{Y}_1$ only contains the admissible set of the first-stage problem.

Proposition

Under (FSN) and (BB), the first-stage decisions of finite cost are

$$\mathcal{Y}_1^{\text{feas}} := \{y_1 \in \mathcal{Y}_1 \mid \ell_1(y_1) + V_2(y_1) < +\infty\} = \{y_1 \in \mathcal{Y}_1 \mid \mathbb{P}(\mathcal{Y}_2(y_1, \xi) \neq \emptyset) = 1\} \subseteq \mathcal{Y}_1 ,$$

and the inclusion may be strict.

Admissible sets

For a given split, we set

$$\mathcal{Y}_1 := \{y_1 \in \mathbb{R}^{n_1} \mid g_1(y_1) \leq 0\}, \qquad \mathcal{Y}_2(y_1, \xi) := \{y_2 \in \mathbb{R}^{n_2} \mid g_2(y_1, \xi, y_2) \leq 0\} .$$

- $\mathcal{Y}_2(y_1, \xi)$ is the admissible set of the second-stage problem; $\dot{V}_2(y_1, \xi) = +\infty$ exactly when it is empty;
- $\mathcal{Y}_1$ only contains the admissible set of the first-stage problem.

Proposition

Under (FSN) and (BB), the first-stage decisions of finite cost are

$$\mathcal{Y}_1^{\text{feas}} := \{y_1 \in \mathcal{Y}_1 \mid \ell_1(y_1) + V_2(y_1) < +\infty\} = \{y_1 \in \mathcal{Y}_1 \mid \mathbb{P}(\mathcal{Y}_2(y_1, \xi) \neq \emptyset) = 1\} \subseteq \mathcal{Y}_1 ,$$

and the inclusion may be strict.

Complete and relatively complete recourse

- Complete recourse: every $y_1 \in \mathbb{R}^{n_1}$ has, almost surely, an admissible recourse:

$$\mathbb{P}(\mathcal{Y}_2(y_1, \xi) \neq \emptyset) = 1, \quad \forall y_1 \in \mathbb{R}^{n_1}.$$

- Relatively complete recourse: every $y_1 \in \mathcal{Y}_1$ has, almost surely, an admissible recourse:

$$\mathbb{P}(\mathcal{Y}_2(y_1, \xi) \neq \emptyset) = 1, \quad \forall y_1 \in \mathcal{Y}_1.$$

♦ Obtaining relatively complete recourse

Take a two-stage problem in first-stage form

$$\min_{y_1 \in \mathcal{Y}_1} \Big\{ \ell_1(y_1) + V_2(y_1) \Big\} \qquad \text{with} \quad \dot{V}_2(y_1, \xi) := \inf_{y_2 \in \mathcal{Y}_2(y_1, \xi)} \ell_2(y_1, \xi, y_2),$$

not necessarily with relatively complete recourse. The induced constraints are

$$\mathcal{Y}_1^{\text{ind}} := \Big\{ y_1 \in \mathbb{R}^{n_1} \quad | \quad \mathbb{P}\big(\mathcal{Y}_2(y_1, \boldsymbol{\xi}) \neq \emptyset\big) = 1 \Big\}.$$

Proposition (Induced-constraint reformulation)

Assume (FSN), (BB), and a finite value. Then

$$\inf_{y_1 \in \mathcal{Y}_1} \Big\{ \ell_1(y_1) + V_2(y_1) \Big\} = \inf_{y_1 \in \mathcal{Y}_1 \cap \mathcal{Y}_1^{\text{ind}}} \Big\{ \ell_1(y_1) + V_2(y_1) \Big\},$$

with the same optimal solutions, if any, and the right-hand problem has relatively complete recourse, with admissible set $\mathcal{Y}_1 \cap \mathcal{Y}_1^{\text{ind}}$.

♦ Obtaining relatively complete recourse

Take a two-stage problem in first-stage form

$$\min_{y_1 \in \mathcal{Y}_1} \Big\{ \ell_1(y_1) + V_2(y_1) \Big\} \qquad \text{with} \quad \dot{V}_2(y_1, \xi) := \inf_{y_2 \in \mathcal{Y}_2(y_1, \xi)} \ell_2(y_1, \xi, y_2),$$

not necessarily with relatively complete recourse. The induced constraints are

$$\mathcal{Y}_1^{\text{ind}} := \Big\{ y_1 \in \mathbb{R}^{n_1} \quad | \quad \mathbb{P}\big(\mathcal{Y}_2(y_1, \boldsymbol{\xi}) \neq \emptyset\big) = 1 \Big\}.$$

Proposition (Induced-constraint reformulation)

Assume (FSN), (BB), and a finite value. Then

$$\inf_{y_1 \in \mathcal{Y}_1} \Big\{ \ell_1(y_1) + V_2(y_1) \Big\} = \inf_{y_1 \in \mathcal{Y}_1 \cap \mathcal{Y}_1^{\text{ind}}} \Big\{ \ell_1(y_1) + V_2(y_1) \Big\},$$

with the same optimal solutions, if any, and the right-hand problem has relatively complete recourse, with admissible set $\mathcal{Y}_1 \cap \mathcal{Y}_1^{\text{ind}}$.

Presentation Outline

1. Stochastic Programming Approach
   - One-stage Problems
   - Two-stage Problems
   - The recourse value function
2. Information and discretization
   - Choosing the information structure
   - The value of information
   - Sample Average Approximation

Two-stage framework: three information structures

In $\min_{y_1, y_2} \mathbb{E}_{\mathbb{P}}[\ell(y_1, \xi, y_2)]$, an information structure constrains the decisions, not the data (cost, constraints and law of ξ unchanged).

- Nonadaptative: y_1 and y_2 deterministic, both chosen without knowing ξ. A small decision set; specific methods (e.g. stochastic gradient).
- Two-stage: y_1 deterministic, $y_2 \preceq \xi$. The problem of stochastic programming.
- Perfect information: $y_1 \preceq \xi$ and $y_2 \preceq \xi$. One deterministic problem per outcome, then the expectation of their values.[1]

Nonadaptive



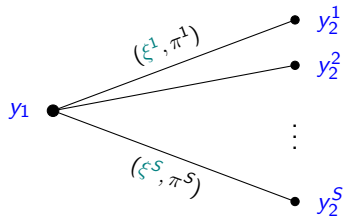
[1] Also called *open-loop* (nonadaptive) and *anticipative* (perfect information). First-stage and recourse decisions are often called *here-and-now* and *wait-and-see* (Shapiro et al.); the wait-and-see value WS of Birge and Louveaux is the perfect-information value.

Two-stage framework: three information structures

In $\min_{\boldsymbol{y}_1, \boldsymbol{y}_2} \mathbb{E}_{\mathbb{P}}[\ell(\boldsymbol{y}_1, \boldsymbol{\xi}, \boldsymbol{y}_2)]$, an information structure constrains the decisions, not the data (cost, constraints and law of $\boldsymbol{\xi}$ unchanged).

- Nonadaptive: y_1 and y_2 deterministic, both chosen without knowing $\boldsymbol{\xi}$. A small decision set; specific methods (e.g. stochastic gradient).
- Two-stage: y_1 deterministic, $\boldsymbol{y}_2 \preceq \boldsymbol{\xi}$. The problem of stochastic programming.
- Perfect information: $\boldsymbol{y}_1 \preceq \boldsymbol{\xi}$ and $\boldsymbol{y}_2 \preceq \boldsymbol{\xi}$. One deterministic problem per outcome, then the expectation of their values.[1]

Two-stage



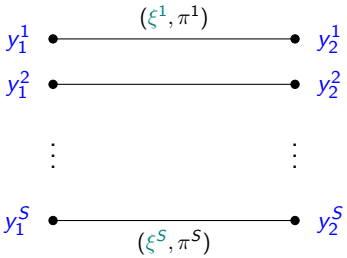
[1] Also called *open-loop* (nonadaptive) and *anticipative* (perfect information). First-stage and recourse decisions are often called *here-and-now* and *wait-and-see* (Shapiro et al.); the wait-and-see value WS of Birge and Louveaux is the perfect-information value.

Two-stage framework: three information structures

In $\min_{\boldsymbol{y}_1, \boldsymbol{y}_2} \mathbb{E}_{\mathbb{P}}[\ell(\boldsymbol{y}_1, \boldsymbol{\xi}, \boldsymbol{y}_2)]$, an information structure constrains the decisions, not the data (cost, constraints and law of $\boldsymbol{\xi}$ unchanged).

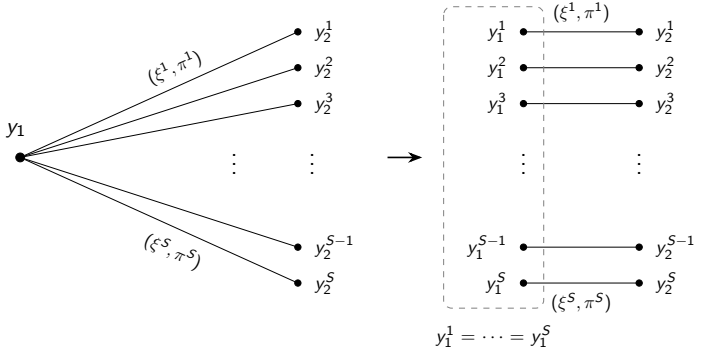
- Nonadaptive: y_1 and y_2 deterministic, both chosen without knowing $\boldsymbol{\xi}$. A small decision set; specific methods (e.g. stochastic gradient).
- Two-stage: y_1 deterministic, $\boldsymbol{y}_2 \preceq \boldsymbol{\xi}$. The problem of stochastic programming.
- Perfect information: $\boldsymbol{y}_1 \preceq \boldsymbol{\xi}$ and $\boldsymbol{y}_2 \preceq \boldsymbol{\xi}$. One deterministic problem per outcome, then the expectation of their values.[1]

Perfect information



[1] Also called *open-loop* (nonadaptive) and *anticipative* (perfect information). First-stage and recourse decisions are often called *here-and-now* and *wait-and-see* (Shapiro et al.); the wait-and-see value WS of Birge and Louveaux is the perfect-information value.

Extensive formulation and scenario form



The same problem written twice. Left, the extensive formulation: the S scenarios form one problem through the shared root decision. Right, the same problem in scenario form: S deterministic problems of the usual size, joined only by the boxed constraint.

Information structures for the newsvendor I

Nonadaptive:

$$\begin{aligned} \min_{y_1, y_2} \quad & \sum_{\omega \in \Omega} \pi^\omega (c\, y_1 - p\, y_2) \\ \text{s.t.} \quad & 0 \leq y_2 \leq y_1 \\ & y_2 \leq d^\omega \qquad \forall \omega \in \Omega \end{aligned}$$

Information structures for the newsvendor II

Two-stage:

$$\min_{y_1,(y_2^\omega)_{\omega\in\Omega}} \quad \sum_{\omega\in\Omega} \pi^\omega (c\, y_1 - p\, y_2^\omega)$$

$$\text{s.t.} \quad 0 \le y_2^\omega \le y_1 \qquad \forall \omega \in \Omega$$

$$y_2^\omega \le d^\omega \qquad \forall \omega \in \Omega$$

Information structures for the newsvendor III

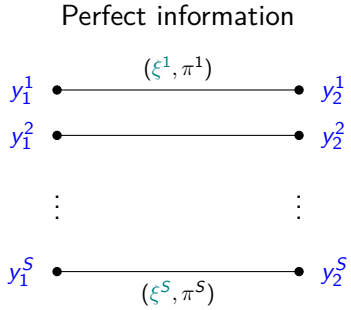
Perfect information:

$$\begin{aligned} \min_{(y_1^\omega, y_2^\omega)_{\omega\in\Omega}} \quad & \sum_{\omega\in\Omega} \pi^\omega \left(c\, y_1^\omega - p\, y_2^\omega\right) \\ \text{s.t.} \quad & 0 \le y_2^\omega \le y_1^\omega && \forall \omega \in \Omega \\ & y_2^\omega \le d^\omega && \forall \omega \in \Omega \end{aligned}$$

Comparing the information structures

The three structures in one formulation:

$$\min_{(y_1^\omega, y_2^\omega)_{\omega\in\Omega}} \quad \sum_{\omega\in\Omega} \pi^\omega \ell\left(y_1^\omega, \xi^\omega, y_2^\omega\right)$$
$$\text{s.t.} \quad g\left(y_1^\omega, \xi^\omega, y_2^\omega\right) \leq 0 \qquad \forall \omega \in \Omega$$



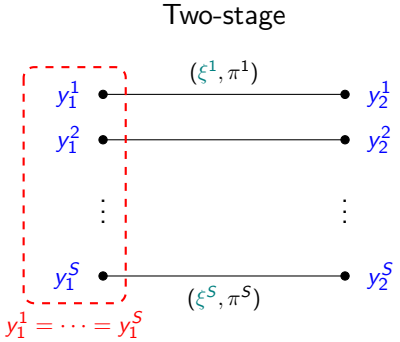
Each added constraint shrinks the feasible set, so more information relaxes: with v^{NA}, v^{SP}, v^{PI} the values of the nonadaptive, two-stage and perfect-information models,

$$v^{\mathrm{PI}}$$

Comparing the information structures

The three structures in one formulation:

$$\begin{aligned} \min_{(y_1^\omega, y_2^\omega)_{\omega\in\Omega}} \quad & \sum_{\omega\in\Omega} \pi^\omega \ell\big(y_1^\omega, \xi^\omega, y_2^\omega\big) \\ \text{s.t.} \quad & g\big(y_1^\omega, \xi^\omega, y_2^\omega\big) \le 0 \qquad \forall \omega \in \Omega \\ & y_1^\omega = y_1^{\omega'} \qquad \forall \omega, \omega' \end{aligned}$$



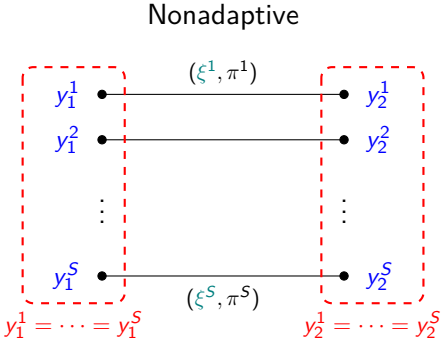
Each added constraint shrinks the feasible set, so more information relaxes: with v^{NA}, v^{SP}, v^{PI} the values of the nonadaptive, two-stage and perfect-information models,

$$v^{\mathrm{PI}} \le v^{\mathrm{SP}}$$

Comparing the information structures

The three structures in one formulation:

$$\begin{aligned} \min_{(y_1^\omega, y_2^\omega)_{\omega\in\Omega}} \quad & \sum_{\omega\in\Omega} \pi^\omega \ell\big(y_1^\omega, \xi^\omega, y_2^\omega\big) \\ \text{s.t.} \quad & g\big(y_1^\omega, \xi^\omega, y_2^\omega\big) \le 0 && \forall \omega \in \Omega \\ & y_1^\omega = y_1^{\omega'} && \forall \omega, \omega' \\ & y_2^\omega = y_2^{\omega'} && \forall \omega, \omega' \end{aligned}$$



Each added constraint shrinks the feasible set, so more information relaxes: with v^{NA}, v^{SP}, v^{PI} the values of the nonadaptive, two-stage and perfect-information models,

$$v^{\mathrm{PI}} \le v^{\mathrm{SP}} \le v^{\mathrm{NA}} .$$

Presentation Outline

1. Stochastic Programming Approach
   - One-stage Problems
   - Two-stage Problems
   - The recourse value function
2. Information and discretization
   - Choosing the information structure
   - The value of information
   - Sample Average Approximation

Value of information

- Expected value of perfect information, $\mathrm{EVPI} := v^{\mathrm{SP}} - v^{\mathrm{PI}} \geq 0$: what knowing ξ in advance would save, hence a ceiling on what any forecast can be worth.
- Expected value problem: the noise replaced by its mean,

$$v^{\mathrm{EV}} := \min_{y_1, y_2} \{\ell(y_1, \mathbb{E}_{\mathbb{P}}[\xi], y_2) \mid g(y_1, \mathbb{E}_{\mathbb{P}}[\xi], y_2) \leq 0\},$$

  with y_1^{EV} the first stage of a solution.
- Expected value of the expected value solution: freeze only the first stage at y_1^{EV}, re-optimize the recourse once ξ is known,

$$v^{\mathrm{EEV}} := \ell_1(y_1^{\mathrm{EV}}) + V_2(y_1^{\mathrm{EV}}).$$

- Value of the stochastic solution, $\mathrm{VSS} := v^{\mathrm{EEV}} - v^{\mathrm{SP}} \geq 0$: what solving the stochastic problem gains over solving at the mean.

Value of information

- Expected value of perfect information, $\mathrm{EVPI} := v^{\mathrm{SP}} - v^{\mathrm{PI}} \geq 0$: what knowing ξ in advance would save, hence a ceiling on what any forecast can be worth.
- Expected value problem: the noise replaced by its mean,

$$v^{\mathrm{EV}} := \min_{y_1, y_2} \{\ell(y_1, \mathbb{E}_{\mathbb{P}}[\xi], y_2) \mid g(y_1, \mathbb{E}_{\mathbb{P}}[\xi], y_2) \leq 0\},$$

  with y_1^{EV} the first stage of a solution.
- Expected value of the expected value solution: freeze only the first stage at y_1^{EV}, re-optimize the recourse once ξ is known,

$$v^{\mathrm{EEV}} := \ell_1(y_1^{\mathrm{EV}}) + V_2(y_1^{\mathrm{EV}}).$$

- Value of the stochastic solution, $\mathrm{VSS} := v^{\mathrm{EEV}} - v^{\mathrm{SP}} \geq 0$: what solving the stochastic problem gains over solving at the mean.

Value of information

- Expected value of perfect information, $\mathrm{EVPI} := v^{\mathrm{SP}} - v^{\mathrm{PI}} \geq 0$: what knowing $\boldsymbol{\xi}$ in advance would save, hence a ceiling on what any forecast can be worth.
- Expected value problem: the noise replaced by its mean,

$$v^{\mathrm{EV}} := \min_{y_1, y_2} \{\ell(y_1, \mathbb{E}_{\mathbb{P}}[\boldsymbol{\xi}], y_2) \mid g(y_1, \mathbb{E}_{\mathbb{P}}[\boldsymbol{\xi}], y_2) \leq 0\},$$

  with y_1^{EV} the first stage of a solution.
- Expected value of the expected value solution: freeze only the first stage at y_1^{EV}, re-optimize the recourse once $\boldsymbol{\xi}$ is known,

$$v^{\mathrm{EEV}} := \ell_1(y_1^{\mathrm{EV}}) + V_2(y_1^{\mathrm{EV}}).$$

- Value of the stochastic solution, $\mathrm{VSS} := v^{\mathrm{EEV}} - v^{\mathrm{SP}} \geq 0$: what solving the stochastic problem gains over solving at the mean.

Value of information

- Expected value of perfect information, $\mathrm{EVPI} := v^{\mathrm{SP}} - v^{\mathrm{PI}} \geq 0$: what knowing $\boldsymbol{\xi}$ in advance would save, hence a ceiling on what any forecast can be worth.
- Expected value problem: the noise replaced by its mean,

$$v^{\mathrm{EV}} := \min_{y_1, y_2} \{\ell(y_1, \mathbb{E}_{\mathbb{P}}[\boldsymbol{\xi}], y_2) \mid g(y_1, \mathbb{E}_{\mathbb{P}}[\boldsymbol{\xi}], y_2) \leq 0\},$$

  with y_1^{EV} the first stage of a solution.
- Expected value of the expected value solution: freeze only the first stage at y_1^{EV}, re-optimize the recourse once $\boldsymbol{\xi}$ is known,

$$v^{\mathrm{EEV}} := \ell_1(y_1^{\mathrm{EV}}) + V_2(y_1^{\mathrm{EV}}).$$

- Value of the stochastic solution, $\mathrm{VSS} := v^{\mathrm{EEV}} - v^{\mathrm{SP}} \geq 0$: what solving the stochastic problem gains over solving at the mean.

Comparing the values

- Without assumption,
$$v^{\mathrm{PI}} \leq v^{\mathrm{SP}} \leq v^{\mathrm{NA}} \qquad \text{and} \qquad v^{\mathrm{SP}} \leq v^{\mathrm{EEV}},$$
- v^{EEV} and v^{NA} are not comparable, the recourse being free in v^{EEV}: the newsvendor below has $v^{\mathrm{EEV}} < v^{\mathrm{NA}}$; the production plan, with no recourse, has $v^{\mathrm{NA}} = v^{\mathrm{SP}} = -100 < v^{\mathrm{EEV}} = 375$.

The newsvendor with $c = 1$, $p = 3$ and a demand of 0 or 100, equiprobable:

v^{PI}	v^{EV}	v^{SP}	v^{EEV}	v^{NA}	EVPI	VSS
-100	-100	-50	-25	0	50	25

The stochastic stock is 100, the expected value stock 50.

Comparing the values

- Without assumption,
$$v^{\mathrm{PI}} \leq v^{\mathrm{SP}} \leq v^{\mathrm{NA}} \qquad \text{and} \qquad v^{\mathrm{SP}} \leq v^{\mathrm{EEV}},$$
- v^{EEV} and v^{NA} are not comparable, the recourse being free in v^{EEV}: the newsvendor below has $v^{\mathrm{EEV}} < v^{\mathrm{NA}}$; the production plan, with no recourse, has $v^{\mathrm{NA}} = v^{\mathrm{SP}} = -100 < v^{\mathrm{EEV}} = 375$.

The newsvendor with $c = 1$, $p = 3$ and a demand of 0 or 100, equiprobable:

v^{PI}	v^{EV}	v^{SP}	v^{EEV}	v^{NA}	EVPI	VSS
-100	-100	-50	-25	0	50	25

The stochastic stock is 100, the expected value stock 50.

♦ Convexity in the right-hand side

- A linear two-stage program with second stage

$$\dot{V}_2(y_1, \xi) = \min_{y_2 \geq 0} \{ c_2^\top y_2 \mid C_2\, y_2 = d_2(\xi) - B_2(\xi)\, y_1 \},$$

with fixed recourse (C_2 deterministic) and c_2 deterministic, d_2 and B_2 affine in ξ, $\boldsymbol{\xi}$ integrable, and a nonempty dual set $D := \{ \lambda \mid C_2^\top \lambda \leq c_2 \}$. Then $\dot{V}_2(y_1, \xi) = \sup_{\lambda \in D} \lambda^\top \big(d_2(\xi) - B_2(\xi) y_1 \big)$ is convex in ξ; by Jensen,

$$V_2(y_1) \geq \dot{V}_2\big(y_1, \mathbb{E}_{\mathbb{P}}[\boldsymbol{\xi}]\big) \quad \text{for every } y_1 \ .$$

- Minimizing over y_1: under these hypotheses,

$$v^{\mathrm{EV}} \leq v^{\mathrm{SP}} \leq v^{\mathrm{NA}}, \qquad v^{\mathrm{PI}} \leq v^{\mathrm{SP}} \leq v^{\mathrm{EEV}} \ .$$

♦ Convexity in the right-hand side

- A linear two-stage program with second stage

$$\dot{V}_2(y_1, \xi) = \min_{y_2 \geq 0} \left\{ c_2^\top y_2 \mid C_2\, y_2 = d_2(\xi) - B_2(\xi)\, y_1 \right\},$$

with fixed recourse (C_2 deterministic) and c_2 deterministic, d_2 and B_2 affine in ξ, $\boldsymbol{\xi}$ integrable, and a nonempty dual set $D := \left\{ \lambda \mid C_2^\top \lambda \leq c_2 \right\}$. Then $\dot{V}_2(y_1, \xi) = \sup_{\lambda \in D} \lambda^\top \big(d_2(\xi) - B_2(\xi) y_1 \big)$ is convex in ξ; by Jensen,

$$V_2(y_1) \geq \dot{V}_2\big(y_1, \mathbb{E}_{\mathbb{P}}[\boldsymbol{\xi}]\big) \quad \text{for every } y_1 \,.$$

- Minimizing over y_1: under these hypotheses,

$$v^{\mathrm{EV}} \leq v^{\mathrm{SP}} \leq v^{\mathrm{NA}}, \qquad v^{\mathrm{PI}} \leq v^{\mathrm{SP}} \leq v^{\mathrm{EEV}} \,.$$

Solving the problems

- y_1^{EV} is easy (one deterministic problem), but v^{EEV} needs a second-stage solve per scenario (or per drawn scenario, by Monte Carlo).
- v^{NA}: one deterministic problem under (FSN); otherwise specific methods (e.g. stochastic gradient).
- v^{SP}: stochastic-programming methods, of two main kinds:
  - Lagrangian decomposition (e.g. progressive hedging);
  - Benders decomposition (e.g. the L-shaped method, nested decomposition).
- v^{PI}: one deterministic problem per scenario; with too many, Monte Carlo: solve for each drawn ξ^n and average the values.
- v^{EV}: easy to compute, but a bound only under convexity.

Solving the problems

- y_1^{EV} is easy (one deterministic problem), but v^{EEV} needs a second-stage solve per scenario (or per drawn scenario, by Monte Carlo).
- v^{NA}: one deterministic problem under (FSN); otherwise specific methods (e.g. stochastic gradient).
- v^{SP}: stochastic-programming methods, of two main kinds:
  - Lagrangian decomposition (e.g. progressive hedging);
  - Benders decomposition (e.g. the L-shaped method, nested decomposition).
- v^{PI}: one deterministic problem per scenario; with too many, Monte Carlo: solve for each drawn ξ^n and average the values.
- v^{EV}: easy to compute, but a bound only under convexity.

Solving the problems

- y_1^{EV} is easy (one deterministic problem), but v^{EEV} needs a second-stage solve per scenario (or per drawn scenario, by Monte Carlo).
- v^{NA}: one deterministic problem under (FSN); otherwise specific methods (e.g. stochastic gradient).
- v^{SP}: stochastic-programming methods, of two main kinds:
  - Lagrangian decomposition (e.g. progressive hedging);
  - Benders decomposition (e.g. the L-shaped method, nested decomposition).
- v^{PI}: one deterministic problem per scenario; with too many, Monte Carlo: solve for each drawn ξ^n and average the values.
- v^{EV}: easy to compute, but a bound only under convexity.

Solving the problems

- y_1^{EV} is easy (one deterministic problem), but v^{EEV} needs a second-stage solve per scenario (or per drawn scenario, by Monte Carlo).
- v^{NA}: one deterministic problem under (FSN); otherwise specific methods (e.g. stochastic gradient).
- v^{SP}: stochastic-programming methods, of two main kinds:
  - Lagrangian decomposition (e.g. progressive hedging);
  - Benders decomposition (e.g. the L-shaped method, nested decomposition).
- v^{PI}: one deterministic problem per scenario; with too many, Monte Carlo: solve for each drawn ξ^n and average the values.
- v^{EV}: easy to compute, but a bound only under convexity.

Solving the problems

- y_1^{EV} is easy (one deterministic problem), but v^{EEV} needs a second-stage solve per scenario (or per drawn scenario, by Monte Carlo).
- v^{NA}: one deterministic problem under (FSN); otherwise specific methods (e.g. stochastic gradient).
- v^{SP}: stochastic-programming methods, of two main kinds:
  - Lagrangian decomposition (e.g. progressive hedging);
  - Benders decomposition (e.g. the L-shaped method, nested decomposition).
- v^{PI}: one deterministic problem per scenario; with too many, Monte Carlo: solve for each drawn ξ^n and average the values.
- v^{EV}: easy to compute, but a bound only under convexity.

Solving the problems

- y_1^{EV} is easy (one deterministic problem), but v^{EEV} needs a second-stage solve per scenario (or per drawn scenario, by Monte Carlo).
- v^{NA}: one deterministic problem under (FSN); otherwise specific methods (e.g. stochastic gradient).
- v^{SP}: stochastic-programming methods, of two main kinds:
  - Lagrangian decomposition (e.g. progressive hedging);
  - Benders decomposition (e.g. the L-shaped method, nested decomposition).
- v^{PI}: one deterministic problem per scenario; with too many, Monte Carlo: solve for each drawn ξ^n and average the values.
- v^{EV}: easy to compute, but a bound only under convexity.

Solving the problems

- y_1^{EV} is easy (one deterministic problem), but v^{EEV} needs a second-stage solve per scenario (or per drawn scenario, by Monte Carlo).
- v^{NA}: one deterministic problem under (FSN); otherwise specific methods (e.g. stochastic gradient).
- v^{SP}: stochastic-programming methods, of two main kinds:
  - Lagrangian decomposition (e.g. progressive hedging);
  - Benders decomposition (e.g. the L-shaped method, nested decomposition).
- v^{PI}: one deterministic problem per scenario; with too many, Monte Carlo: solve for each drawn ξ^n and average the values.
- v^{EV}: easy to compute, but a bound only under convexity.

Presentation Outline

1. Stochastic Programming Approach
   - One-stage Problems
   - Two-stage Problems
   - The recourse value function
2. Information and discretization
   - Choosing the information structure
   - The value of information
   - Sample Average Approximation

How to deal with continuous distributions?

Under finite support, the two-stage problem

$$\begin{aligned} \min_{y_1, \boldsymbol{y}_2} \quad & \mathbb{E}_{\mathbb{P}}\Big[\ell(y_1, \boldsymbol{\xi}, \boldsymbol{y}_2)\Big] \\ \text{s.t.} \quad & g(y_1, \boldsymbol{\xi}, \boldsymbol{y}_2) \leq 0, \qquad \mathbb{P}\text{-a.s.} \\ & \boldsymbol{y}_2 \preceq \boldsymbol{\xi} \end{aligned}$$

becomes

$$\begin{aligned} \min_{y_1, (y_2^\omega)_{\omega \in \Omega}} \quad & \sum_{\omega \in \Omega} \pi^\omega \ell(y_1, \xi^\omega, y_2^\omega) \\ \text{s.t.} \quad & g(y_1, \xi^\omega, y_2^\omega) \leq 0, \qquad \forall \omega \in \Omega. \end{aligned}$$

For a continuous law (e.g. Gaussian), the extensive formulation would need an infinite number of recourse variables, one per outcome.

Simplest idea: sample ξ

First-stage problem: $v^{\mathrm{SP}} = \inf_{y_1 \in \mathcal{Y}_1} J(y_1)$, with

$$J(y_1) := \ell_1(y_1) + \mathbb{E}_{\mathbb{P}}\left[\dot{V}_2(y_1, \xi)\right] .$$

- Draw independent copies $(\xi^n)_{n\in[N]}$ of ξ, of empirical distribution $\widehat{\mathbb{P}}_N := \frac{1}{N}\sum_{n=1}^N \delta_{\xi^n}$.
- Replacing $\mathbb{P}$ by $\widehat{\mathbb{P}}_N$ gives an (FSN) problem, with N scenarios of probability $1/N$, solved by its extensive formulation: the sample average approximation

$$\widehat{v}_N := \inf_{y_1 \in \mathcal{Y}_1} \widehat{J}_N(y_1) , \qquad \widehat{J}_N(y_1) := \ell_1(y_1) + \frac{1}{N}\sum_{n=1}^N \dot{V}_2(y_1, \xi^n) ,$$

  with $\widehat{y}_1^N$ an optimal solution.
- The sample has law $\mathbb{P}^{\otimes N}$, so $\widehat{\mathbb{P}}_N$, $\widehat{J}_N$, $\widehat{v}_N$ and $\widehat{y}_1^N$ are random; $\mathbb{E}_{\mathbb{P}^{\otimes N}}$ is the expectation over the draw.

Estimating the cost of a fixed decision

- For y_1 fixed *before* the draw, each ξ^n has the law of ξ, so

$$\mathbb{E}_{\mathbb{P}^{\otimes N}}\left[\widehat{J}_N(y_1)\right] = J(y_1) \qquad \text{for every } y_1 \in \mathcal{Y}_1 \text{ with } J(y_1) \in \mathbb{R}\,:$$

  $\widehat{J}_N(y_1)$ is an unbiased estimator of $J(y_1)$.

- Recall (CLT). If $(\boldsymbol{C}_n)_{n\in\mathbb{N}}$ are independent, identically distributed, with finite variance, then, in law,

$$\sqrt{N}\Big(\frac{1}{N}\sum_{n=1}^{N}\boldsymbol{C}_n - \mathbb{E}[\boldsymbol{C}_1]\Big) \Longrightarrow \mathcal{N}\big(0, \mathrm{Var}[\boldsymbol{C}_1]\big)\,.$$

Estimating the cost of a fixed decision: confidence interval

- With $\boldsymbol{C}_n := \dot{V}_2(y_1, \xi^n)$, their sample mean $\bar{V}_N := \frac{1}{N}\sum_{n=1}^N \boldsymbol{C}_n$ and empirical variance $\hat{\sigma}^2 := \frac{1}{N-1}\sum_{n=1}^N \left(\boldsymbol{C}_n - \bar{V}_N\right)^2$,

$$\mathbb{P}^{\otimes N}\left(J(y_1) \in \left[\widehat{J}_N(y_1) - \frac{z\,\hat{\sigma}}{\sqrt{N}},\ \widehat{J}_N(y_1) + \frac{z\,\hat{\sigma}}{\sqrt{N}}\right]\right) \longrightarrow 1-\alpha\,,$$

  with z the $(1-\alpha/2)$-quantile of $\mathcal{N}(0,1)$ ($z = 1.96$ at 95%): the Monte Carlo estimate, of width $\propto 1/\sqrt{N}$.

- A decision obtained *from* the sample depends on it: unbiasedness does not apply.

Biased estimator

Theorem (Downward and decreasing bias)

Let $\mathcal{Y}_1$ be nonempty, let $(\boldsymbol{\xi}^n)_{n\in\mathbb{N}}$ be independent copies of $\boldsymbol{\xi}$, and assume $\widehat{v}_N$ integrable for every N. Then, for every N,

$$\mathbb{E}_{\mathbb{P}^{\otimes N}}\left[\widehat{v}_N\right] \;\leq\; \mathbb{E}_{\mathbb{P}^{\otimes(N+1)}}\left[\widehat{v}_{N+1}\right] \;\leq\; v^{\mathrm{SP}}\,.$$

Proof of the bound:

- Fix $y_1 \in \mathcal{Y}_1$: $\widehat{J}_N(y_1) \geq \widehat{v}_N$ for every draw.
- Take expectations; $\widehat{J}_N(y_1)$ is unbiased:

$$J(y_1) = \mathbb{E}_{\mathbb{P}^{\otimes N}}\left[\widehat{J}_N(y_1)\right] \geq \mathbb{E}_{\mathbb{P}^{\otimes N}}\left[\widehat{v}_N\right]\,.$$

- Take the infimum over $y_1 \in \mathcal{Y}_1$: $v^{\mathrm{SP}} \geq \mathbb{E}_{\mathbb{P}^{\otimes N}}\left[\widehat{v}_N\right]$.

Biased estimator

Theorem (Downward and decreasing bias)

Let $\mathcal{Y}_1$ be nonempty, let $(\boldsymbol{\xi}^n)_{n\in\mathbb{N}}$ be independent copies of $\boldsymbol{\xi}$, and assume $\widehat{v}_N$ integrable for every N. Then, for every N,

$$\mathbb{E}_{\mathbb{P}^{\otimes N}}\left[\widehat{v}_N\right] \;\leq\; \mathbb{E}_{\mathbb{P}^{\otimes(N+1)}}\left[\widehat{v}_{N+1}\right] \;\leq\; v^{\mathrm{SP}} \,.$$

Proof of the bound:

- Fix $y_1 \in \mathcal{Y}_1$: $\widehat{J}_N(y_1) \geq \widehat{v}_N$ for every draw.
- Take expectations; $\widehat{J}_N(y_1)$ is unbiased:

$$J(y_1) = \mathbb{E}_{\mathbb{P}^{\otimes N}}\left[\widehat{J}_N(y_1)\right] \geq \mathbb{E}_{\mathbb{P}^{\otimes N}}\left[\widehat{v}_N\right] \,.$$

- Take the infimum over $y_1 \in \mathcal{Y}_1$: $v^{\mathrm{SP}} \geq \mathbb{E}_{\mathbb{P}^{\otimes N}}\left[\widehat{v}_N\right]$.

Decreasing bias

Let $\widehat{J}_N^{-n}$ be the sampled objective on the N draws $(\boldsymbol{\xi}^k)_{k\in[N+1],\, k\neq n}$. Each draw appears in N of them, so

$$\widehat{J}_{N+1}(y_1) = \frac{1}{N+1}\sum_{n=1}^{N+1} \widehat{J}_N^{-n}(y_1) \qquad \text{for every } y_1 \ .$$

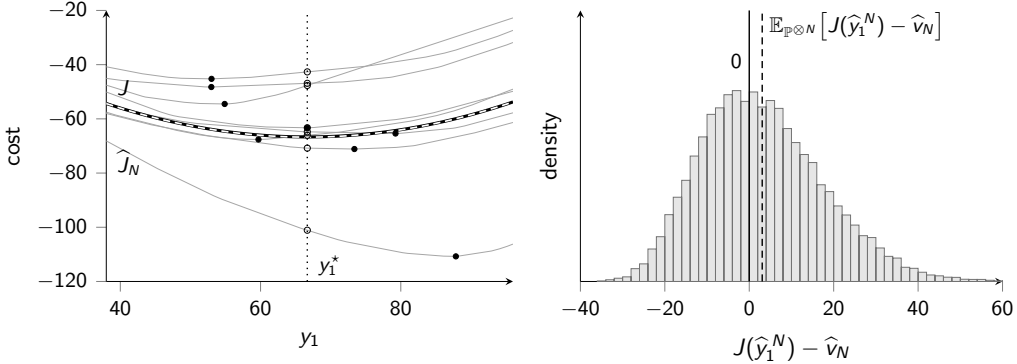
The infimum of an average is at least the average of the infima:

$$\widehat{v}_{N+1} = \inf_{y_1\in\mathcal{Y}_1} \frac{1}{N+1}\sum_{n=1}^{N+1} \widehat{J}_N^{-n}(y_1) \geq \frac{1}{N+1}\sum_{n=1}^{N+1} \inf_{y_1\in\mathcal{Y}_1} \widehat{J}_N^{-n}(y_1) \ .$$

Each $\widehat{J}_N^{-n}$ uses N independent copies of $\boldsymbol{\xi}$, so its infimum has the law of $\widehat{v}_N$; taking expectations,

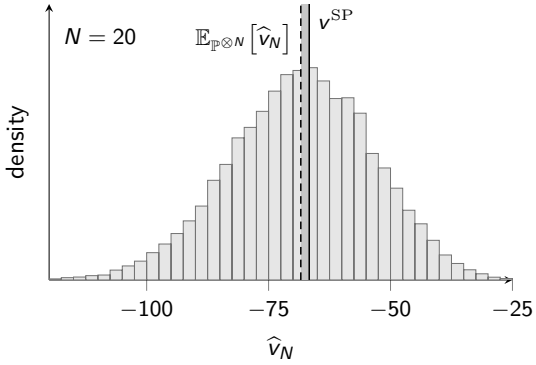
$$\mathbb{E}_{\mathbb{P}^{\otimes(N+1)}}\left[\widehat{v}_{N+1}\right] \geq \frac{1}{N+1}\sum_{n=1}^{N+1} \mathbb{E}_{\mathbb{P}^{\otimes N}}\left[\widehat{v}_N\right] = \mathbb{E}_{\mathbb{P}^{\otimes N}}\left[\widehat{v}_N\right] \ .$$

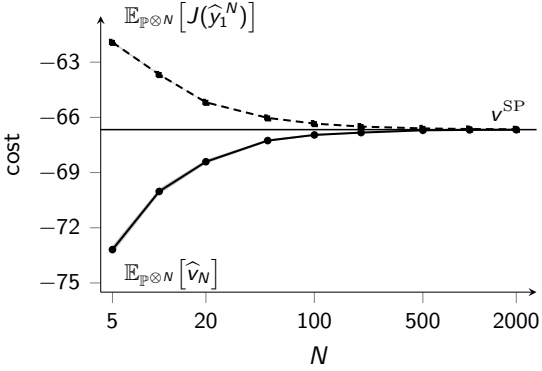
Why the minimization biases the value



Newsvendor, $c = 1$, $p = 3$, demand uniform on $[0, 100]$, $N = 20$. *Left*: eight sampled objectives $\widehat{J}_N$ (grey), each dotted at its minimum $(\widehat{y}_1^N, \widehat{v}_N)$, and their expectation J (black); each dot lies at or below its curve's value at $y_1^\star = 200/3$ (dotted). *Right*: $J(\widehat{y}_1^N) - \widehat{v}_N$ over 20 000 runs: positive on average, yet only in 55% of runs.

The sampled value is random, and optimistic on average





Same instance, $v^{\mathrm{SP}} = -200/3$, 20 000 samples. *Left*: $\widehat{v}_N$ at $N = 20$, with v^{SP} (solid), $\mathbb{E}_{\mathbb{P}^{\otimes N}}\left[\widehat{v}_N\right]$ (dashed) and the bias shaded: the spread (standard deviation 14.7) dwarfs the bias (1.7). *Right*: as N grows, $\mathbb{E}_{\mathbb{P}^{\otimes N}}\left[\widehat{v}_N\right]$ rises to v^{SP}; above it, the expected true cost $\mathbb{E}_{\mathbb{P}^{\otimes N}}\left[J(\widehat{y}_1{}^N)\right]$ of the sampled decision.

What you must remember

- Start from the timing: for each decision, say whether it is first or second stage. Write the extensive formulation and give it to a solver first.
- Once you have solved your problem, compute the VSS.
- Never report the value of a sampled problem as the value of its solution: judge every candidate on scenarios the optimizer never saw.

What you should remember

- A recourse decision is a function of the observation. Under finite support the extensive formulation is deterministic, its scenario blocks coupled only through the first-stage decision.
- Under relatively complete recourse every induced constraint is already implied by the first-stage constraints.
- More information relaxes the problem: $v^{\mathrm{PI}} \leq v^{\mathrm{SP}} \leq v^{\mathrm{NA}}$. When v^{SP} is finite, EVPI and VSS are nonnegative, and measure the data and the model.
- The value of a sampled problem is optimistic on average, less so as the sample grows.

Questions?



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