Stochastic Programming: A tutorial – Part I DORS Tutorials 14/02/2023

Giovanni Pantuso

Department of Mathematical Sciences University of Copenhagen Copenhagen, Denmark gp@math.ku.dk

Table of Contents

A promise

General formulations

Two-stage problems

A (very special) two-stage case

A closer look

A closer look at ξ

A closer look: continuous distributions

A closer look: discrete distributions

Approximations

Probability Metrics

Property Matching

Monte Carlo

Wrapping up

Mathematical properties

Bibliography

How to formulate general stochastic programs

How to formulate general stochastic programs

What makes them difficult

How to formulate general stochastic programs

What makes them difficult

How to formulate general stochastic programs:two-stage linear (mixed-integer) recourse problems

What makes them difficult

How to formulate general stochastic programs:two-stage linear (mixed-integer) recourse problems for a risk neutral decision maker

What makes them difficult

How to formulate general stochastic programs:two-stage linear (mixed-integer) recourse problems for a risk neutral decision maker

What makes them difficult and how to address the difficulty

How to formulate general stochastic programs:two-stage linear (mixed-integer) recourse problems for a risk neutral decision maker

What makes them difficult and how to address the difficulty

How to solve (approximations of) two-stage stochastic programs

Limitations ...

- ► Risk-aversion
- ► Chance constraints
- ► Stochastic dominance
- ► Multi-stage problems
- ► Endogenous uncertainty
- ► Robust/Distributionally robust
- **>** ...

Plan

- ► Introduction (45 min)
- ▶ break (15 min)
- ► Exercise 1 (10 min)
- ► More info on approximations (15 min)
- ► Exercise 2 (30 min)
- ▶ break (10 min)
- ► L-Shaped method (45 min)
- ► Exercise 3 (50 min)

Table of Contents

A promise

General formulations

Two-stage problems

A (very special) two-stage case

A closer look

A closer look at ξ

A closer look: continuous distributions

A closer look: discrete distributions

Approximations

Probability Metrics

Property Matching

Monte Carlo

Wrapping up

Mathematical properties

Bibliography

Uncertainty

Something is uncertain

$$(\Omega,\mathcal{F},\mathbb{P})$$

Uncertainty

Something is uncertain

$$(\Omega,\mathcal{F},\mathbb{P})$$

$$\omega \in \Omega \to \xi(\omega) \in \Xi$$

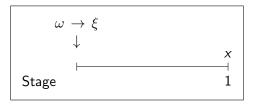
Uncertainty

Something is uncertain

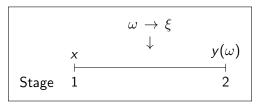
$$(\Omega,\mathcal{F},\mathbb{P})$$

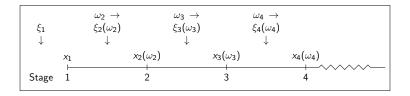
$$\omega \in \Omega \to \xi(\omega) \in \Xi$$

Notation can vary









The decision maker:

1. Makes decisions $x \in \mathcal{X} \subseteq \mathbb{R}^{n_1}$ (first decision stage)

The decision maker:

- 1. Makes decisions $x \in \mathcal{X} \subseteq \mathbb{R}^{n_1}$ (first decision stage)
- 2. Waits for the outcome $\omega \in \Omega$ of some random experiment.
- 3. ω determines $\boldsymbol{\xi}(\omega)$ (our random data)

The decision maker:

- 1. Makes decisions $x \in \mathcal{X} \subseteq \mathbb{R}^{n_1}$ (first decision stage)
- 2. Waits for the outcome $\omega \in \Omega$ of some random experiment.
- 3. ω determines $\xi(\omega)$ (our random data)
- 4. Makes decisions $y(\omega) \in \mathcal{Y} \subseteq \mathbb{R}^{n_2}$, given ξ and x (second decision stage)

min
$$c^{\top}x + \mathbb{E}_{\boldsymbol{\xi}}[\min \boldsymbol{q}(\omega)^{T}y(\omega)]$$

$$\min z = c^{\top} x + \mathbb{E}_{\xi}[\min \boldsymbol{q}(\omega)^{T} y(\omega)]$$

s.t. $Ax = b$

$$\begin{aligned} \min z &= c^\top \ x + \mathbb{E}_{\pmb{\xi}}[\min \pmb{q}(\omega)^T y(\omega)] \\ \text{s.t. } Ax &= b \\ \pmb{T}(\omega)x + \pmb{W}(\omega)y(\omega) &= \pmb{h}(\omega) \ \ a.s. \\ x &\in \mathcal{X}, y(\omega) \in \mathcal{Y} \end{aligned}$$

Parameters $c \in \mathbb{R}^{n_1}$, $b \in \mathbb{R}^{m_1}$, and $A \in \mathbb{R}^{n_1 \times m_1}$ are known.

Parameters $c \in \mathbb{R}^{n_1}$, $b \in \mathbb{R}^{m_1}$, and $A \in \mathbb{R}^{n_1 \times m_1}$ are known.

Parameters $\boldsymbol{q}(\omega) \in \mathbb{R}^{n_2}$, $\boldsymbol{h}(\omega) \in \mathbb{R}^{m_2}$, $\boldsymbol{W}(\omega) \in \mathbb{R}^{m_2 \times n_2}$ and $\boldsymbol{T}(\omega) \in \mathbb{R}^{m_2 \times n_1}$ are uncertain.

Parameters $c \in \mathbb{R}^{n_1}$, $b \in \mathbb{R}^{m_1}$, and $A \in \mathbb{R}^{n_1 \times m_1}$ are known.

Parameters $\boldsymbol{q}(\omega) \in \mathbb{R}^{n_2}$, $\boldsymbol{h}(\omega) \in \mathbb{R}^{m_2}$, $\boldsymbol{W}(\omega) \in \mathbb{R}^{m_2 \times n_2}$ and $\boldsymbol{T}(\omega) \in \mathbb{R}^{m_2 \times n_1}$ are uncertain.

$$\boldsymbol{\xi}(\omega) = \left(\boldsymbol{q}(\omega)^\top, \boldsymbol{h}(\omega)^\top, \ \boldsymbol{W}^{1}(\omega), \ldots, \boldsymbol{W}(\omega)^{m_2}, \boldsymbol{T}(\omega)^1, \ldots, \boldsymbol{T}(\omega)^{m_2}\right).$$

Parameters $c \in \mathbb{R}^{n_1}$, $b \in \mathbb{R}^{m_1}$, and $A \in \mathbb{R}^{n_1 \times m_1}$ are known.

Parameters $\boldsymbol{q}(\omega) \in \mathbb{R}^{n_2}$, $\boldsymbol{h}(\omega) \in \mathbb{R}^{m_2}$, $\boldsymbol{W}(\omega) \in \mathbb{R}^{m_2 \times n_2}$ and $\boldsymbol{T}(\omega) \in \mathbb{R}^{m_2 \times n_1}$ are uncertain.

$$\boldsymbol{\xi}(\omega) = \left(\boldsymbol{q}(\omega)^{\top}, \boldsymbol{h}(\omega)^{\top}, \ \boldsymbol{W}^{1}(\omega), \ldots, \boldsymbol{W}(\omega)^{m_{2}}, \boldsymbol{T}(\omega)^{1}, \ldots, \boldsymbol{T}(\omega)^{m_{2}}\right).$$

 ξ is a realization of $\xi(\omega)$.

min
$$c^{\top}x + Q(x)$$

s.t. $Ax = b$
 $x \in \mathcal{X}$

min
$$c^{\top}x + Q(x)$$

s.t. $Ax = b$
 $x \in \mathcal{X}$

where

$$Q(x) = \mathbb{E}_{\boldsymbol{\xi}}[Q(x,\xi)]$$

$$\min c^{\top}x + Q(x)$$

s.t. $Ax = b$
 $x \in \mathcal{X}$

where

$$Q(x) = \mathbb{E}_{\boldsymbol{\xi}}[Q(x,\xi)]$$

$$Q(x,\xi) = \min_{y} \{ q^{\top} y | Wy = h - Tx, y \in \mathcal{Y} \}.$$

Consider $\Xi = \{\xi_1, \dots, \xi_S\}$ with probabilities π_s , $s = 1 \dots, S$.

Consider $\Xi = \{\xi_1, \dots, \xi_S\}$ with probabilities π_s , $s = 1 \dots, S$.

$$\xi_s \implies q_s, T_s, W_s, h_s.$$

Consider $\Xi = \{\xi_1, \dots, \xi_S\}$ with probabilities π_s , $s = 1 \dots, S$.

$$\xi_s \implies q_s, T_s, W_s, h_s.$$

$$y(\omega)$$
 becomes y_1, \ldots, y_S .

min
$$c^{\top}x + \sum_{s=1}^{S} \pi_s q_s^{\top} y_s$$

s.t. $Ax = b$
 $T_s x + W_s y_s = h_s$ $s = 1, \dots, S$
 $x \in \mathcal{X}$
 $y_s \in \mathcal{Y}$ $s = 1, \dots, S$.

Table of Contents

A promise

General formulations

Two-stage problems

A (very special) two-stage case

A closer look

A closer look at ξ

A closer look: continuous distributions

A closer look: discrete distributions

Approximations

Probability Metrics

Property Matching

Monte Carlo

Wrapping up

Mathematical properties

Bibliography

Where do we get ξ ?

Where do we get ξ ?

- ► Number of failures ≈ Weibull
- ► Wind speed ≈ Weibull,Rayleigh
- ► Forecast error (linear regression) ≈ Normal
- ▶ Hospitalization in certain epidemics \approx LogNormal
- ightharpoonup Repair times pprox LogNormal
- ightharpoonup Choice model pprox Gumbel, Normal, EV Type I
- ▶ Waiting times \approx Beta
- ▶ ...

Some randomness is discrete

Some randomness is discrete

- ► Number of occurrences ≈ Poisson
- ▶ Number of trials before success \approx Geometric
- ► Number of successes ≈ HyperGeometric
- ▶ ...

High dimensions and co-dependencies are problematic

A closer look at the constraints

If ξ is continuous...

A closer look at the constraints

If ξ is continuous...

Constraints must hold a.s. ...

A closer look at the constraints

If ξ is continuous...

Constraints must hold a.s. ...

Possibly ∞ constraints

The recourse function ...

$$Q(x) = \mathbb{E}_{\boldsymbol{\xi}} \big[Q(x, \xi) \big] = \int_{\Omega} Q(x, \xi(\omega)) \mathbb{P}(d\omega)$$

Why is this difficult?

The recourse function ...

$$Q(x) = \mathbb{E}_{\boldsymbol{\xi}} \big[Q(x, \xi) \big] = \int_{\Omega} Q(x, \xi(\omega)) \mathbb{P}(d\omega)$$

Why is this difficult?

Ingredient 1: a closed form expression $Q(x,\xi)$

The recourse function ...

$$Q(x) = \mathbb{E}_{\boldsymbol{\xi}} \big[Q(x, \xi) \big] = \int_{\Omega} Q(x, \xi(\omega)) \mathbb{P}(d\omega)$$

Why is this difficult?

Ingredient 1: a closed form expression $Q(x,\xi)$

Ingredient 2: an antiderivative

The recourse function ...

$$Q(x) = \mathbb{E}_{\boldsymbol{\xi}} \big[Q(x, \xi) \big] = \int_{\Omega} Q(x, \xi(\omega)) \mathbb{P}(d\omega)$$

Why is this difficult?

Ingredient 1: a closed form expression $Q(x,\xi)$

Ingredient 2: an antiderivative

Observe:
$$Q(x) = \int \int \cdots \int Q(x,\xi) \mathbb{D}(\xi) d\xi_1 d\xi_2 \cdots d\xi_N$$

The recourse function ...

$$Q(x) = \mathbb{E}_{\boldsymbol{\xi}}\big[Q(x,\xi)\big] = \int_{\Omega} Q(x,\xi(\omega))\mathbb{P}(d\omega)$$

Idea! Numerical integration!

In one dimension (i.e., ${\it N}=1$): Riemann sums, Trapezoidal rule, Simpson's rule

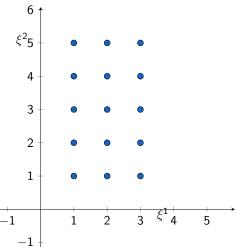
Ex. Riemann Sums

$$\int_{a}^{b} f(x) dx$$

- ▶ Partition [a, b] using K points $x_0 = a, x_1, ..., x_K = b$ equally spaced Δx
- $\blacktriangleright \int_a^b f(x) dx \approx \sum_k f(x_k) \Delta x$
- ► As *K* increases we improve the approximation.

In multiple dimensions: Quadrature methods.

Same principle, harder partition



 $\label{lem:numerical} \mbox{Numerical integration: it is already an approximation}$

Numerical integration: it is already an approximation all this work for one x...

Numerical integration: it is already an approximation

all this work for one x...

we still have to solve the stochastic program..

When ξ is discrete...

$$Q(x) = \sum_{s=1}^{S} \pi_s Q(x, \xi_s)$$

When ξ is discrete...

$$Q(x) = \sum_{s=1}^{S} \pi_s Q(x, \xi_s)$$

Finitely many linear constraints

When ξ is discrete...

$$Q(x) = \sum_{s=1}^{S} \pi_s Q(x, \xi_s)$$

Finitely many linear constraints

When ξ is discrete ... but large ...

Table of Contents

A promise

General formulations

Two-stage problems

A (very special) two-stage case

A closer look

A closer look at ξ

A closer look: continuous distributions

A closer look: discrete distributions

Approximations

Probability Metrics

Property Matching

Monte Carlo

Wrapping up

Mathematical properties

Bibliography

Approximations

Continuous/discrete but large \rightarrow discrete and small

Approximations

Continuous/discrete but large \rightarrow discrete and small

Three categories of methods (loosely speaking)

- ► Probability Metrics
- ► Property Matching
- ► Monte Carlo

Results from research on stability, see,e.g., [Dup90, Pfl01, RR02, Röm03].

$$|z(\mathbb{P})-z(\mathbb{Q})| \leq Ld(\mathbb{P},\mathbb{Q})$$

► Start from large *N* scenarios (e.g., sampled)

- ► Start from large *N* scenarios (e.g., sampled)
- ► Remove one scenario at a time to minimize the distance between the new and old distribution

- ► Start from large *N* scenarios (e.g., sampled)
- ► Remove one scenario at a time to minimize the distance between the new and old distribution
- ► Add the probability of the deleted scenarios to the closest scenarios (in the sense of the probability metric)

Scenario reduction/generation, see, e.g., [HR03, DGKR03].

Idea: Replicate only the statistical properties that are important for the problem [HW01].

Create a small distribution that replicates only those properties.

Not always necessary to increase the size of the distribution.

Not always necessary to increase the size of the distribution.

Problem driven.

Not always necessary to increase the size of the distribution.

Problem driven.

Observe: requires an NLP (heuristics exist [HKW03])

Not always necessary to increase the size of the distribution.

Problem driven.

Observe: requires an NLP (heuristics exist [HKW03])

Which properties?

Monte Carlo in a nutshell

Make K identical copies ξ_1, \ldots, ξ_K of ξ .

Monte Carlo in a nutshell

Make K identical copies ξ_1, \ldots, ξ_K of ξ .

From each take, independently, a realization ξ_k .

Monte Carlo in a nutshell

Make K identical copies ξ_1, \ldots, ξ_K of ξ .

From each take, independently, a realization ξ_k .

Write the Sample Average Approximation (SAA)

$$z^{K} = \min f(x) := c^{T} x + \sum_{k=1}^{K} \frac{1}{K} q_{k}^{T} y_{k}$$
s.t. $Ax = b$,
$$T_{k}x + W_{k}y_{k} = h_{k}, \qquad k = 1, \dots, K$$

$$x \in \mathcal{X},$$

$$y_{k} \in \mathcal{Y}, \qquad k = 1, \dots, K.$$

Sample Average Approximation (SAA)

$$z^{K} = \min f^{K}(x) := c^{T} x + \sum_{k=1}^{K} \frac{1}{K} q_{k}^{T} y_{k}$$
s.t. $Ax = b$,
$$T_{k}x + W_{k}y_{k} = h_{k}, \qquad k = 1, \dots, K$$

$$x \in \mathcal{X},$$

$$y_{k} \in \mathcal{Y}, \qquad k = 1, \dots, K.$$

or

$$z^{K} = \min f^{K}(x) := c^{T} x + \sum_{k=1}^{K} \frac{1}{K} Q(x, \xi_{k})$$

s.t. $Ax = b$,
 $x \in \mathcal{X}$

 z^K is a stochastic estimator of z^*

$$f^K(ar{x}) = c^{ op} \ ar{x} + \sum_{k=1}^K rac{1}{K} Q(ar{x}, \xi_k)$$

is an **unbiased** estimator (pointwise $x = \bar{x}$) of

$$f(\bar{x}) = c^{\top} \bar{x} + \mathbb{E}_{\boldsymbol{\xi}}[Q(\bar{x}, \xi)]$$

Finally, $f^K(\bar{x})$ a consistent estimator of $f(\bar{x})$, that is

$$\lim_{K o\infty}rac{1}{K}\sum_{k=1}^KQ(ar{x},\xi_k) o \mathbb{E}_{m{\xi}}\left[Q(ar{x},\xi)
ight]=Q(ar{x})$$

We say that $f^K(\bar{x})$ converges pointwise.

Finally, $f^K(\bar{x})$ a consistent estimator of $f(\bar{x})$, that is

$$\lim_{K o\infty}rac{1}{K}\sum_{k=1}^KQ(ar x,\xi_k) o \mathbb{E}_{m{\xi}}\left[Q(ar x,\xi)
ight]=Q(ar x)$$

We say that $f^K(\bar{x})$ converges pointwise.

But observe

$$\mathbb{S}td[rac{1}{\mathcal{K}}\sum_{k=1}^{\mathcal{K}}Q(ar{x},\xi_k)]=rac{\mathbb{S}td[Q(ar{x},\xi)]}{\sqrt{\mathcal{K}}}$$

Exercise 1: 15 minutes

https://tinyurl.com/sptutorial1

Under certain conditions $z^K \to z^*$ as $k \to \infty$ (exponentially fast!)

Under certain conditions $z^K \to z^*$ as $k \to \infty$ (exponentially fast!)

 $\mathbb{E}[z^K]$ gives a statistical lower bound! (obs! z^K is biased)

Proof

Under certain conditions $z^K \to z^*$ as $k \to \infty$ (exponentially fast!)

 $\mathbb{E}[z^K]$ gives a statistical lower bound! (obs! z^K is biased)

Proof

$$\mathbb{E} \left[c^{\top} \bar{x} + \frac{1}{K} \sum_{k=1}^{K} [Q(\bar{x}, \xi_k)] \right]$$
 gives a statistical upper bound!

See, e.g., [Sha91, MMW99, Sha03].

Exercise 2: 30 minutes

https://tinyurl.com/sptutorial2

Approximations

Getting a good solution vs Estimating its value $% \left(1\right) =\left(1\right) \left(1\right$

Table of Contents

A promise

General formulations

Two-stage problems

A (very special) two-stage case

A closer look

A closer look at ξ

A closer look: continuous distributions

A closer look: discrete distributions

Approximations

Probability Metrics

Property Matching

Monte Carlo

Wrapping up

Mathematical properties

Bibliography

$$\mathcal{K}_2(\xi) = \{x | \exists y \ge 0, \text{s.t.} W(\omega)y = h(\omega) - T(\omega)x\}$$

$$\mathcal{K}_2(\xi) = \{x | \exists y \geq 0, \text{s.t.} W(\omega)y = h(\omega) - T(\omega)x\}$$

Convex and polyhedral!

$$\mathcal{K}_2 = \bigcap_{\xi \in \Xi} \mathcal{K}_2(\xi)$$

$$\mathcal{K}_2 = \bigcap_{\xi \in \Xi} \mathcal{K}_2(\xi)$$

Convex and polyhedral!

Useful jargon

Complete recourse

$$\mathcal{K}_2 = \mathbb{R}^{n_1}$$

Useful jargon

Complete recourse

$$\mathcal{K}_2 = \mathbb{R}^{n_1}$$

Relatively complete recourse

$$\mathcal{K}_2 \subseteq \mathcal{K}_1 = \{x | Ax = b, x \geq 0\}$$

.

$$Q(x,\xi)$$
 is:

- a. piece-wise linear convex in h, T and x,
- b. piece-wise linear concave in q.

$$Q(x) = \mathbb{E}_{\boldsymbol{\xi}} Q(x, \xi) = \sum_{s=1}^{S} \pi_s Q(x, \xi_s)$$

piece-wise linear convex in x.

Proof of lower bound

The expectation of the optimal objective value of SAAs of size K is a lower bound for the true optimal objective value, that is:

$$\mathbb{E}[z^K] \leq z^*$$

Proof.

First, observe that

 $z^* = \min_{x \in \mathcal{K}_1} f(x) = \min_{x \in \mathcal{K}_1} c^\top x + \mathbb{E}\left[K^{-1} \sum_{k=1}^K Q(x, \xi_k)\right]$. We can use a direct proof. We know that for all $x' \in \mathcal{K}_1$

$$\min_{x \in \mathcal{K}_1} K^{-1} \sum_{k=1}^K c^\top x + Q(x, \xi_k) \le K^{-1} \sum_{k=1}^K c^\top x' + Q(x', \xi_k)$$

by taking the expectation of both sides we obtain continues...

Proof of lower bound

The expectation of the optimal objective value of SAAs of size K is a lower bound for the true optimal objective value, that is:

$$\mathbb{E}[z^K] \le z^*$$

Proof.

continuing...

$$\mathbb{E}\left[\min_{x \in \mathcal{K}_1} K^{-1} \sum_{k=1}^K c^\top x + Q(x, \xi_k)\right] \leq \mathbb{E}\left[K^{-1} \sum_{k=1}^K c^\top x' + Q(x', \xi_k)\right]$$

which corresponds to

$$\mathbb{E}\left[z^{K}\right] \leq \mathbb{E}\left[K^{-1}\sum_{k=1}^{K}c^{\top}x' + Q(x', \xi_{k})\right]$$

continues...



Proof of lower bound

The expectation of the optimal objective value of SAAs of size K is a lower bound for the true optimal objective value, that is:

$$\mathbb{E}[z^K] \le z^*$$

Proof.

continuing... Now, since the inequality holds for all x', it holds also for the solution x' that minimizes expectation on the right-hand-side and which yields z^* , that is

$$\mathbb{E}\left[z^{K}\right] \leq \min_{x' \in \mathcal{K}_{1}} \mathbb{E}\left[K^{-1} \sum_{k=1}^{K} c^{\top} x' + Q(x', \xi_{k})\right] = z^{*}$$

Table of Contents

A promise

General formulations

Two-stage problems

A (very special) two-stage case

A closer look

A closer look at ξ

A closer look: continuous distributions

A closer look: discrete distributions

Approximations

Probability Metrics

Property Matching

Monte Carlo

Wrapping up

Mathematical properties

Bibliography

References I

- [DGKR03] Jitka Dupačová, Nicole Gröwe-Kuska, and Werner Römisch. Scenario reduction in stochastic programming. *Mathematical programming*, 95(3):493–511, 2003.
 - [Dup90] Jitka Dupačová. Stability and sensitivity-analysis for stochastic programming. *Annals of operations research*, 27(1):115–142, 1990.
 - [HKW03] Kjetil Høyland, Michal Kaut, and Stein W Wallace. A heuristic for moment-matching scenario generation. Computational optimization and applications, 24(2):169–185, 2003.
 - [HR03] Holger Heitsch and Werner Römisch. Scenario reduction algorithms in stochastic programming. *Computational optimization and applications*, 24(2):187–206, 2003.

References II

- [HW01] Kjetil Høyland and Stein W Wallace. Generating scenario trees for multistage decision problems. Management science, 47(2):295–307, 2001.
- [MMW99] Wai-Kei Mak, David P Morton, and R Kevin Wood. Monte carlo bounding techniques for determining solution quality in stochastic programs. *Operations* research letters, 24(1-2):47–56, 1999.
 - [PfI01] G Ch Pflug. Scenario tree generation for multiperiod financial optimization by optimal discretization. *Mathematical programming*, 89(2):251–271, 2001.
 - [Röm03] Werner Römisch. Stability of stochastic programming problems. *Handbooks in operations research and management science*, 10:483–554, 2003.

References III

- [RR02] Svetlozar T Rachev and Werner Römisch. Quantitative stability in stochastic programming: The method of probability metrics. *Mathematics of Operations Research*, 27(4):792–818, 2002.
- [Sha91] Alexander Shapiro. Asymptotic analysis of stochastic programs. *Annals of Operations Research*, 30(1):169–186, 1991.
- [Sha03] Alexander Shapiro. Monte carlo sampling methods. In Stochastic Programming, volume 10 of Handbooks in Operations Research and Management Science, pages 353–425. Elsevier, 2003.