Data-driven distributionally robust optimization with Wasserstein metric, moment conditions and robust constraints

Adrián Esteban-Pérez, Juan Miguel Morales

adrianesteban@uma.es

juan.morales@uma.es

Department of Applied Mathematics
University of Malaga
OASYS GROUP
oasys.uma.es

EURO 2018.
29th European Conference on Operational Research.
July 11th, 2018, Valencia (Spain)

Optimization and Analytics for Sustainable energY Systems

European Research Council
Established by the European Commission

Distributionally Robust Optimization

Distributionally Robust Optimization (DRO). Introduction

- Stochastic Programming: $\inf_{\mathbf{x} \in X} \mathbb{E}_Q f(\mathbf{x}, \boldsymbol{\xi})$
- Robust Optimization: $\inf_{\mathbf{x} \in X} \sup_{\boldsymbol{\xi} \in \Xi} f(\mathbf{x}, \boldsymbol{\xi})$
- DRO is essentially Stochastic Programming + Robust Optimization.

Data-Driven Distributionally Robust Optimization

Data-Driven Distributionally Robust Optimization (DDRO)

- Input: Training samples: $\widehat{\boldsymbol{\xi}}_1, \ldots, \widehat{\boldsymbol{\xi}}_N$
- Construct a set of probability distributions $\mathcal{Q}_N$ using the traininig samples (ambiguity set)

How to construct an ambiguity set?

- Seek probability distributions *close* to the empirical distribution $\widehat{P}$ based on the training samples.
- How close?
- We use probability metrics. A usually choice is Wasserstein metric.

Data-Driven Distributionally Robust Optimization

The goal is to compute:

$$\inf_{\mathbf{x} \in X} \sup_{Q \in \mathcal{Q}_N} \mathbb{E}_Q f(\mathbf{x}, \boldsymbol{\xi})$$

and the optimal solution $\mathbf{x}^*$.

DDRO

Wasserstein metric of order p

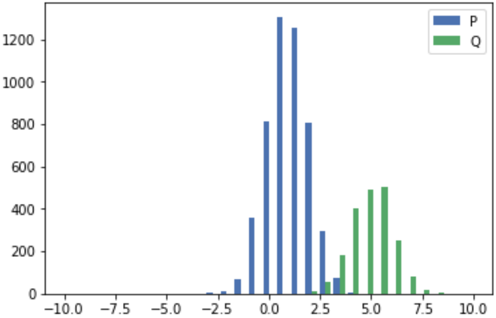
Wasserstein metric (of order p) between two probability distributions P and Q:

$$W_p(P, Q) = \left(\inf_{(X,Y): X \rightsquigarrow P, Y \rightsquigarrow Q} \mathbb{E}(\|X - Y\|^p) \right)^{1/p}$$

DDRO

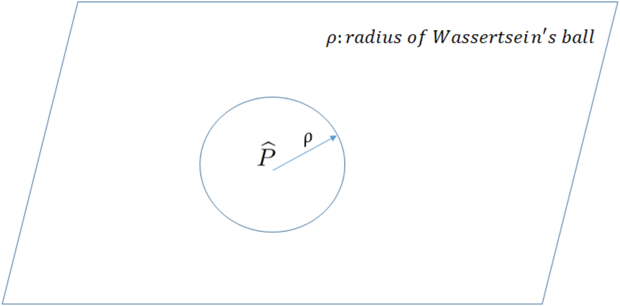
Wasserstein metric

Equivalently, Wasserstein metric between two probability distributions P and Q, $W(P, Q)$, is the minimum cost of moving P to Q. In the discrete case in 1-D:



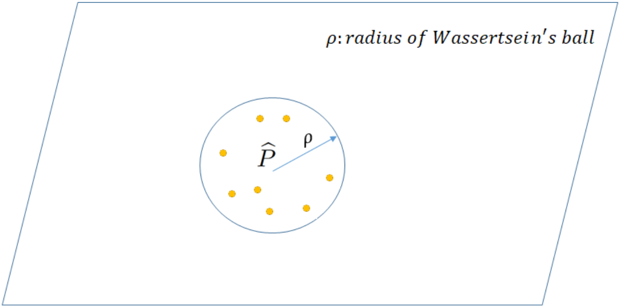
DDRO

- A problem: DDRO with Wasserstein's metric is too conservative:



DDRO

- A problem: DDRO with Wasserstein's metric is too conservative:
- A common approach to solving this problem is to add a priori information!!



DDRO

Wassersein metric

Wasserstein metric paradigm

- *Advantages*: Good theoretical properties: E.g.: Convergence with respect to Wasserstein metric (of order p) is equivalent to the usual weak convergence of measures plus convergence of the first p-th moments, rates of convergence of the empirical distribution to the true distribution.
- *Disadvantages*: We get too conservative distributions.
- Idea behind this approach: *the mass of the empirical distribution is moved to the worst case location points with the worst case mass in such points.*

DDRO

Some remarks about DDRO

A brief scheme

- Formulate the problem and the ambiguity set.
- Reformulate the inner supremum problem in a nice form using duality arguments in order to join it with the infimum outer problem.
- We get a minimization problem with a constraint of the form:

$$\sup_{\boldsymbol{\xi}} F(\mathbf{x}, \boldsymbol{\xi}) \leqslant 0$$

- So, the assumptions of ambiguity set and the objective function are essential!!

DDRO

Remarks and commom assumptions in DDRO

Reformulation of robust constraints in DDRO paradigm

Nowadays, sometimes

$$\sup_{\boldsymbol{\xi} \in \Xi} F(\mathbf{x}, \boldsymbol{\xi}) \leqslant 0$$

can be reformulated in a nice form applying the results existing in:

- *Deriving robust counterparts of nonlinear uncertain inequalities*, Ben-Tal et al. (2015), Mathematical Programming.

Motivation

> *A Wasserstein ball around the empirical distribution includes distributions with different support and allows (in a sense) robustness to unseen data.*

(Sinha et al. (2018), *Certifying Some Distributional Robustness with Principled Adversarial Training*)

- Thus, we consider a Wasserstein's ball.
- Using an ambiguity set, our goal is to find *good* hidden distributions (distributions *closer* to the true distributions with similar features) which reflects the random phenomena of our model.
- How do we do?
- We consider conic constraints in order to add a priori shape information.

Our approach

We split the support set in K regions and we introduce K decision variables (the mass in each region) subject to:

- $\sum_{i=1}^{K} p_i = 1$ and $p_i \geqslant 0$
- The array $(p_i)_{i=1}^{K}$ is in a cone $\mathcal{C}$ which reflects the shape of the distribution.

An example: The Univariate Newsvendor problem

DDRO. Formulation and notation

- x: order quantity (decision variable).
- ξ: demand of the item (random variable).
- h: unit holding cost.
- b: unit backorder cost.

An example: The Univariate Newsvendor problem

DDRO. Formulation and notation

- x: order quantity (decision variable).
- ξ: demand of the item (random variable).
- h: unit holding cost.
- b: unit backorder cost.

An example: The Univariate Newsvendor problem

DDRO. Formulation and notation

- x: order quantity (decision variable).
- ξ: demand of the item (random variable).
- h: unit holding cost.
- b: unit backorder cost.

An example: The Univariate Newsvendor problem

DDRO. Formulation and notation

- x: order quantity (decision variable).
- ξ: demand of the item (random variable).
- h: unit holding cost.
- b: unit backorder cost.

An example: The Univariate Newsvendor problem

DDRO. Formulation and notation

- x: order quantity (decision variable).
- ξ: demand of the item (random variable).
- h: unit holding cost.
- b: unit backorder cost.

An example: The Univariate Newsvendor problem

DDRO. Formulation and notation

- x: order quantity (decision variable).
- ξ: demand of the item (random variable).
- h: unit holding cost.
- b: unit backorder cost.
- $(x-\xi)^+$: quantity in stock, where $z^+ = \max(z, 0)$.
- $(\xi - x)^+$: shortage quantity.

Thus, the problem is the following:

$$\inf_{x \geqslant 0} \sup_{Q \in \mathcal{Q}_N} \mathbb{E}_Q h(x-\xi)^+ + b(\xi - x)^+$$

$\mathcal{Q}_N$ is the ambiguity set which is constructed using a Wasserstein ball (using the Wasserstein metric of order 1) and we consider the conic constraints approach presented before.

Numerical experiments

Data set and assumptions

Parameters of the model: $h = b = 20$.

Numerical experiments

Data set and assumptions

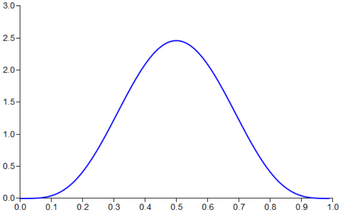
Parameters of the model: $h = b = 20$.
The true distribution:

Numerical experiments

Data set and assumptions

Parameters of the model: $h = b = 20$.
The true distribution:

- The probability distribution is a Beta distribution $B(5,5)$

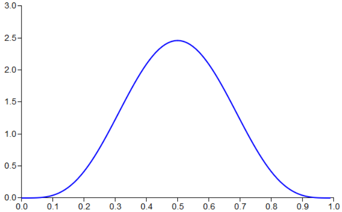


Numerical experiments

Data set and assumptions

Parameters of the model: $h = b = 20$.
The true distribution:

- The probability distribution is a Beta distribution $B(5, 5)$

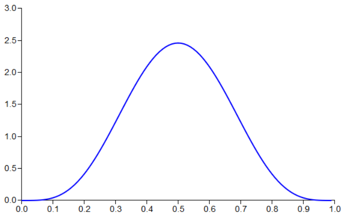


Numerical experiments

Data set and assumptions

Parameters of the model: $h = b = 20$.
The true distribution:

- The probability distribution is a Beta distribution $B(5, 5)$



- The probability distribution is *unimodal*.

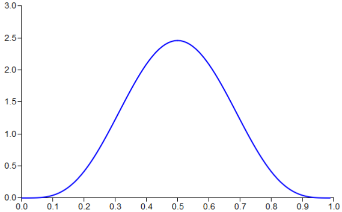
Numerical experiments

Data set and assumptions

Parameters of the model: $h = b = 20$.

The true distribution:

- The probability distribution is a Beta distribution $B(5, 5)$



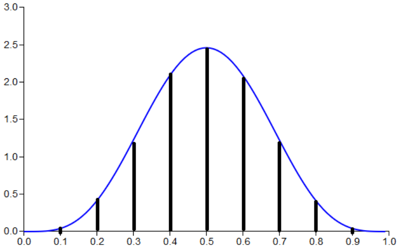
The decision-maker knows:

- The probability distribution is *unimodal*.
- The support set is the interval $[0, 1]$.

Numerical experiments

Data set

We construct K regions over the support set of the demand:

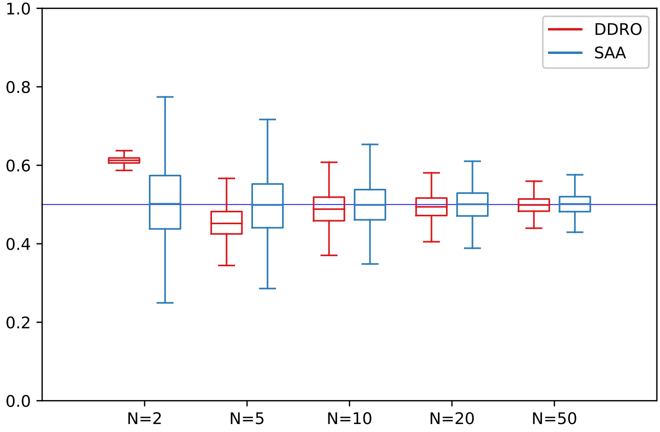


For each $i = 1, \ldots, K$, we assign a probability mass p_i to the i-th region. We consider the cone

$$\mathcal{C} = \{p \in \mathbb{R}^K : p_1 \leqslant p_2 \leqslant \ldots \leqslant p_m \geqslant p_{m+1} \geqslant \ldots \geqslant p_K\}$$

Numerical experiments

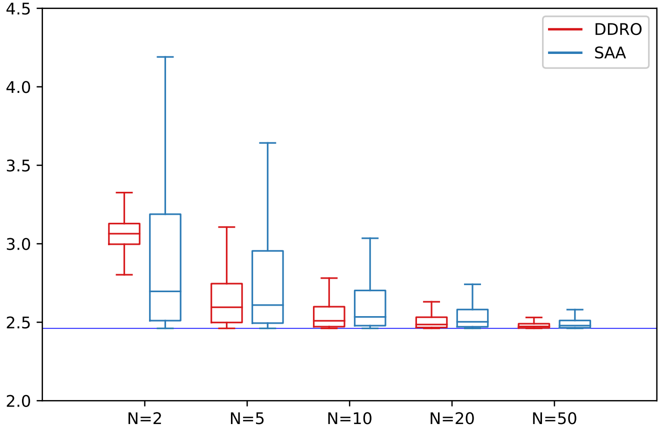
Order Quantity



Our DDRO approach obtains less variance

Numerical experiments

Actual Expected Cost



Our DDRO approach obtains less variance

Conclusions

- Shape information helps us to get better solutions than SAA method.

- Shape information is added in an easy way using conic constraints which becomes linear!!

Questions?

Thanks for the attention!

Supported by the project: FlexAnalytics - Advanced Analytics to Empower the Small Flexible Consumers of Electricity.

European Research Council

Established by the European Commission