Probability minimization using Decision-Dependent Distributionally Robust Optimization

Diego. F. Fonseca. V. Department of Mathematics, Universidad de los Andes, df.fonseca@uniandes.edu.co

Introduction

We analyze the problem of minimizing the probability that a function does not exceed a given level where this function depends on a random vector and the decision. We study this via Distributionally Robust Optimization (DRO) approach using Wasserstein metrics where the ambiguity set depends on the decision. In contrast to other approaches in which the set of ambiguity does not depend on the decision, which triggers mixed integer programming problems, we show that for certain functions our resulting problem can be convex.

Probability stochastic programming

A *probability stochastic program* is an optimization program given by

$$J^* = \min_{x\in\mathbb{X}} \mathbb{P}\left(F(x,\xi) \leq \rho\right)$$

where $F : \mathbb{X} \times \Xi \to \mathbb{R}$, $\mathbb{X}$ is the feasible region and ξ is random parameter with distribution $\mathbb{P}$ supported in Ξ. The objective is to estimate J^*.

Wasserstein distance

Definition [Wasserstein Metric]

The *p-Wasserstein distance* $W_p(\mu,\nu)$ between $\mu,\nu \in \mathcal{P}_p(\Xi)$ is defined by

$$W_p^p(\mu,\nu) := \inf_{\Pi\in\mathcal{P}(\Xi\times\Xi)} \left\{ \int_{\Xi\times\Xi} \|\xi-\zeta\|_q^p \,\Pi(d\xi,d\zeta) \;\middle|\; \begin{array}{l}\Pi(\cdot\times\Xi)=\mu(\cdot),\\ \Pi(\Xi\times\cdot)=\nu(\cdot)\end{array} \right\}$$

where d is a metric on Ξ and $\mathcal{P}_p(\Xi)$ is the set of probability measures μ such that $\int_\Xi d^p(\xi,\zeta_0)\mu(d\xi) < \infty$ for some $\zeta_0 \in \Xi$.

In $\mathcal{P}_p(\Xi)$ the ball of ratio $\varepsilon > 0$ and centered at $\mu \in \mathcal{P}_p(\Xi)$ is $\mathcal{B}_\varepsilon(\mu) = \left\{\nu \in \mathcal{P}_p(\Xi) \,\middle|\, W_p^p(\mu,\nu) \leq \varepsilon^p\right\}$.

Let $\widehat{\xi}_1,\ldots,\widehat{\xi}_N$ be a sample of $\mathbb{P}$ and $\widehat{\mathbb{P}}_N$ the empirical distribution determined by this sample. Given $\varepsilon > 0$ we consider $\mathcal{B}_\varepsilon\left(\widehat{\mathbb{P}}_N\right)$.

Distributionally Robust Optimization DRO

The Distributionally Robust Optimization problem that estimates J^* with hight probability is

$$\widehat{J}_N := \min_{x\in\mathbb{X}} \sup_{\mathbb{Q}\in\mathcal{B}_\varepsilon(\widehat{\mathbb{P}}_N)} \mathbb{Q}\left(F(x,\xi)\leq\rho\right). \tag{1}$$

Theorem [Hota et al. [2019], Gao [2022]]

For any p-Wasserstein distance and $q \geq 1$, the problem (1) can be considered equivalent to the following optimization problem:

$$\begin{cases} \inf\limits_{x\in\mathcal{X},\,\lambda\geq 0,\,s} & \lambda\varepsilon^p + \dfrac{1}{N}\displaystyle\sum_{i=1}^N s_i, \\ \text{subject to} & 1-\lambda\left(\textbf{dist}\left(\widehat{\xi}_i, \{\xi\in\Xi \,:\, F(x,\xi)\leq\rho\}\right)\right)^p \leq s_i, \\ & s_i \geq 0, \\ & \forall i\in[N]. \end{cases} \tag{2}$$

Decision-Dependent DRO

The Decision-Dependent Distributionally Robust Optimization problem that estimates J^* with hight probability is

$$\widehat{J}^{\mathrm{Dp}}_{N,p,q}(\varepsilon) := \min_{x\in\mathcal{X}} \sup_{\mathbb{Q}\in\mathcal{B}_{\varepsilon\gamma_{x,F,q}}(\widehat{\mathbb{P}}_N^{x,F})} \mathbb{Q}\left(\zeta\leq\rho\right). \tag{3}$$

where

- For $x\in\mathbb{R}^m$, we define $\zeta^{x,F} := F(x,\xi)$. We call $\mathbb{P}^{x,F}$ to its probability distribution. Because it depends on $\mathbb{P}$, $\mathbb{P}^{x,F}$ is also unknown.
- We define $\widehat{\zeta}_i^{x,F} := F\left(x,\widehat{\xi}_i\right)$, so $\widehat{\zeta}_1^{x,F},\ldots,\widehat{\zeta}_N^{x,F}$ is a sample of $\zeta^{x,F}$. This allows us to define the empirical distribution of $\zeta^{x,F}$, $\widehat{\mathbb{P}}_N^{x,F}$.
- We assume that F is a q-Lipschitz function with respect to ξ. This is, for each x, there exists $\gamma_{x,F,q} > 0$ such that $|F(x,\xi)-F(x,\zeta)| \leq \gamma_{x,F,q}\|\xi-\zeta\|_q$ for all $\xi,\zeta\in\Xi$.
- We define $\mathcal{D} := \mathcal{B}_{\varepsilon\gamma_{x,F,q}}\left(\widehat{\mathbb{P}}_N^{x,F}\right)$, a ball centered at $\widehat{\mathbb{P}}_N^{x,F}$ with radius $\varepsilon\gamma_{x,F,q}$ with respect to the W_p in $\mathbb{R}$.

Reformulation Decision-Dependent DRO

Theorem [Main theorem]

For any p-Wasserstein distance and $q \geq 1$, the problem (3) can be considered equivalent to the following optimization problem:

$$\begin{cases} \inf\limits_{x\in\mathcal{X},\,\lambda\geq 0,\,s,\,w} & \lambda\varepsilon^p\gamma_{x,F,q}^p + \dfrac{1}{N}\displaystyle\sum_{i=1}^N s_i \\ \text{subject to} & 1+w_i(\rho-F(x,\widehat{\xi}_i))\left|\rho-F(x,\widehat{\xi}_i)\right|^{p-1} \leq s_i \;\forall i\in[N], \\ & 0\leq w_i\leq\lambda \qquad \forall i\in[N], \\ & s_i\geq 0 \qquad \forall i\in[N]. \end{cases} \tag{4}$$

Algorithm to solve decision-dependent DRO

Assuming $p=1$ and $q=2$, the objective is to solve (4) for x. For a fixed x, note that the dual problem of (4) is given by

$$\begin{cases} \sup\limits_{u,v\in\mathbb{R}^N,\,\tau\geq -1} & \displaystyle\sum_{i=1}^N u_i \\ \text{subject to} & (\rho-F(x,\widehat{\xi}_i))u_i+v_i\leq 0 \;\forall i\in[N], \\ & \sum_{i=1}^N v_i+\tau\varepsilon\gamma_{x,F,q}\leq 0 \quad \forall i\in[N], \\ & \frac{-1}{N}\leq u_i\leq 0,\;\; v_i\leq 0 \quad \forall i\in[N]. \end{cases} \tag{5}$$

The algorithm to solve (4) for x is based on the subgradient method, which uses the iteration

$$x^{(k+1)} = \mathrm{Proy}_{\mathcal{X}}\left(x^{(k)}+\alpha_k g^{(k)}\right)$$

where $\alpha_k > 0$ is the kth step size, and $g^{(k)} := g(x^{(k)})$ is a vector whose components are given by

$$g(x)_i = \left\langle u^*\circ w^*, \left[\frac{\partial F(x,\widehat{\xi}_1)}{\partial x_i},\ldots,\frac{\partial F(x,\widehat{\xi}_N)}{\partial x_i}\right]\right\rangle - \varepsilon\frac{\partial\gamma_{x,F,q}}{\partial x_i}\lambda^*\tau^*$$

where $\circ$ is the element-wise product (Hadamard product), w^* and λ^* are solutions of (4), and u^* and τ^* are solutions of (5), in both cases for fixed x.

Portfolio Optimization

We consider a portfolio of m assets whose rate of return are described by the random vector $\xi\in\mathbb{R}^m$ and we denote by $x\in\mathbb{R}^m$ the vector of weights associated with the portfolio. In this context, the investor wants to minimize the probability that his return will not exceed a given level.

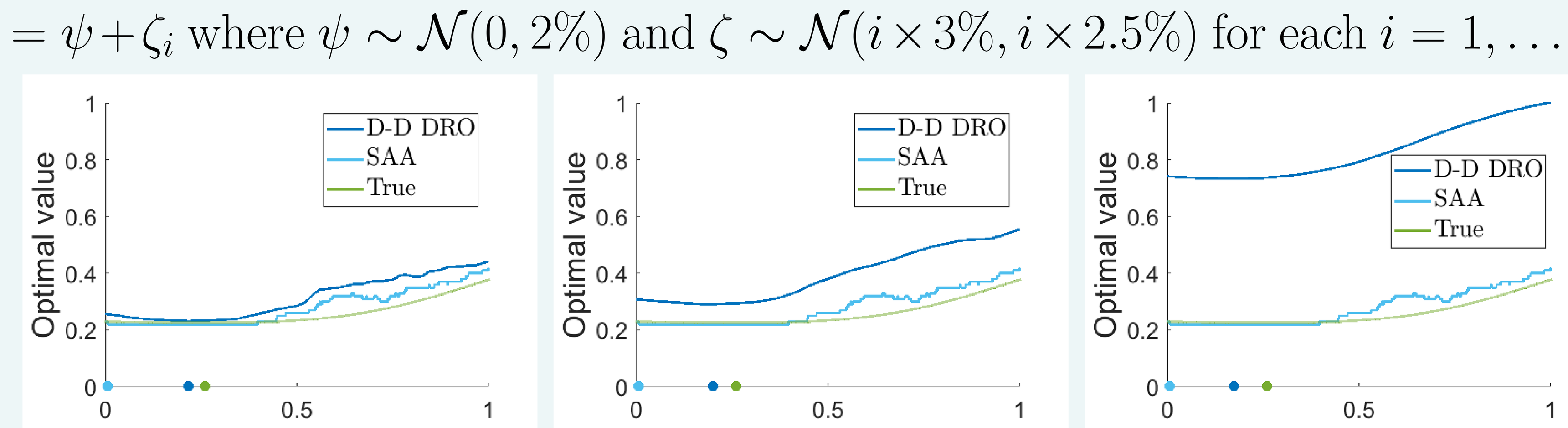
$$J^* = \min_{x\in\mathcal{X}} \mathbb{P}(\langle x,\xi\rangle\leq\rho) \tag{6}$$

where $\mathcal{X} := \{x\in\mathbb{R}^m : \sum_{i=1}^m x_i = 1,\; x_i\geq 0\;\forall i\}$. The decision-dependent DRO version for this case is of the form

$$J^{\mathrm{Dp}}_{N,1,2}(\varepsilon) = \min_{x\in\mathcal{X}} \sup_{\mathbb{Q}\in\mathcal{B}_{\varepsilon\|x\|}(\widehat{\mathbb{P}}_N^{x,F})} \mathbb{Q}(\langle x,\xi\rangle\leq\rho) = \min_{x\in\mathcal{X}} J^{\mathrm{Dp}}_{N,1,2}(\varepsilon,x). \tag{7}$$

Note that, in this case, $F(x,\xi) = \langle x,\xi\rangle$.

Numerical analysis: We assume that the random variable ξ can be expressed as $\xi = \psi+\zeta_i$ where $\psi\sim\mathcal{N}(0,2\%)$ and $\zeta\sim\mathcal{N}(i\times 3\%, i\times 2.5\%)$ for each $i=1,\ldots,m$.



(a) $\varepsilon = 0.0001$ (b) $\varepsilon = 0.001$ (c) $\varepsilon = 0.02$.

Figure 1:In this case, $m=2$ and $\rho=0.02$ are assumed, so $\mathcal{X} = \{x\in\mathbb{R}^2_{\geq 0} : x_1+x_2=1\}$. Then the optimization focuses on x_1. The figure shows $J^{\mathrm{Dp}}_{N,1,2}(\varepsilon,x)$ (dark blue), Sample Average Approximation $\widehat{\mathbb{P}}_N(\langle x,\xi\rangle\leq\rho)$ (light blue), and $\mathbb{P}(\langle x,\xi\rangle\leq\rho)$ (green).

References

R. Gao. Wasserstein regularization for 0-1 loss. *Optimization Online*, 2022.

A. Hota, A. Cherukuri, and J. Lygeros. Data-Driven Chance Constrained Optimization under Wasserstein Ambiguity Sets. *2019 American Control Conference (ACC)*, pages 1501–1506, 2019.