# EPFL | MGT-418 : Convex Optimization | Project 6
## Distributionally Robust Portfolio Optimization

Charles Vuichard, Erick Maraz, Gloria Dal Santo

December 2020

## 1 Sample Average Approximation

The sample average approximation problem (SAA, (1)) is:

$$\max_{\boldsymbol{x}\in\mathcal{X}} \frac{1}{N}\sum_{i=1}^{N} u(\boldsymbol{x}^\top \boldsymbol{\xi}_i) = \max_{\boldsymbol{x}\in\mathcal{X}} \sum_{i=1}^{N} \left[\min_{\ell\,=\,1,\dots,L} a_\ell \boldsymbol{x}^\top \boldsymbol{\xi}_i + b_\ell\right] \tag{1}$$

In order to rewrite the inner minimization problem, we use an epigraphical variable $\boldsymbol{t}\in\mathbb{R}^N$ such that:

$$a_\ell \boldsymbol{x}^\top \boldsymbol{\xi}_i + b_\ell \geq t_i \quad \forall \ell = 1,\dots,L \quad \forall i = 1,\dots,N \tag{2}$$

Hence, we can now write the (SAA, (1)) as a linear problem:

$$\begin{aligned} \max_{\boldsymbol{x}\in\mathcal{X},\,\boldsymbol{t}\in\mathbb{R}^N} \quad & \frac{1}{N}\sum_{i=1}^{N} t_i \\ \text{s.t.} \quad & a_\ell x^\top \boldsymbol{\xi}_i + b_\ell \geq t_i \quad \forall \ell = 1,\dots,L \quad \forall i = 1,\dots,N \\ & \mathbf{1}^\top \boldsymbol{x} = 0 \\ & \boldsymbol{x} \geq 0 \end{aligned} \tag{3}$$

## 2 Distributionally Robust Optimization

### a) Inner Linear Program

The distributionally robust optimization problem (DRO, (5)) is:

$$\max_{\boldsymbol{x}\in X} \min_{\mathbb{Q}\in\mathcal{B}(\hat{\mathbb{P}}_N)} \sum_{i=1}^{N} u(\boldsymbol{x}^\top \boldsymbol{\xi}_i) q_i \tag{4}$$

Recall that: $\mathcal{B}(\hat{\mathbb{P}}_N) = \{\boldsymbol{\xi}_1,\dots,\boldsymbol{\xi}_N \mid d(\hat{\mathbb{P}}_N,\mathbb{Q}) \leq \rho\}$, where $d(\hat{\mathbb{P}}_N,\mathbb{Q})$ is the Wasserstein distance and that we are optimizing the expected utility with respect to $\mathbb{Q}$ on the training samples with $d(\hat{\mathbb{P}}_N,\mathbb{Q})$ at most $\rho > 0$. Therefore, for a fixed $\boldsymbol{x}$, we use $\boldsymbol{\pi}\in\mathbb{R}^{N\times N}$ and $\boldsymbol{q}\in\mathbb{R}^N$ as the decision variables in the inner optimization problem by adding the squared Wasserstein distance in the contraints:

$$
\begin{aligned}
\min_{\boldsymbol{q}\in\mathbb{R}^N,\,\boldsymbol{\pi}\in\mathbb{R}^{N\times N}} \quad & \sum_{i=1}^{N} u(\boldsymbol{x}^\top \boldsymbol{\xi}_i) q_i \\
\text{s.t.} \quad & \sum_{i=1}^{N}\sum_{j=1}^{N} \|\boldsymbol{\xi}_i - \boldsymbol{\xi}_j\|_2^2 \pi_{ij} \leq \rho^2 \\
& \sum_{j=1}^{N} \pi_{ij} = \hat{p}_i && \forall i = 1,\dots,N \\
& \sum_{i=1}^{N} \pi_{ij} = q_j && \forall j = 1,\dots,N \\
& \pi_{ij} \geq 0 && \forall i = 1,\dots,N \quad \forall j = 1,\dots,N
\end{aligned} \tag{5}
$$

## b) Dualization

The Lagrangian of problem (5) is:

$$
\begin{aligned}
\mathcal{L}(\boldsymbol{q},\boldsymbol{\pi},\lambda,\hat{\boldsymbol{\mu}},\boldsymbol{\mu}) =& \sum_{j=1}^{N} u(\boldsymbol{x}^\top \boldsymbol{\xi}_j) q_j + \lambda \left( \sum_{i=1}^{N}\sum_{j=1}^{N} \|\boldsymbol{\xi}_i - \boldsymbol{\xi}_j\|_2^2 \pi_{ij} - \rho^2 \right) \\
& - \sum_{i=1}^{N} \mu_i \left( \left( \sum_{j=1}^{N} \pi_{ij} \right) - \hat{p}_i \right) - \sum_{j=1}^{N} \hat{\mu}_j \left( \left( \sum_{i=1}^{N} \pi_{ij} \right) - q_j \right) \\
=& - \lambda\rho^2 + \sum_{i=1}^{N}\sum_{j=1}^{N} \left( \lambda \|\boldsymbol{\xi}_i - \boldsymbol{\xi}_j\|_2^2 - \mu_i - \hat{\mu}_j \right) \pi_{ij} \\
& + \sum_{j=1}^{N} \left( u(\boldsymbol{x}^\top \boldsymbol{\xi}_j) + \hat{\mu}_j \right) q_j + \sum_{i=1}^{N} \mu_i \hat{p}_i
\end{aligned} \tag{6}
$$

*Note that we change the index of the objective function from $i$ to $j$ for better readability.*

By analyzing the grouped terms of the decision variables in (6), we can find the dual objective of (5):

$$
g(\lambda,\hat{\boldsymbol{\mu}},\boldsymbol{\mu}) = \inf_{\boldsymbol{q}\in\mathbb{R}^N,\,\boldsymbol{\pi}\in\mathbb{R}^{N\times N}} \mathcal{L}(\boldsymbol{q},\boldsymbol{\pi},\lambda,\hat{\boldsymbol{\mu}},\boldsymbol{\mu}) \tag{7}
$$

$$
= \begin{cases} -\lambda\rho^2 + \sum_{i=1}^{N} \mu_i \hat{p}_i & \text{if} \quad \begin{cases} \hat{\mu}_j & = -u(\boldsymbol{x}^\top \boldsymbol{\xi}_j) \quad \text{and} \\ \mu_i + \hat{\mu}_j & \leq \lambda \|\boldsymbol{\xi}_i - \boldsymbol{\xi}_j\|_2^2 \end{cases} \\ -\infty & \text{otherwise} \end{cases} \tag{8}
$$

Finally, since we know that $\hat{p}_i = 1/N$, the dual problem is:

$$
\begin{aligned}
\max_{\lambda\in\mathbb{R}_+,\,\hat{\boldsymbol{\mu}}\in\mathbb{R}^N,\,\boldsymbol{\mu}\in\mathbb{R}^N} \quad & - \lambda\rho^2 + \frac{1}{N}\sum_{i=1}^{N} \mu_i \\
\text{s.t.} \quad & \hat{\mu}_i = -u(\boldsymbol{x}^\top \boldsymbol{\xi}_i) && \forall i = 1,\dots,N \\
& \mu_i + \hat{\mu}_j \leq \lambda \|\boldsymbol{\xi}_i - \boldsymbol{\xi}_j\|_2^2 && \forall i = 1,\dots,N \quad \forall j = 1,\dots,N \\
& \lambda \geq 0
\end{aligned} \tag{9}
$$

## c) Slater Point

A Slater point $(\pi, q)$ holds is the following conditions are satisfied:

- A : $\sum_{i=1}^{N}\sum_{j=1}^{N}(\|\boldsymbol{\xi}_i - \boldsymbol{\xi}_j\|_2^2)\pi_{ij} < \rho^2$
- B : $\sum_{j=1}^{N} \pi_{ij} = \hat{p}_i$
- C : $\sum_{i=1}^{N} \pi_{ij} = q_j$
- D : $\pi_{ij} > 0$

Where $\rho > 0$.
For A : $\sum_{i=1}^{N}\sum_{j=1}^{N}(\|\boldsymbol{\xi}_i - \boldsymbol{\xi}_j\|_2^2)\pi_{ij} < \rho^2$, let's take $\pi_{ij,i\neq j} = \frac{\rho^2}{\|\boldsymbol{\xi}_i - \boldsymbol{\xi}_j\|_2^2 (N-1)^2}\frac{1}{a}$

$$\Rightarrow \sum_{i=1}^{N}\sum_{j=1}^{N}(\|\boldsymbol{\xi}_i - \boldsymbol{\xi}_j\|_2^2)\pi_{ij} = \frac{\rho^2}{a}$$
$$\Rightarrow a > 1$$

For B : $\sum_{j=1}^{N} \pi_{ij} = \hat{p}_i$, let's take $\pi_{ij,i=j} = \hat{p}_i - \sum_{j=1,i\neq j}^{N} \frac{\rho^2}{\|\boldsymbol{\xi}_i - \boldsymbol{\xi}_j\|_2^2 (N-1)^2}\frac{1}{a}$, $\hat{p}_i = \frac{1}{N}$

$$\Rightarrow \sum_{j=1}^{N} \pi_{ij} = \hat{p}_i$$

Hence, $\pi$ is symmetric.
For C : $\sum_{i=1}^{N} \pi_{ij} = q_j$ :

$$\sum_{i=1}^{N} \pi_{ij} = \hat{p}_i = q_j = \frac{1}{N}$$

For D $\pi_{ij} > 0$, we have $\pi_{ij,i\neq j} = \frac{\rho^2}{\|\boldsymbol{\xi}_i - \boldsymbol{\xi}_j\|_2^2 (N-1)^2}\frac{1}{a} > 0$.
$\pi_{ij,i=j} = \frac{1}{N} - \sum_{j=1,i\neq j}^{N} \frac{\rho^2}{\|\boldsymbol{\xi}_i - \boldsymbol{\xi}_j\|_2^2 (N-1)^2}\frac{1}{a} > 0$ if :

$$\frac{1}{N} > \sum_{j=1,i\neq j}^{N} \frac{\rho^2}{\|\boldsymbol{\xi}_i - \boldsymbol{\xi}_j\|_2^2 (N-1)^2}\frac{1}{a}$$
$$\Rightarrow a > \sum_{j=1,i\neq j}^{N} \frac{\rho^2 N}{\|\boldsymbol{\xi}_i - \boldsymbol{\xi}_j\|_2^2 (N-1)^2}$$
$$a = max\left(1, \sum_{j=1,i\neq j}^{N} \frac{\rho^2 N}{\|\boldsymbol{\xi}_i - \boldsymbol{\xi}_j\|_2^2 (N-1)^2)}\right) + 1$$

Hence $\pi_{ij} > 0$.

For the Slater point $(\pi, q)$ with $\pi_{ij,i\neq j} = \frac{\rho^2}{\|\boldsymbol{\xi}_i - \boldsymbol{\xi}_j\|_2^2 (N-1)^2}\frac{1}{a}$, $\pi_{ij,i=j} = \hat{p}_i - \sum_{j=1,i\neq j}^{N} \frac{\rho^2}{\|\boldsymbol{\xi}_i - \boldsymbol{\xi}_j\|_2^2 (N-1)^2}\frac{1}{a}$, $\hat{p}_i = \frac{1}{N}$, $q_j = \frac{1}{N}$ all the conditions A, B, C and D hold. For $\rho > 0$, strong duality holds between the Inner Linear Program and the Dual Problem

### d) Final Linear Program

Since strong duality holds for the inner minimization problem (5), we reformulate the problem (4) by maximizing over the outer decision variables of (4) and the inner decision variables of (9). We deal with the utility function $u(\boldsymbol{x}^\top \boldsymbol{\xi})$ constraint with the epigraphical variable (2).

$$
\begin{array}{lll}
\max\limits_{\substack{\boldsymbol{x}\in\mathcal{X},\boldsymbol{t}\in\mathbb{R}^N \\ \lambda\in\mathbb{R}_+,\hat{\boldsymbol{\mu}}\in\mathbb{R}^N,\boldsymbol{\mu}\in\mathbb{R}^N}} & -\lambda\rho^2 + \dfrac{1}{N}\sum_{i=1}^{N}\mu_i & \\
\text{s.t.} & \mu_i + \hat{\mu}_j \leq \lambda\|\boldsymbol{\xi}_i - \boldsymbol{\xi}_j\|_2^2 & \forall i = 1,\dots,N \quad \forall j = 1,\dots,N \\
& t_i \leq a_\ell x^\top \boldsymbol{\xi}_i + b_\ell & \forall \ell = 1,\dots,L \quad \forall i = 1,\dots,N \\
& t_i = -\hat{\mu}_i & \forall i = 1,\dots,N \\
& \mathbf{1}^\top \boldsymbol{x} = 1 & \\
& \boldsymbol{x} \geq 0 & \\
& \lambda \geq 0 &
\end{array} \tag{10}
$$

# 3 Perfect Distributional Information

The optimal value of the SAA problem is: `0.57124`

# 4 DRO Implementation

The optimal value of the DRO problem is: `0.52546`

# 5 SAA vs. DRO

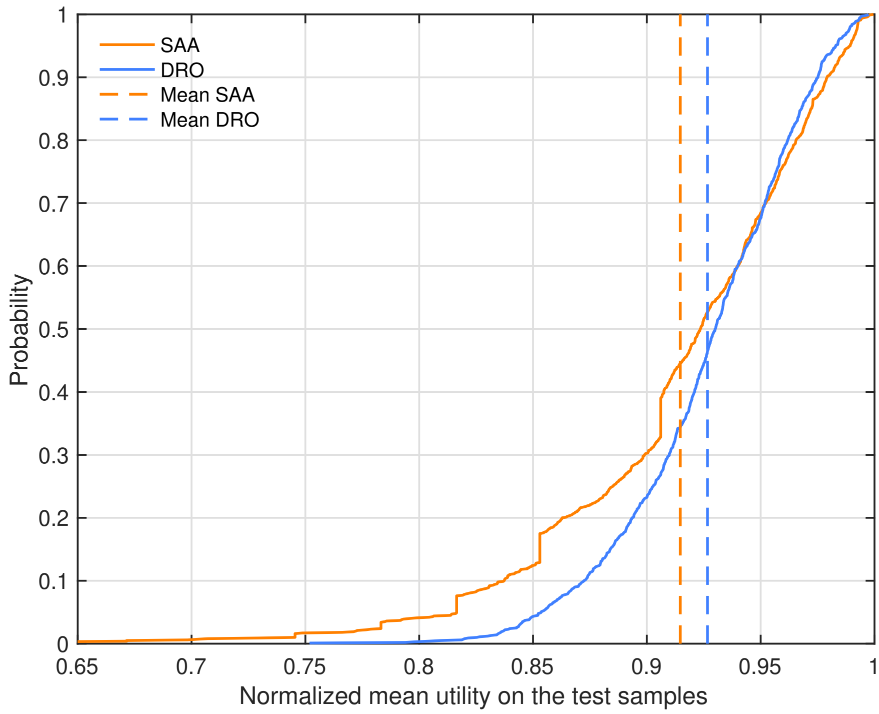


Figure 1: SAA vs DRO

# 6 Interpretation

From the graph depicted in Figure 1 we can obtain the information summarized in the following table:

| | *worst-case* | *best-case* | *mean* |
|---|---|---|---|
| SAA | 0.6332 | 0.9991 | 0.9137 |
| DRO | 0.7518 | 0.9974 | 0.9268 |

Table 1: Normalized mean utility values obtained from SAA and DRO problems

We can see that the SAA leads the DRO curve for mean utility values greater than 0.9535 thus reaching a slightly higher best-case value. However, the difference between the worst-case values obtained with the two approximations is much greater, for which we have the DRO taking a higher mean value. Indeed, solving the robust optimization problem we are trying to maximize the worst-case expected utility. In this way we are also shifting the mean over the testing samples to a higher value. In conclusion we can say that, despite the slightly lower best-case value, to apply the DRO problem ensures a utility that is favorable in most of the conditions, confirming the robust nature of the optimization problem.