EPFL | MGT-418 : Convex Optimization | Project 6

Distributionally Robust Portfolio Optimization

(Graded)

Description

An investor wishes to invest into a portfolio of K assets with uncertain rates of return $\boldsymbol{\xi} \in [-1, +\infty)^K$. The investor's decision is the weight x_k to attribute to each asset k. Since short sales are forbidden, the vector of portfolio weights $\boldsymbol{x}$ ranges over the feasible set $\mathcal{X} = \left\{\boldsymbol{x} \in \mathbb{R}_+^K : \mathbf{1}^\top \boldsymbol{x} = 1\right\}$, where $\mathbf{1}$ stands for the vector of ones. The rate of return of the portfolio is given by $\boldsymbol{x}^\top \boldsymbol{\xi}$. Assume that the investor's objective is to maximize the expected value of the concave piecewise linear utility function

$$u(\boldsymbol{x}^\top \boldsymbol{\xi}) = \min_{l=1,\ldots,L} \left\{a_l \boldsymbol{x}^\top \boldsymbol{\xi} + b_l\right\},$$

where $a_l \in \mathbb{R}_+$ and $b_l \in \mathbb{R}$ for all $l = 1, \ldots, L$. The expectation should be calculated under the distribution $\mathbb{P}$ of $\boldsymbol{\xi}$. This distribution, however, is unknown, and the investor has only access to N independent training samples $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_N$ from $\mathbb{P}$. These samples must be used to estimate the expected utility of an investment decision. A naive approach is to assume that the N samples provide perfect information about $\mathbb{P}$. A more conservative approach would be to assume that the samples provide only partial information about $\mathbb{P}$. In the following, we will compare both approaches. In particular, we will compare (i) the sample average approximation, which approximates $\mathbb{P}$ by the empirical distribution $\hat{\mathbb{P}}_N$ that assigns the same probability $\hat{p}_i = 1/N$ to each training sample, $i = 1, \ldots, N$, with (ii) the distributionally robust approach, which evaluates the worst-case expected utility with respect to all discrete distributions $\mathbb{Q}$ on the training samples that have a Wasserstein distance $d(\hat{\mathbb{P}}, \mathbb{Q})$ of at most $\rho > 0$ from the empirical distribution $\hat{\mathbb{P}}_N$. The Wasserstein distance between two discrete probability distributions $\hat{\mathbb{P}}_N$ and $\mathbb{Q}$ is defined as the square root of the optimal value of a transportation problem (OT). Formally, we set

$$d^2(\hat{\mathbb{P}}_N, \mathbb{Q}) = \left\{ \begin{array}{lll} \displaystyle\min_{\boldsymbol{\pi} \in \mathbb{R}^{N \times N}} & \displaystyle\sum_{i=1}^N \sum_{j=1}^N \|\boldsymbol{\xi}_i - \boldsymbol{\xi}_j\|_2^2 \, \pi_{ij} & \\ \text{s.t.} & \displaystyle\sum_{j=1}^N \pi_{ij} = \hat{p}_i & \forall i = 1, \ldots, N \\ & \displaystyle\sum_{i=1}^N \pi_{ij} = q_j & \forall j = 1, \ldots, N \\ & \pi_{ij} \geq 0 & \forall i = 1, \ldots, N, \quad \forall j = 1, \ldots, N, \end{array} \right. \tag{OT}$$

where q_j denotes the probability of the training sample ξ_j under the distribution $\mathbb{Q}$. In the following, we denote by $\mathcal{B}_\rho(\hat{\mathbb{P}}_N)$ the family of all N-point distributions on $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_N$ with $d(\hat{\mathbb{P}}_N, \mathbb{Q}) \leq \rho$. We would expect the distributionally robust solution to be more conservative than the sample average approximation. In question 5 you will perform a numerical case-study to confirm or refute this expectation.

Questions

1. **Sample Average Approximation (10 points):** Formulate the following sample average approximation problem as a linear program.

$$\max_{\boldsymbol{x}\in\mathcal{X}} \mathbb{E}_{\boldsymbol{\xi}\sim\hat{\mathbb{P}}_N}\left[u(\boldsymbol{x}^\top\boldsymbol{\xi})\right] = \max_{\boldsymbol{x}\in\mathcal{X}} \frac{1}{N}\sum_{i=1}^{N} u\left(\boldsymbol{x}^\top\boldsymbol{\xi}_i\right) \tag{SAA}$$

2. **Distributionally Robust Optimization (40 points):** Formulate the following distributionally robust optimization problem as a linear program.

$$\max_{\boldsymbol{x}\in\mathcal{X}} \min_{\mathbb{Q}\in\mathcal{B}_\rho(\hat{\mathbb{P}}_N)} \mathbb{E}_{\boldsymbol{\xi}\sim\mathbb{Q}}\left[u(\boldsymbol{x}^\top\boldsymbol{\xi})\right] = \max_{\boldsymbol{x}\in\mathcal{X}} \min_{\mathbb{Q}\in\mathcal{B}_\rho(\hat{\mathbb{P}}_N)} \sum_{i=1}^{N} u\left(\boldsymbol{x}^\top\boldsymbol{\xi}_i\right) q_i \tag{DRO}$$

   (a) **Inner Linear Program (5 points):** Show that the inner minimization problem can be reformulated as the following linear program over $\boldsymbol{\pi}$ and $\boldsymbol{q}$.

$$\begin{array}{lll} \min\limits_{\boldsymbol{q}\in\mathbb{R}^N,\,\boldsymbol{\pi}\in\mathbb{R}^{N\times N}} & \sum\limits_{i=1}^{N} u\left(\boldsymbol{x}^\top\boldsymbol{\xi}_i\right) q_i & \\ \text{s.t.} & \sum\limits_{i=1}^{N}\sum\limits_{j=1}^{N} \|\boldsymbol{\xi}_i-\boldsymbol{\xi}_j\|_2^2\pi_{ij} \le \rho^2 & \\ & \sum\limits_{j=1}^{N} \pi_{ij} = \hat{p}_i & \forall i=1,\dots,N \\ & \sum\limits_{i=1}^{N} \pi_{ij} = q_j & \forall j=1,\dots,N \\ & \pi_{ij}\ge 0 & \forall i=1,\dots,N \quad \forall j=1,\dots,N \end{array} \tag{1}$$

   (b) **Dualization (20 points):** Denote the Lagrange multipliers for the upper bound on the weighted sum of all elements of $\boldsymbol{\pi}$ by λ, for the equality on the row-sums by $\hat{\boldsymbol{\mu}}$, and for the equality on the column-sums by $\boldsymbol{\mu}$. Show that the dual problem to (1) is

$$\begin{array}{lll} \max\limits_{\lambda\ge 0,\,\hat{\boldsymbol{\mu}}\in\mathbb{R}^N,\,\boldsymbol{\mu}\in\mathbb{R}^N} & -\lambda\rho^2 + \frac{1}{N}\sum\limits_{i=1}^{N}\hat{\mu}_i & \\ \text{s.t.} & \mu_i = -u\left(\boldsymbol{x}^\top\boldsymbol{\xi}_i\right) & \forall i=1,\dots,N \\ & \mu_i + \hat{\mu}_j \le \lambda\|\boldsymbol{\xi}_i-\boldsymbol{\xi}_j\|_2^2 & \forall i=1,\dots,N \quad \forall j=1,\dots,N. \end{array} \tag{2}$$

   (c) **Slater Point (10 points):** Show that if $\rho > 0$, then strong duality holds between Problems (1) and (2), by finding a Slater point to (1).

   (d) **Final Linear Program (5 points):** Combine the inner maximization problem (2) over the dual variables with the outer maximization problem over $\boldsymbol{x}$ into a single linear program.

3. **Perfect Distributional Information (10 points):** Implement the SAA problem for the 10,000 samples provided in the file `test.mat` and report its optimal value. Use the skeleton code in the file `p6q3.m`.

4. **DRO Implementation (10 points):** Implement the DRO problem for the 30 training samples provided in the file `train.mat` for a Wasserstein radius of $\rho = 0.9$. Report the mean utility of the optimal decision $\boldsymbol{x}^\star$ on the 10,000 test samples from the file `test.mat`. Use the skeleton code in the file `p6q4.m`.

5. **SAA vs. DRO (20 points):** Solve the SAA and the DRO problems for 1,000 independent datasets, each containing 30 training samples, and plot the cumulative distribution function of the corresponding mean utilities on the 10,000 test samples, normalized by the mean utility under perfect distributional information, for both problems. Use the skeleton code in the file `p6q5.m`.

6. **Interpretation (10 points):** Compare the SAA and the DRO cumulative distribution functions in terms of worst-case, mean, and best-case performance. Interpret your results.