Portfolio Optimization with Wasserstein Distances

Arnav Hiray
Mentored by Sebastian Gutierrez Hernandez

Spring 2024 Georgia Tech Directed Reading Program (Mathematics)

April 18^{th}, 2024

Portfolio Optimization

- **Definition:** Portfolio optimization is a mathematical framework aimed at selecting the optimal mix of investments to achieve a certain objective
- **Examples of Different Objectives:**
  - *Minimize Risk for a Given Return:* Target a specific return level while reducing the overall risk as much as possible.
  - *Maximize Return for a Given Risk Level:* Achieve the highest possible return without exceeding a predefined risk threshold.
  - *Minimize Costs:* Construct a portfolio that minimizes transaction costs and tax liabilities.
  - *Maintain Liquidity:* Ensure that the portfolio includes assets that can be quickly and easily converted to cash.

Global Minimum Variance

$$\min_{\mathbf{w}} \quad \mathbf{w}^T \Sigma \mathbf{w}$$
$$s.t \quad \mathbf{w}^T \mathbf{1} = 1$$

The matrix Σ is the covariance matrix that quantifies the relationship between returns of different assets, measuring how much the returns of one asset move together with another.

However, estimating Σ relies on historical returns, and this can be highly unreliable and inaccurate for future estimation.

Global Minimum Variance

$$\min_{\mathbf{w}} \quad \mathbf{w}^T \Sigma \mathbf{w}$$
$$s.t \quad \mathbf{w}^T \mathbf{1} = 1$$

The matrix Σ is the covariance matrix that quantifies the relationship between returns of different assets, measuring how much the returns of one asset move together with another.

However, estimating Σ relies on historical returns, and this can be highly unreliable and inaccurate for future estimation.

Wasserstein Distance

Wasserstein Distance

We define the 2−Wasserstein Distance between two probability measures α, β as

$$W_2(\alpha, \beta) = \inf_{\pi \in \mathcal{U}(\alpha,\beta)} \left(\int_{\mathcal{X} \times \mathcal{X}} ||x - y||_2^2 d\pi(x, y) \right)^{\frac{1}{2}}$$

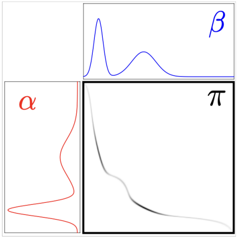


Figure: Peyre, G., & Cuturi, M. (2019)

Wasserstein Ball

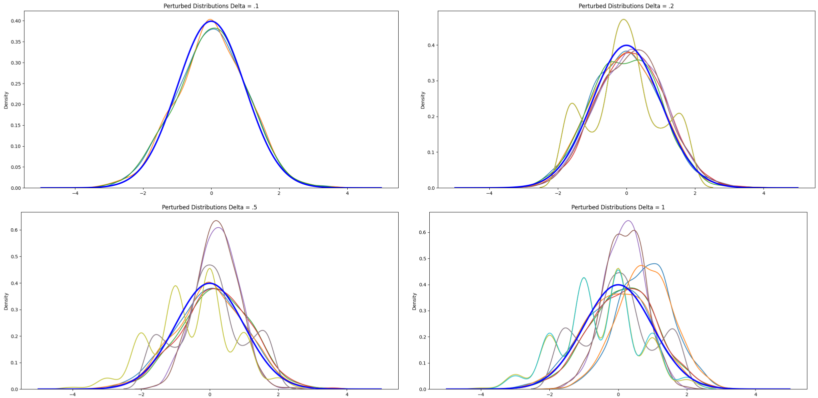
Wasserstein Ball

The Wasserstein Ball of radius $\delta > 0$ centered at a measure P_T is the set

$$\mathcal{B}_\delta(P_T) = \{P : W_p(P_T, P) \leq \delta\}$$

It can get wild.

Visualizing Wasserstein Balls: Standard Normal



Distributionally Robust Optimization Problem

The goal of DRO is to minimize portfolio risk under the worst-case scenario subject to constraints.

DRO Formulation of Global Minimum Variance

$$\min_{\mathbf{w}} \max_{P \in \mathcal{B}_\delta(P_T)} \mathbf{w}^T \Sigma_P \mathbf{w}$$

$$s.t \quad \mathbf{w}^T \mathbf{1} = 1$$

- P_T: Empirical expected return
- B_δ: W_2 ball of radius δ
- w: Admissible allocation
- $\mathbf{w}^T \Sigma_P \mathbf{w}$: Variance of allocation w and density P

Distributionally Robust Optimization Problem

The goal of DRO is to minimize portfolio risk under the worst-case scenario subject to constraints.

DRO Formulation of Global Minimum Variance

$$\min_{\mathbf{w}} \max_{P \in \mathcal{B}_\delta(P_T)} \mathbf{w}^T \Sigma_P \mathbf{w}$$
$$s.t \quad \mathbf{w}^T \mathbf{1} = 1$$

- P_T: Empirical expected return
- B_δ: W_2 ball of radius δ
- w: Admissible allocation
- $\mathbf{w}^T \Sigma_P \mathbf{w}$: Variance of allocation w and density P

This is an infinite-dimensional optimization problem.

Distributionally Robust Optimization Problem

The goal of DRO is to minimize portfolio risk under the worst-case scenario subject to constraints.

DRO Formulation of Global Minimum Variance

$$\min_{\mathbf{w}} \max_{P \in \mathcal{B}_\delta(P_T)} \mathbf{w}^T \Sigma_P \mathbf{w}$$
$$s.t \quad \mathbf{w}^T \mathbf{1} = 1$$

- P_T: Empirical expected return
- B_δ: W_2 ball of radius δ
- w: Admissible allocation
- $\mathbf{w}^T \Sigma_P \mathbf{w}$: Variance of allocation w and density P

This is an infinite-dimensional optimization problem.

Reformulating DRO (Blanchet, Chen, and Zhou 2022)

DRO Global Minimum Optimization Reformulation

The previous optimization can be formulated as the following:

$$\min_{\mathbf{w} \in \mathbb{R}^d} \sqrt{\mathbf{w}^T \Sigma \mathbf{w}} + \sqrt{\delta} ||\mathbf{w}||_2$$

$$s.t \quad \mathbf{w}^T \mathbf{1} = 1$$

Note the above is a Convex Optimization problem.

The Regularization Penalty

- **Purpose of Regularization:**
  - *Enhance Sparsity:* Promotes fewer stock selections in portfolios by adding a penalty term.
  - *Address Overfitting:* Helps prevent the model from fitting too closely to the historical data, enhancing predictive performance.
- **Data-Informed Parameter Adjustment:**
  - *Endogenous Parameter δ:* The level of regularization is informed and adjusted based on empirical data analysis techniques.
  - Larger δ correlated with more conservative DRO.

References

- Jose Blanchet, Lin Chen, Xun Yu Zhou (2022) Distributionally Robust Mean-Variance Portfolio Selection with Wasserstein Distances. Management Science 68(9):6382-6410. https://doi.org/10.1287/mnsc.2021.4155
- Zhang, Z.; Jing, H.; Kao, C. High-Dimensional Distributionally Robust Mean-Variance Efficient Portfolio Selection. Mathematics 2023, 11, 1272. https://doi.org/10.3390/math11051272
- Peyre, G., & Cuturi, M. (2019). Computational Optimal Transport. Foundations and Trends in Machine Learning, 11(5-6), 355-607.

Reformulation DRO (Continue)

Proposition A.2

For any cost function c that is lower semi-continuous and nonnegative, the optimal value of the following problem

$$\max_{P \in \mathcal{B}_\delta(P_T)} \mathbf{w}^T \Sigma_P \mathbf{w}$$

can be reformulated as

$$\inf_{\lambda \geq 0} \left[\frac{1}{n} \sum_{i=1}^{n} \Phi(R_i) + \lambda\delta \right]$$

where $\Phi(R_i) = \sup_u \left[(\mathbf{w}^T u)^2 - \lambda c(u, R_i) \right]$

Reformulation DRO (Continue)

Proposition A.3 + Theorem 1

Let the cost function be $||x-y||_2^2$. Then, the optimal value of the following problem

$$\max_{P\in\mathcal{B}_\delta(P_T)} \mathbf{w}^T\Sigma_P\mathbf{w}$$

is equal to

$$\mathbb{E}_{P_T}[(\mathbf{w}^TR)^2] + \delta||\mathbf{w}||_2^2 + 2\sqrt{\delta||\mathbf{w}_2^2}\sqrt{\mathbf{w}^T\Sigma\mathbf{w}}$$

Take $\Phi(R_i) = \sup_u \left[(\mathbf{w}^Tu)^2 - \lambda||u-R_i||^2\right]$ and use Proposition A.2.

$$\mathbb{E}_{P_T}[(\mathbf{w}^TR)^2] + \delta||\mathbf{w}||_2^2 + 2\sqrt{\delta||\mathbf{w}_2^2}\sqrt{\mathbf{w}^T\Sigma\mathbf{w}} = (\sqrt{\mathbf{w}^T\Sigma\mathbf{w}} + \sqrt{\delta}||\mathbf{w}||_2)^2$$