Data-Driven Optimization with Distributionally Robust Stochastic Dominance

Chun Peng

HEC Montréal & GERAD, Montréal, Canada

(Joint work with Prof. Erick Delage)

chun.peng@hec.ca

2020 Virtual INFORMS

Outline

Stochastic Dominance

Definition (Stochastic Dominance, SD)

Given any two random variables X and Y capturing some earnings, we consider that X stochastically dominates Y to the k-th order, denoted by $X \succeq_{(k)} Y$, if and only if

$$F_X^{(k)}(\eta) \leq F_Y^{(k)}(\eta), \forall \eta \in \mathbb{R},$$

where $F_X^{(1)}(\eta) = \mathbb{P}(X \leq \eta)$ and $F_X^{(k)}(\eta) = \int_{-\infty}^{\eta} F_X^{(k-1)}(t)\, dt$, $k = 2, 3, \ldots$. Furthermore, the dominance is known as FSD when $k = 1$ and SSD when $k = 2$.

- $X \succeq_{(1)} Y \Leftrightarrow F_X(\eta) \leq F_Y(\eta), \forall \eta \in \mathbb{R}.$
- $X \succeq_{(2)} Y \Leftrightarrow \mathbb{E}[(\eta - X)^+] \leq \mathbb{E}[(\eta - Y)^+], \forall \eta \in \mathbb{R}.$
- $X \succeq_{(1)} Y \Leftrightarrow \mathbb{E}[u(X)] \geq \mathbb{E}[u(Y)]$ for all non-decreasing functions u.
- $X \succeq_{(2)} Y \Leftrightarrow \mathbb{E}[u(X)] \geq \mathbb{E}[u(Y)]$ for all non-decreasing concave functions u.

Optimization with SD Constraints

Consider problem with a k-th order SD constraint:

$$\text{[SDCPk]} \quad \underset{\boldsymbol{x}\in\mathcal{X}}{\text{minimize}} \; \boldsymbol{c}^\top \boldsymbol{x} \tag{1a}$$

$$\text{subject to} \; f(\boldsymbol{x},\boldsymbol{\xi}) \succeq_{(k)}^{\mathbb{P}} f_0(\boldsymbol{\xi}) \tag{1b}$$

- $f(\boldsymbol{x},\boldsymbol{\xi})$ is the random controlled performance function, and $f_0(\boldsymbol{\xi})$ is the random reference performance function. Let $f_0(\boldsymbol{\xi})$ be $f(\boldsymbol{x}_0,\boldsymbol{\xi})$ with $\boldsymbol{x}_0 \in \mathcal{X}$;
- SDCP1 and SDCP2 are widely studied in SP context: $\mathbb{P}$ has a finite empirical distribution with $\{\hat{\xi}_i\}_{i=1}^{M}$, e.g. [Dentcheva and Ruszczyński. 2003], [Luedtke. 2008], [Hu et al. 2012], [Dentcheva and Wolfhagen. 2015], $\cdots$
- decision-making under uncertainty: SDCPk assumes the full information of probability distribution of $\boldsymbol{\xi}$.

Distributionally Robust Stochastic Dominance[1]

Definition (Distributionally Robust Stochastic Dominance, DRSD)

Given two random variables X and Y, we say that X robustly stochastically dominates Y in the k-th order if and only if:

$$X \succeq_{(k)}^{\mathbb{P}} Y \quad \forall \mathbb{P} \in \mathcal{P},$$

where $k = 1, 2, \cdots$, and in particular, this relation is referred as *DRFSD* and *DRSSD* if $k = 1$ and 2, respectively.

[1] Dentcheva, D., Ruszczyński, A. Robust stochastic dominance and its application to risk-averse optimization. *Mathematical Programming*, 2010, 123(1), 85-100.

Optimization with DRSD Constraints

Consider the distributionally robust stochastic dominance constrained problem:

$$[\text{DRSDCPk}] \quad \underset{\boldsymbol{x}\in\mathcal{X}}{\text{minimize}} \;\; \boldsymbol{c}^\top \boldsymbol{x} \tag{2a}$$

$$\text{subject to} \;\; f(\boldsymbol{x},\boldsymbol{\xi}) \succeq_{(k)}^{\mathbb{P}} f_0(\boldsymbol{\xi}) \qquad \forall \mathbb{P} \in \mathcal{P}. \tag{2b}$$

- relevant studies:

- [Dentcheva and Ruszczyński. 2010]: the definition of DRSSD relation, derive the optimality conditions from a theoretical viewpoint.
- [Chen and Jiang. 2018]: derive quantitative stability results for problem with k-th order DRSD constraints induced by full random recourse.
  Cons: do not propose a numerical solution scheme (formulation, algorithm).
- [Guo et al. 2017]: use a discretization scheme to approximate DRSSDCP under a moment-based ambiguity set.
- [Sehgal and Mehra, 2020]: study a robust portfolio selection with SSD under a budget uncertainty set of stock return.
  Note: both assume reference performance function $f_0(\boldsymbol{\xi})$ is finitely known.

Data-Driven DRSSDCP

Consider DRSDCP in the second-order:

$$\text{[DRSSDCP]} \quad \underset{\boldsymbol{x}\in\mathcal{X}}{\text{minimize}} \quad \boldsymbol{c}^\top \boldsymbol{x} \tag{3a}$$

$$\text{subject to} \quad f(\boldsymbol{x},\boldsymbol{\xi}) \succeq_{(2)}^{\mathbb{P}} f_0(\boldsymbol{\xi}) \qquad \forall \mathbb{P} \in \mathcal{P}_{\mathsf{W}}^r(\hat{\mathbb{P}}, \epsilon). \tag{3b}$$

Definition (Wasserstein Ambiguity Set)

The type-r Wasserstein ambiguity set of radius ϵ centered at $\bar{\mathbb{P}}$ is defined by

$$\mathcal{P}_{\mathsf{W}}^r(\hat{\mathbb{P}}, \epsilon) := \left\{\mathbb{P} \in \mathcal{M}(\Xi) \,\middle|\, d_{\mathsf{W}}^r(\mathbb{P}, \bar{\mathbb{P}}) \le \epsilon \right\},$$

where $\mathcal{M}(\Xi)$ is the space of all distributions supported on Ξ and d_{W} is the Wasserstein metric.

Assumption 1: Ξ is nonempty compact convex and $\mathcal{X}$ is nonempty convex.
Assumption 2: $f(\boldsymbol{x},\boldsymbol{\xi})$ and $f_0(\boldsymbol{\xi})$ are piecewise linear concave in both $\boldsymbol{x}$ and $\boldsymbol{\xi}$, e.g. $f(\boldsymbol{x},\boldsymbol{\xi}) := \min_{n\in[N]} a_n(\boldsymbol{x})^\top \boldsymbol{\xi} + b_n(\boldsymbol{x})$, $f_0(\boldsymbol{\xi}) := \min_{n\in[N]} {a_n^0}^\top \boldsymbol{\xi} + b_n^0$.

Special Cases of DRSSDCP (3)

Proposition (Reduction to SDCP2)

DRSSDCP (3) with $\mathcal{P} := \mathcal{P}^r_W(\hat{\mathbb{P}}, 0)$ *reduces to SDCP2 (1) with* $\mathbb{P} := \hat{\mathbb{P}}$*. Moreover, it can be reformulated as a linear program if* $\mathcal{X}$ *is polyhedral.*

Proposition (Reduction to DFSDCP)

DRSSDCP (3) with $\mathcal{P} := \mathcal{P}^r_W(\hat{\mathbb{P}}, \infty)$ *reduces to a distribution-free statewise dominance constrained problem (DFSDCP),*

$$\begin{aligned} \underset{\boldsymbol{x} \in \mathcal{X}}{\text{minimize}} \quad & \boldsymbol{c}^\top \boldsymbol{x} \\ \text{subject to} \quad & f(\boldsymbol{x}, \boldsymbol{\xi}) \geq f_0(\boldsymbol{\xi}) \qquad \forall \boldsymbol{\xi} \in \Xi. \end{aligned}$$

Moreover, it can be reformulated as a linear program if $\mathcal{X}$ *and* Ξ *are polyhedral.*

Multistage Robust Optimization Reformulation

Proposition

Under type-1 Wasserstein ambiguity set $\mathcal{P}^1_W(\hat{\mathbb{P}}, \epsilon)$ with $\epsilon \in (0, \infty)$, DRSSDCP (3) coincides with the optimal value of the multistage robust optimization problem (5),

$$
\begin{aligned}
&\underset{\boldsymbol{x}\in\mathcal{X}}{\text{minimize}} && \boldsymbol{c}^\top \boldsymbol{x} && && \text{(5a)}\\
&\text{subject to} && L(\boldsymbol{x}, t) \le 0 && \forall t \in \bar{\mathcal{T}} && \text{(5b)}\\
&\textit{where} && L(\boldsymbol{x}, t) := \inf_{\lambda, \boldsymbol{q}} \lambda\epsilon + \frac{1}{M}\sum_{i=1}^{M} q_i && && \text{(5c)}\\
&\textit{s.t.} && g(\boldsymbol{x}, \boldsymbol{\xi}, t) - \lambda\|\boldsymbol{\xi} - \hat{\xi}_i\| \le q_i && \forall i \in [M], \boldsymbol{\xi} \in \Xi && \text{(5d)}\\
& && \boldsymbol{\lambda} \ge 0, \boldsymbol{q} \in \mathbb{R}^M, && && \text{(5e)}
\end{aligned}
$$

where $\bar{\mathcal{T}} := \left[\inf_{\boldsymbol{\xi}\in\Xi} f_0(\boldsymbol{\xi}), \sup_{\boldsymbol{\xi}\in\Xi} f_0(\boldsymbol{\xi})\right]$ *and* $g(\boldsymbol{x}, \boldsymbol{\xi}, t) := (t - f(\boldsymbol{x}, \boldsymbol{\xi}))^+ - (t - f_0(\boldsymbol{\xi}))^+$. *Moreover, it can be reformulated as a multistage robust LP when $\mathcal{X}$ and Ξ are polyhedral.*

- "2.5"-stage robust optimization problem: $\min_{\boldsymbol{x}}\text{-}\sup_{t}\text{-}\inf_{\boldsymbol{\lambda},\boldsymbol{q}}\text{-}\sup_{\boldsymbol{\xi}}$.
- Remark: use a type-1 Wasserstein metric with $d(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2) := \|\boldsymbol{\xi}_1 - \boldsymbol{\xi}_2\|_p$, $p \in \{1, \infty\}$.

Conservative Approximation via Finite Adaptability

Given the partitions $\mathcal{T}_k$ such that $\bigcup_{k\in[K]} \mathcal{T}_k = \bar{\mathcal{T}}$,

piecewise static policy: $\boldsymbol{\lambda}(t) = \sum_{k\in[K]} \lambda_k \mathbf{1}\{t \in \mathcal{T}_k\}_{k=1}^K$ and,

piecewise linear policy: $\boldsymbol{q}_i(t) = \sum_{k\in[K]} (\bar{q}_{ik} + q_{ik}t)\mathbf{1}\{t \in \mathcal{T}_k\}_{k=1}^K$.

DRSSDCP (5) can be conservatively approximated by

$$\underset{\boldsymbol{x}\in\mathcal{X},\boldsymbol{\lambda}\geq 0,\boldsymbol{q},\bar{\boldsymbol{q}}\in\mathbb{R}}{\text{minimize}} \quad \boldsymbol{c}^\top \boldsymbol{x} \tag{6a}$$

$$\text{subject to} \quad \lambda_k \epsilon + \frac{1}{M}\sum_{i=1}^{M}(\bar{q}_{ik} + q_{ik}t) \leq 0 \qquad \forall t \in \mathcal{T}_k, k \in [K] \tag{6b}$$

$$g(\boldsymbol{x},\boldsymbol{\xi},t) - \lambda_k\|\boldsymbol{\xi} - \hat{\xi}_i\| \leq \bar{q}_{ik} + q_{ik}t \quad \forall \boldsymbol{\xi} \in \Xi, t \in \mathcal{T}_k, \forall i, k. \tag{6c}$$

Proposition

Problem (6) admits an upper bound for problem (5). Under assumptions 1 and 2, it is equivalent to a finite-dimensional convex optimization problem, which further admits a LP if $\mathcal{X}$ and Ξ are polyhedral.

Lower Bounding Approximation via Finite Scenarios

Given a finite scenarios set $\hat{\mathcal{T}} := \{\hat{t}_1, \cdots, \hat{t}_{k'}, \cdots, \hat{t}_{K'}\}$, consider the following optimziation problem,

$$\begin{aligned}
\underset{\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{q}}{\text{minimize}} \quad & \boldsymbol{c}^\top \boldsymbol{x} && && \text{(7a)}\\
\text{subject to} \quad & \lambda_{k'}\epsilon + \frac{1}{M}\sum_{i\in[M]} q_{ik'} \le 0 && \forall k' \in [K'] && \text{(7b)}\\
& \sup_{\boldsymbol{\xi}\in\Xi} g(\boldsymbol{x}, \boldsymbol{\xi}, \hat{t}_{k'}) - \lambda_{k'}\|\boldsymbol{\xi} - \hat{\xi}_i\| \le q_{ik'} && \forall i \in [M], k' \in [K'] && \text{(7c)}\\
& \boldsymbol{x} \in \mathcal{X};\ \boldsymbol{\lambda} \ge 0; \boldsymbol{q} \in \mathbb{R}^{M\times K'}. && && \text{(7d)}
\end{aligned}$$

Proposition

Problem (7) provides a lower bound for problem (5). Under assumptions 1 and 2, problem (7) is equivalent to a finite-dimensional convex optimization problem, which further admits a LP representative problem if $\mathcal{X}$ and Ξ are polyhedral.

Iterative Partition based Solution Scheme (I)

Algorithm 1 Iterative Partition based Solution Algorithm

Initialize: $\bar{\mathcal{T}} := \left[\inf_{\boldsymbol{\xi}\in\Xi} f_0(\boldsymbol{\xi}), \sup_{\boldsymbol{\xi}\in\Xi} f_0(\boldsymbol{\xi})\right]$; Tlimit, Iter, θ.
Initialize: $\mathsf{LB}^0 = -\infty$, $\mathsf{UB}^0 = +\infty$, $\mathscr{P}^1 := \{\bar{\mathcal{T}}\}, \ell = 1$, $\hat{\mathcal{T}}^0 := \{\hat{t}_0\}$.
while $time \leq$ Tlimit or $\ell <$ Iter or $|(\mathsf{UB}^{\ell-1} - \mathsf{LB}^{\ell-1})/\mathsf{UB}^{\ell-1}| > \theta$ **do**
    Solve the upper bound problem (6) with the partitions $\mathscr{P}^\ell$.
    Record $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{q}^*, \bar{\boldsymbol{q}}^*)^\ell$ and $\mathsf{UB}^\ell := {z_{ub}^{\ell}}^*$.
    Calculate the active scenarios set $\hat{\mathcal{A}}^\ell$ for all members of $\mathscr{P}^\ell$.
    Construct the finite scenarios set $\hat{\mathcal{T}}^\ell \leftarrow \hat{\mathcal{A}}^\ell \bigcup \hat{\mathcal{T}}^{\ell-1}$.
    Solve lower bound problem (7) with $\hat{\mathcal{T}}^\ell$ and record $\mathsf{LB}^\ell := z_{lb}^*$
    Update the partitions $\mathscr{P}^{\ell+1} \leftarrow \mathscr{P}^\ell$ by using $\hat{\mathcal{A}}^\ell$, and $l := l + 1$.
return optimal objective value z^* and optimal solution $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{q}^*, \bar{\boldsymbol{q}}^*)$.

Given $\hat{\mathcal{T}}^{\ell+1} \supseteq \hat{\mathcal{T}}^\ell$, then $\mathsf{LB}^{\ell+1} \geq \mathsf{LB}^\ell$; and for all $\mathcal{T} \in \mathscr{P}^{\ell+1}$, there exists $\mathcal{T}' \in \mathscr{P}^\ell$ such that $\mathcal{T} \subseteq \mathcal{T}'$, then $\mathsf{UB}^\ell \geq \mathsf{UB}^{\ell+1}$.

Iterative Partition based Solution Scheme (II)

- Deriving the active scenarios set: $\hat{\mathcal{A}} = \bigcup_{k\in[K]} \left\{\hat{\mathcal{A}}_k^1 \bigcup \hat{\mathcal{A}}_k^2\right\}$.

$$\hat{\mathcal{A}}_k^1 := \arg\min_{t\in\mathcal{T}_k} \left\{-\lambda_k^*\epsilon - \frac{1}{M}\sum_{i\in[M]}(\bar{q}_{ik}^* + q_{ik}^* t)\right\}, \quad (8)$$

$$\hat{\mathcal{A}}_k^2 := \arg\min_{t\in\mathcal{T}_k,\boldsymbol{\xi}\in\Xi} \left\{\bar{q}_{ik}^* + q_{ik}^* t - g_n(\boldsymbol{x}^*, \boldsymbol{\xi}, t) + \lambda_k^*\|\boldsymbol{\xi} - \hat{\xi}_i\|, \;\; \forall i, n'\right\}. \quad (9)$$

- Deriving the partitioning sets via a nested Voronoi diagram partition
  Given partitions $\mathscr{P}$ and active scenarios $\hat{t} \in \hat{\mathcal{A}}$, a new partition is given by

$$\mathcal{V}\left(\mathscr{P}, \hat{\mathcal{A}}\right) = \bigcup_{\mathcal{T}\in\mathscr{P}} \bigcup_{\hat{t}\in\hat{\mathcal{A}}} \left(\mathcal{T} \cap \left\{t \mid \|\hat{t} - t\|_2 \le \|\hat{t}' - t\|_2, \;\; \hat{t} \neq \hat{t}'\right\}\right),$$

  where we perform the two union operators in sequence for all the members of $\mathscr{P}$ and $\hat{\mathcal{A}}$.

Application to Portfolio Selection

Consider DRSSD constrained portfolio selection problem:

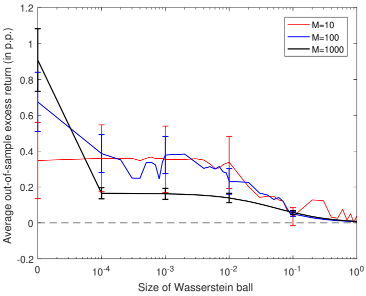
$$\begin{aligned} \underset{\boldsymbol{x}\in\mathcal{X}}{\text{maximize}} \quad & \mathbb{E}_{\hat{\mathbb{P}}}[\boldsymbol{\xi}]^\top \boldsymbol{x} \\ \text{subject to} \quad & \sup_{\mathbb{P}\in\mathcal{P}^1_{\mathsf{W}}(\hat{\mathbb{P}},\epsilon)} \mathbb{E}_{\mathbb{P}}\left[(t-\boldsymbol{\xi}^\top\boldsymbol{x})^+ - (t-\boldsymbol{\xi}^\top\boldsymbol{x}_0)^+\right] \le \phi \qquad \forall t\in\mathbb{R}, \end{aligned}$$

where $\mathcal{X} := \left\{\boldsymbol{x}\in\mathbb{R}^{\mathcal{J}} \;\middle|\; \sum_{j\in[\mathcal{J}]} x_j = 1, x_j \ge 0, \forall j\in[\mathcal{J}]\right\}$.

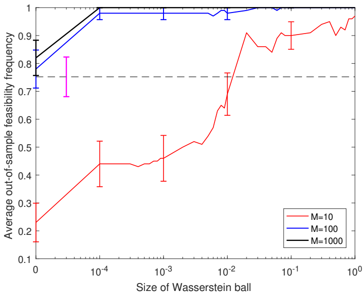
- $\boldsymbol{x}_0$ is the reference portfolo.
- assume Ξ to be interval, i.e. $\Xi := \{\boldsymbol{\xi}^- \le \boldsymbol{\xi} \le \boldsymbol{\xi}^+\}$;
- restrict a small positive ϕ to satisfy Slater's condition, e.g. [Guo et al. 2017, Chen and Jiang. 2018, Hu et al. 2012];
- Numerical experiments with synthetic data and real stock data.

Synthetic Data: Out-of-Sample Performance

- Consider 3 stocks with $M \in \{10, 100, 1000\}$ training samples and 10,000 testing samples from log-normal distribution; reference portfolio $\boldsymbol{x}_0 := [1, 0, 0]$.
- All the performance is averaged over 100 runs.



(a)

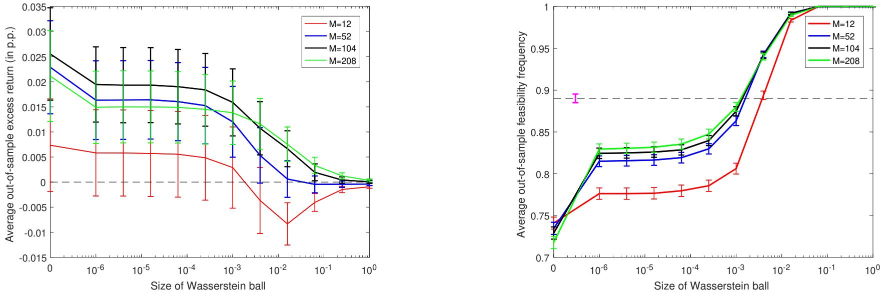


(b)

- out-of-sample feasibility frequency shows the proportion of out-of-sample problems to be feasible from SSD feasibility viewpoint.
- the dashed line in (b) shows the average probability that is achieved by sampling from the reference distribution that satisfies the DRSSD constraint.

Real Stock Data: Out-of-Sample Performance

- randomly choose 5 stocks of S&P 500 [Delage and Li. 2018] with continuous $M \in \{12, 52, 104, 208\}$ weekly stock returns as in-sample data and the subsequent 26 weeks' return as testing data.
- $\boldsymbol{x}_0$: use the equally weighted portfolio as the benchmark.
- All the performance is averaged over 10,000 runs.



(a) (b)

- the bars shows the 90% confidence interval.

Concluding Remarks

Main Takeaways:

- we reformulate data-driven DRSSDCP as a multistage robust optimization problem under mild conditions, and further propose a partition-based conservative approximation (CA) and a scenario-based lower bounding problem (LBP).
- we develop an exact global optimization solution scheme by integrating CA and LBP in an iterative algorithm.
- we show how our data-driven DRSSDCP can be used in practice by portfolio selection problems in terms of out-of-sample performance.

Future Work:

- explore the benefits of using our data-driven DRSSDCP with practical applications.
- extend the framework to the case of multivariate stochastic dominance [Hu et al. 2012, Dentcheva and Wolfhagen. 2015].

Bibliography

- Dentcheva, D., Ruszczyński, A. (2003). Optimization with stochastic dominance constraints. *SIAM Journal on Optimization*, 14(2), 548-566.
- Luedtke, J. (2008). New formulations for optimization under stochastic dominance constraints. *SIAM Journal on Optimization*, 19(3), 1433-1450.
- Guo, S., Xu, H., Zhang, L. (2017). Probability approximation schemes for stochastic programs with DRSSD constraints. *Optimization Methods and Software*, 32(4), 770-789.
- Chen, Z., Jiang, J. (2018). Stability analysis of optimization problems with k-th order stochastic and distributionally robust dominance constraints induced by full random recourse. *SIAM Journal on Optimization*, 28(2), 1396-1419.
- Dentcheva, D., Ruszczyński, A. (2010). Robust stochastic dominance and its application to risk-averse optimization. *Mathematical Programming*, 123(1), 85-100.
- Hu, J., Homem-de-Mello, T., Mehrotra, S. (2012). Sample average approximation of stochastic dominance constrained programs. *Mathematical Programming*, 133(1-2), 171-201.
- Sehgal, R., Mehra, A. (2020). Robust portfolio optimization with sSecond order stochastic dominance constraints. *Computers & Industrial Engineering*, 106396.
- Delage, E., Li, J. Y. M. (2018). Minimizing risk exposure when the choice of a risk measure is ambiguous. *Management Science*, 64(1), 327-344.
- Dentcheva D, Wolfhagen E. (2015) Optimization with multivariate stochastic dominance constraints. *SIAM Journal on Optimization*, 25(1): 564-588.

Questions & Comments...

...Thank you!