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

Introduction

Solution Framework

Portfolio Selection

Conclusion
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 $k = 2$.

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

Consider problem with a $k$-th order SD constraint:

\[
\begin{align*}
[S D C P_k] & \quad \text{minimize} \quad \mathbf{c}^\top \mathbf{x} \\
& \quad \text{subject to} \quad f(x, \xi) \preceq_{(k)} f_0(\xi)
\end{align*}
\]

- $f(x, \xi)$ is the random controlled performance function, and $f_0(\xi)$ is the random reference performance function. Let $f_0(\xi)$ be $f(x_0, \xi)$ with $x_0 \in \mathcal{X}$;
- 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 Ruszczynski. 2003], [Luedtke. 2008], [Hu et al. 2012], [Dentcheva and Wolfhagen. 2015], · · ·
- decision-making under uncertainty: SDCPk assumes the full information of probability distribution of $\xi$. 
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:

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

where \(k = 1, 2, \ldots\), and in particular, this relation is referred as DRFSD and DRSSD if \(k = 1\) and \(2\), respectively.

Optimization with DRSD Constraints

Consider the distributionally robust stochastic dominance constrained problem:

\[ \text{minimize} \quad \mathbf{c}^\top \mathbf{x} \]
\[ \text{subject to} \quad f(x, \xi) \succeq^P f_0(\xi) \quad \forall P \in \mathcal{P}. \]

- 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 \( f_0(\xi) \) is finitely known.
Data-Driven DRSSDCP

Consider DRSSDCP in the second-order:

\[
\text{[DRSSDCP]} \quad \min_{x \in X} \mathbf{c}^\top x \quad \text{(3a)}
\]

subject to

\[
f(x, \xi) \succeq_{(2)} f_0(\xi) \quad \forall \mathbb{P} \in \mathcal{P}_W^r(\hat{\mathbb{P}}, \epsilon). \quad \text{(3b)}
\]

Definition (Wasserstein Ambiguity Set)

The type-\( r \) Wasserstein ambiguity set of radius \( \epsilon \) centered at \( \hat{\mathbb{P}} \) is defined by

\[
\mathcal{P}_W^r(\hat{\mathbb{P}}, \epsilon) := \left\{ \mathbb{P} \in \mathcal{M}(\Xi) \mid d_W^r(\mathbb{P}, \hat{\mathbb{P}}) \leq \epsilon \right\},
\]

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

Assumption 1: \( \Xi \) is nonempty compact convex and \( X \) is nonempty convex.

Assumption 2: \( f(x, \xi) \) and \( f_0(\xi) \) are piecewise linear concave in both \( x \) and \( \xi \), e.g.

\[
f(x, \xi) := \min_{n \in [N]} a_n(x)^\top \xi + b_n(x), \quad f_0(\xi) := \min_{n \in [N]} a_n^0 \xi + b_n^0.
\]
Special Cases of DRSSDCP (3)

Proposition (Reduction to SDCP2)

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

Proposition (Reduction to DFSDCP)

\[ \text{DRSSDCP} \ (3) \text{ with } \mathcal{P} := \mathcal{P}_W^r(\hat{\mathcal{P}}, \infty) \text{ reduces to a distribution-free statewise dominance constrained problem (DFSDCP),} \]

\[
\begin{align*}
\text{minimize} & \quad c^\top x \\
\text{subject to} & \quad f(x, \xi) \geq f_0(\xi) \quad \forall \xi \in \Xi.
\end{align*}
\]

Moreover, it can be reformulated as a linear program if \( \mathcal{X} \) and \( \Xi \) are polyhedral.
**Multistage Robust Optimization Reformulation**

**Proposition**

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

\[
\begin{align*}
\text{minimize} & \quad \mathbf{c}^\top \mathbf{x} \\
\text{subject to} & \quad L(\mathbf{x}, t) \leq 0 \quad \forall t \in \bar{T} \\
\text{where} & \quad L(\mathbf{x}, t) := \inf_{\lambda, \mathbf{q}} \lambda \epsilon + \frac{1}{M} \sum_{i=1}^{M} q_i \\
\text{s.t.} & \quad g(\mathbf{x}, \xi, t) - \lambda \|\xi - \hat{\xi}_i\| \leq q_i \quad \forall i \in [M], \xi \in \Xi \\
& \quad \lambda \geq 0, \mathbf{q} \in \mathbb{R}^M,
\end{align*}
\]

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

- “2.5”-stage robust optimization problem: \(\min \sup_{\mathbf{x}} \inf_{t} \sup_{\lambda, \mathbf{q}} \inf_{\xi} \).

- **Remark**: use a type-1 Wasserstein metric with \(d(\xi_1, \xi_2) := \|\xi_1 - \xi_2\|_p, p \in \{1, \infty\}\).
Conservative Approximation via Finite Adaptability

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

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

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

DRSSDCP (5) can be conservatively approximated by

$$\min_{x \in \mathcal{X}, \lambda \geq 0, q, \bar{q} \in \mathbb{R}} \mathbf{c}^\top x \quad (6a)$$

subject to

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

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

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 $\Xi$ are polyhedral.
Lower Bounding Approximation via Finite Scenarios

Given a finite scenarios set \( \hat{T} := \{\hat{t}_1, \ldots, \hat{t}_{k'}, \ldots, \hat{t}_{K'}\} \), consider the following optimization problem,

\[
\begin{align*}
\text{minimize} \quad & \mathbf{c}^\top \mathbf{x} \\
\text{subject to} \quad & \lambda_{k'} \epsilon + \frac{1}{M} \sum_{i \in [M]} q_{ik'} \leq 0 \quad \forall k' \in [K'] \\
& \sup_{\xi \in \Xi} g(\mathbf{x}, \xi, \hat{t}_{k'}) - \lambda_{k'} \|\xi - \hat{\xi}_i\| \leq q_{ik'} \quad \forall i \in [M], k' \in [K'] \\
& \mathbf{x} \in \mathcal{X}; \quad \lambda \geq 0; \quad \mathbf{q} \in \mathbb{R}^{M \times K'}.
\end{align*}
\]

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 \( \Xi \) are polyhedral.
Iterative Partition based Solution Scheme (I)

**Algorithm 1** Iterative Partition based Solution Algorithm

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

Given \( \hat{T}^{l+1} \supseteq \hat{T}^l, \) then \( LB^{l+1} \geq LB^l; \) and for all \( \mathcal{T} \in \mathcal{P}^{l+1}, \) there exists \( \mathcal{T}' \in \mathcal{P}^l \) such that \( \mathcal{T} \subseteq \mathcal{T}', \) then \( UB^l \geq UB^{l+1}. \)
Iterative Partition based Solution Scheme (II)

- Deriving the active scenarios set: \( \hat{A} = \bigcup_{k \in [K]} \{ \hat{A}^1_k \cup \hat{A}^2_k \} \).

\[
\hat{A}^1_k := \arg \min_{t \in T_k} \left\{-\lambda^*_k \epsilon - \frac{1}{M} \sum_{i \in [M]} (\bar{q}^*_ik + q^*_ik t) \right\},
\]

\[
\hat{A}^2_k := \arg \min_{t \in T_k, \xi \in \Xi} \left\{ \bar{q}^*_ik + q^*_ik t - g_n(x^*, \xi, t) + \lambda^*_k \| \xi - \hat{\xi}_i \|, \quad \forall i, n' \right\}.
\]

- Deriving the partitioning sets via a nested Voronoi diagram partition

Given partitions \( \mathcal{P} \) and active scenarios \( \hat{t} \in \hat{A} \), a new partition is given by

\[
\mathcal{V} \left( \mathcal{P}, \hat{A} \right) = \bigcup_{\mathcal{T} \in \mathcal{P}} \bigcup_{\hat{t} \in \hat{A}} \left( \mathcal{T} \cap \left\{ t \| \| \hat{t} - t \|_2 \leq \| \hat{t}' - t \|_2, \quad \hat{t} \neq \hat{t}' \right\} \right),
\]

where we perform the two union operators in sequence for all the members of \( \mathcal{P} \) and \( \hat{A} \).
Application to Portfolio Selection

Consider DRSSD constrained portfolio selection problem:

$$\begin{align*}
\text{maximize} \quad & \mathbb{E}_{\hat{P}}[\xi]^\top x \\
\text{subject to} \quad & \sup_{P \in \mathcal{P}_{W}(\hat{P}, \epsilon)} \mathbb{E}_P \left[ (t - \xi^\top x)^+ - (t - \xi^\top x_0)^+ \right] \leq \phi \quad \forall t \in \mathbb{R},
\end{align*}$$

where $\mathcal{X} := \left\{ x \in \mathbb{R}^J \mid \sum_{j \in [J]} x_j = 1, x_j \geq 0, \forall j \in [J] \right\}$.

- $x_0$ is the reference portfolio.
- Assume $\Xi$ to be interval, i.e. $\Xi := \{ \xi^- \leq \xi \leq \xi^+ \}$;
- Restrict a small positive $\phi$ 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 \( x_0 := [1, 0, 0] \).
- All the performance is averaged over 100 runs.

- 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.
- \( x_0 \): use the equally weighted portfolio as the benchmark.
- All the performance is averaged over 10,000 runs.

- 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


Questions & Comments...

...Thank you!