Multivariate robust second-order stochastic dominance and resulting risk-averse optimization

Jia Liu

Xi'an Jiaotong University, China

TEL:086-82663661, E-mail: jialiu@xjtu.edu.cn

Joint work with Zhiping Chen and Yu Mei (Xi'an Jiaotong University)

(Dec, 17, 2019, Shanghai)

Outline

- Introduction
- Multivariate robust SSD
- Optimization problem with multivariate robust SSD
- Discretization and stability analysis
- Conclusion

Let $(\Omega, \mathscr{F}, P_0)$ denote an abstract probability space. Denote $\mathscr{L}_p = \mathscr{L}_p(\Omega, \mathscr{F}, P_0; R) \ (p \ge 1),$

Definition (FSD)

 $X \in \mathscr{L}_p$ dominates $Y \in \mathscr{L}_p$ in the first order, denoted $X \succeq_{(1)} Y$, if

 $P\{X \le \eta\} \le P\{Y \le \eta\}, \quad \forall \eta \in R$

We define expected shortfall function $F_2(X;\eta) = \int_{-\infty}^{\eta} F(X;\alpha) d\alpha = \mathbb{E}_{P_0}[(\eta - X)_+].$

Definition (SSD)

 $X \in \mathscr{L}_p$ dominates $Y \in \mathscr{L}_p$ in the second order, denoted $X \succeq_{(2)} Y$, if

 $F_2(X;\eta) \le F_2(Y;\eta), \, \forall \eta \in \mathbb{R}$

Second-order stochastic dominance is particularly popular in industry since it models risk-averse preferences.

Proposition

- X ≥₍₁₎ Y iff E_{P0}[u(X)] ≥ E_{P0}[u(Y)] for all u ∈ U₁, here U₁ denotes the set of all nondecreasing functions u: R → R.
- $X \geq_{(2)} Y$ iff $\mathbb{E}_{P_0}[u(X)] \geq \mathbb{E}_{P_0}[u(Y)]$ for all $u \in \mathscr{U}_2$, here \mathscr{U}_2 denotes the set of all concave and nondecreasing functions $u : R \to R$.
- Dentcheva and Ruszczyński (2003) first considered optimization problem with SSD and derived the optimality conditions.
- Dentcheva and Ruszczyński (2006) developed duality relations and solved the dual problem by utilizing the piecewise linear structure of the dual functional
- Luedtke (2008) get new linear formulations for SSD with finite distributed benchmark
- Drapkin, Gollmer, Gotzes, Schultz, et al. (2011a,2011b) study cases where the random variables are induced by mixed-integer linear recourse

Solution methods

- Sampling approachesa are the most popular solution method (see, Dentcheva and Ruszczyński, 2003, Liu, Sun and Xu, 2016)
- Cut plane methods are the most efficient solution algorithm (see, e.g., Rudolf and Ruszczyński, 2003; Homem-de-Mello and Mehrotra, 2009; Sun, Xu, et al., 2013).

Strong application background in finance

• e.g., portfolio selection applications (Dentcheva and Ruszczyński, 2006, Meskarian, Fliege and Xu 2014; Chen, Zhuang, L., 2019)

Our focus:

- Multivariate extensions: compare random vectors
- Distributionally robust counterparts: ambiguous distribution

Multivariate extensions:

compare random vectors in $\mathscr{L}_p^m = \mathscr{L}_p(\Omega, \mathscr{F}, P_0; \mathbb{R}^m)$,

- Define by order relationship between the expected multivariate utility functions (Müller and Stoyan, 2002; Armbruster and Luedtke, 2015).
 - * Random vectors X, Y such that $\mathbb{E}_{P_0}[u(X)] \ge \mathbb{E}_{P_0}[u(Y)]$ for all $u \in \mathscr{U}_2$, here \mathscr{U}_2 denotes the set of all concave and nondecreasing functions $u: \mathbb{R}^n \to \mathbb{R}$.
- Introduce a scalarization function and model as a univariate SD (Dentcheva and Ruszczyński, 2010; Noyan and Rudolf, 2013, 2018)

$$\theta(c, X) \geq_{(2)} \theta(c, Y), \ \forall c \in C.$$

Multivariate SSD

Definition (Multivariate SSD)

Random vector $X \in \mathscr{L}_p^m$ dominates $Y \in \mathscr{L}_p^m$ in the second order with respect to the scalarization function θ and a set *C*, denoted as $X \succeq_{(C)}^{\theta,C} Y$, if

$$\theta(c, X) \ge_{(2)} \theta(c, Y), \ \forall c \in C.$$
 (1)

where θ is the min-biaffine scalarization function, $c \in C \subset R^m$ plays the role of a scalarization vector.

- Linear scalarization function $\theta(c, x) = a^T(c)x + b(c)$ (Dentcheva and Ruszczyński, 2010, together with $C = R_+^m$)
- Min-biaffine scalarization function $\theta(c, x) = \min_{1 \le t \le T} \{a_t^T(c)x + b_t(c)\}$ (Noyan and Rudolf, 2018)

Robust SSD

Ambiguity of the distribution

- Fully distributional information is hardly known in practice
- Estimated distribution is usually imprecise

 \Rightarrow Find solution feasible for all possible distribution (Distributionally robust technique)

Definition (Robust SSD)

A random variable $X \in \mathcal{L}_p$ dominates robustly a random variable $Y \in \mathcal{L}_p$ in the second-order with respect to a set of probability measures Q if

$$\mathbb{E}_{P}[u(X)] \ge \mathbb{E}_{P}[u(Y)], \quad \forall u \in \mathcal{U}, \ \forall P \in Q.$$
(2)

Here \mathscr{U} is the set of concave and nondecreasing utility functions defined above. Denote by $X \geq_{(2)}^{Q} Y$.

- Dentcheva and Ruszczyński (2010) proposed the notion of robust second-order stochastic dominance, investigated the optimization problem with this kind of constraints and derived the corresponding conditions of optimality under different cases.
- Guo, Xu and Zhang (2017) studied the efficient solution method for the problems with robust stochastic dominance constraints.
- Chen and Jiang (2018) studied stability Analysis of Optimization Problems with *k*-th order distributionally robust dominance constraints induced by full random recourse

Uncertainty set is key point in robust optimization

- Box uncertainty (Natarajan et al., 2010)
- Ellipsoidal uncertainty (Ermoliev et al., 1985)
- Known first two order moments (El Ghaoui et al., 2003; Natarajan, Sim 2011; Chen, He, Zhang, 2010)
- Imprecise first two order moments (Delage and Ye, 2010; Cheng and Lisser, 2014)
- Mixture distribution uncertainty (Zhu and Fukushima, 2009)
- Probabilistic distance based uncertainty (Wasserstein distance, Pflug and Wozabal, 2012,2014; Phi-divergence, Ben-Tal et al. 2013, Guan and Jiang, 2017; K-L distance, Hu and Hong, 2014)

Assumption (Assumption 1)

Q is convex, closed, and bounded.

Definition (Multivariate robust SSD)

A random vector $X \in \mathscr{L}_p^m$ dominates robustly a random vector $Y \in \mathscr{L}_p^m$ in the second-order with respect to a set of probability measures Q if

 $\mathbb{E}_{P}[u(\theta(c,X))] \ge \mathbb{E}_{P}[u(\theta(c,Y))], \quad \forall c \in C, \forall u \in \mathscr{U}, \ \forall P \in Q.$ (3)

Denoted shortly by $X \geq_{(2)}^{\theta,C,Q} Y$.

If Q is a singleton set, θ(c, x) = c^Tx and C = R^m₊ ⇒ linearly multivariate SSD

• If
$$m = 1$$
 and $\theta(c, x) = x \implies$ robust SSD

• If m = 1, Q is a singleton set and $\theta(c, x) = x \Longrightarrow$ classical SSD

Our motivation: multivariate version of robust SSD

- Multivariate extensions: compare random vectors
- Distributionally robust counterparts: ambiguous distribution
- Our contributions
 - Study mathematical properties for min-biaffine scalarization
 - Analyze their optimality conditions
 - Examine the approximation scheme with stability results

Introduction Definition Properties Optimization problems Discretization Con

Mathematical properties

We adopt min-biaffine scalarization function (Noyan and Rudolf, 2018OR) $\theta(c, x) = \min_{1 \le t \le T} \{a_t^T(c)x + b_t(c)\}$

Assumption (Assumption 2)

For any fixed $c \in C$, $\theta(c, \cdot)$ is nondecreasing in the sense that $x \geq^{sep} y$ entails $\theta(c, x) \geq \theta(c, y)$.

Holds automatically for portfolio selection optimizations

Lemma (*pth* integrability)

For fixed $c \in C$, $X \in \mathscr{L}_p^m$ implies $\theta(c, X) \in \mathscr{L}_p$.

Lemma (Lipschitz continuity)

For any fixed $c \in C$, there exists a constant $C_0(c) := \max_{1 \le t \le T} ||a_t(c)||_1$ such that for any $X, Y \in \mathscr{L}_p^m$, we have

 $\|\theta(c,X)-\theta(c,Y)\|_p\leq C_0(c)\|X-Y\|_p.$

Relations to utility functions

Theorem

For $X, Y \in \mathscr{L}_p^m$, the following conditions are equivalent:

$$X \geq_{(2)}^{\theta,C,Q} Y;$$

$$\theta(c,X) \geq^{Q}_{(2)} \theta(c,Y), \forall c \in C;$$

$$\Phi = \left\{ \int_C [Q(c)](\theta(c, x))\mu(dc) : \mu \in \mathcal{M}_+(C), Q \colon C \to \mathscr{U}, \text{ such that} \\ (c, x) \to [Q(c)](\theta(c, x)) \text{ is Lebesgue measurable on } C \times \mathbb{R}^m \right\}$$

③ *CVaR*_{*α,P*}(*θ*(*c*, *X*)) ≥ *CVaR*_{*α,P*}(*θ*(*c*, *Y*)), $\forall c \in C, \forall P \in Q, \forall \alpha \in (0, 1],$ where *CVaR*_{*α,P*}(*X*) = sup_η{η − $\frac{1}{\alpha} \mathbb{E}_P[(\eta - X)_+]$ }.

Reformulation

We know from above theorem that $X \succeq_{(2)}^{\theta, \tilde{C}, Q} Y$ is equivalent to

$$\sup_{P \in Q} \mathbb{E}_{P}[(\eta - \theta(c, X))_{+} - (\eta - \theta(c, Y))_{+}] \le 0, \ \forall (c, \eta) \in \tilde{C} \times R.$$

We introduce a functional $\sigma \colon \mathscr{L}_p \to \overline{R}$ defined as

$$\sigma(V) = \sup_{P \in Q} \mathbb{E}_P[V],$$

We define $\rho_{c,\eta} \colon \mathscr{L}_p^m \to \bar{R}$ as

$$\rho_{c,\eta}(X) = \sigma[(\eta - \theta(c, X))_+ - (\eta - \theta(c, Y))_+]$$

 $X \geq_{(2)}^{\theta, \tilde{C}, Q} Y$ is equivalent to

$$\rho_{c,\eta}(X) \leq 0, \ \forall (c,\eta) \in \tilde{C} \times R.$$

Lipschitz continuity of $\sigma(\cdot)$

Lemma

If the set Q is convex, closed and bounded, then $\sigma(\cdot)$ is convex and subdifferentiable everywhere. Moreover for any $V \in \mathscr{L}_p$, we have $\partial \sigma(V) = \{P \in Q : \mathbb{E}_P[V] = \sigma(V)\}$, and $\sigma(\cdot)$ is Lipschitz continuous on \mathscr{L}_p with modulus $B := \sup_{P_1 \in Q} \sup_{P_2 \in Q} \left\| \frac{dP_1}{dP_2} \right\|_q$.

Convexity and Lipschitz continuous of $\rho_{c,\eta}(\cdot)$

Proposition

Given Assumptions 1 and 2, $\rho_{c,\eta}(\cdot)$ has the following properties:

•
$$\rho_{c,\eta}(\cdot)$$
 is convex;

2 $\rho_{c,\eta}(\cdot)$ is nonincreasing in the sense that

 $X_1(\omega) \geq^{sep} X_2(\omega), \forall \omega \in \Omega \Rightarrow \rho_{c,\eta}(X_1) \leq \rho_{c,\eta}(X_2);$

(a) $\rho_{c,\eta}(\cdot)$ is Lipschitz continuous with modulus $B \cdot C_0(c)$.

Subdifferentiability of $\rho_{c,\eta}(\cdot)$

Proposition (subdifferential)

Given Assumptions 1 and 2, for any $(c, \eta) \in \tilde{C} \times R$, the functional $\rho_{c,\eta}(\cdot)$ is continuous and subdifferretiable on \mathscr{L}_p^m , and its subdifferential at a point $X \in \mathscr{L}_p^m$ is

$$\partial \rho_{c,\eta}(X) = \left\{ Q \in \mathscr{L}_q^m : \exists P_{c,\eta} \in \mathcal{A}_{c,\eta}(X), \exists \lambda(\omega) \in D_{c,\eta}(X,\omega) \text{ such that} \\ Q = \int_{\Omega} \lambda(\omega) dP_{c,\eta}(\omega) \right\}$$
(4)
$$= D_{c,\eta}(X, \cdot) \circ \mathcal{A}_{c,\eta}(X).$$

Subdifferentiability of $\rho_{c,\eta}(\cdot)$

where

$$D_{c,\eta}(X,\omega) = \begin{cases} \operatorname{conv}\{-a_{i_1}(c), \cdots, -a_{i_l}(c)\}, \\ \text{if } \theta_c(X(\omega)) < \eta \text{ and } \{i_1, \cdots, i_l\} = \operatorname{argmin}_t \{a_t^T(c)X(\omega) + b_t(c)\}, \\ l = 1, \cdots, T, \\ \operatorname{conv}\{\mathbf{0}, -a_{i_1}(c), \cdots, -a_{i_l}(c)\}, \\ \text{if } \theta_c(X(\omega)) = \eta \text{ and } \{i_1, \cdots, i_l\} = \operatorname{argmin}_t \{a_t^T(c)X(\omega) + b_t(c)\}, \\ l = 1, \cdots, T, \\ \{\mathbf{0}\}, \text{ if } \theta_c(X(\omega)) > \eta. \end{cases}$$

$$\mathcal{A}_{c,\eta}(X) = \partial \sigma[(\eta - \theta_c(X))_+ - (\eta - \theta_c(Y))_+] : \mathcal{L}_p^m \to 2^{\mathcal{L}_q}, \ X \in \mathcal{L}_p^m.$$

 $\theta_c(\cdot) := \theta(c, \cdot). 2^A$ represents the power set of a set *A*.

Introduction Definition Properties Optimization problems Discretization Con

Optimization problems with multivariate robust SSD

Consider an optimization problem with a multivariate robust SSD constraint:

$$\min_{z \in Z_0} \phi(H(z,\xi)) \tag{5}$$

s.t.
$$G(z,\xi) \geq_{(2)}^{\theta,C,Q} Y(\xi),$$
 (6)

Assumption (Assumption 3)

- The uncertainty is exogenous, i.e., ξ ∈ ℒ^l_p does not depend on decision z.
- 2) Z_0 is a nonempty, convex and compact subset of a Banach space \mathscr{Z}
- Solution For almost all ω ∈ Ω, z ↦ [H(z, ξ(ω))] and z ↦ [G(z, ξ(ω))] are continuous and concave mappings, here G is concave in the sense that: G(λz₁ + (1 − λ)z₂, ξ(ω)) ≥^{sep} λG(z₁, ξ(ω)) + (1 − λ)G(z₂, ξ(ω)) for any λ ∈ [0, 1];

φ(·) is a continuous, nonincreasing and convex functional. Y(·) is continuous.

Relaxation to a compact set

Let $\mathfrak{M}: \tilde{C} \rightrightarrows R$ be a multifunction with a nonempty compact graph. In what follows, we only consider (c, η) in the set graph (\mathfrak{M}) , i.e., we consider the following relaxation problem of problem (5)-(6):

$$\min_{z \in \mathbb{Z}_0} \phi(H(z,\xi)) \tag{7}$$

s.t.
$$\rho_{c,\eta}(G(z,\xi)) \le 0, \ \forall (c,\eta) \in \operatorname{graph}(\mathfrak{M}) \subset \tilde{C} \times R.$$
 (8)

The reason for this relaxation is to satisfy the Slater constraint qualification.

Assumption (Assumption 4: uniformity)

There exists a point $\tilde{z} \in Z_0$ such that

$$\max_{P \in \mathcal{Q}} \max_{(c,\eta) \in graph(\mathfrak{M})} \mathbb{E}_{P}[(\eta - \theta(c, G(\tilde{z}, \xi)))_{+} - (\eta - \theta(c, Y(\xi)))_{+}] < 0.$$

Optimality condition

Theorem (Optimality condition)

Given Assumptions 1-4. If \hat{z} is an optimal solution to problem (7)-(8), then there exist measures $\hat{S} \in \partial \phi[H(\hat{z},\xi)]$, $P_{c,\eta} \in \mathcal{A}_{c,\eta}(G(\hat{z},\xi))$, a measurable selection $\lambda_{c,\eta} \in D_{c,\eta}(G(\hat{z},\xi),\omega), (c,\eta) \in \operatorname{graph}(\mathfrak{M}), \omega \in \Omega$, and a measure $\hat{v} \in \mathcal{M}_+(\operatorname{graph}(\mathfrak{M}))$, such that \hat{z} is an optimal solution to the problem

$$\min_{z \in \mathbb{Z}_0} \left\{ \int_{\Omega} H(z,\xi) \hat{S}(d\omega) + \int_{graph(\mathfrak{M})} \int_{\Omega} \lambda_{c,\eta}(\omega) \cdot G(z,\xi) dP_{c,\eta}(\omega) d\hat{\nu} \right\}$$
(9)

and the following complementary condition is satisfied

$$\int_{graph(\mathfrak{M})} \mathbb{E}_{P_{c,\eta}}[(\eta - \theta(c, G(\hat{z}, \xi)))_+] d\hat{\nu} = \int_{graph(\mathfrak{M})} \mathbb{E}_{P_{c,\eta}}[(\eta - \theta(c, Y(\xi)))_+] d\hat{\nu}.$$
(10)

Conversely, if for some $\hat{S} \in \partial \phi[H(\hat{z},\xi)], P_{c,\eta} \in \mathcal{A}_{c,\eta}(G(\hat{z},\xi)), \lambda_{c,\eta}(\omega) \in D_{c,\eta}(G(\hat{z},\xi),\omega)$ and $\hat{v} \in \mathcal{M}_+(graph(\mathfrak{M})), \text{ the optimal solution to problem (9) satisfies (10) and}$

(8), then \hat{z} is an optimal solution to problem (7)-(8).

Discretization and stability analysis: A case study for moment-based uncertainty set

Recall the optimization problem with a multivariate robust SSD constraint:

$$\min_{z \in Z_0} \phi(H(z,\xi))$$

s.t. $\mathbb{E}_P[(\eta - \theta(c, G(z,\xi)))_+ - (\eta - \theta(c, Y(\xi)))_+] \le 0, \ \forall (c,\eta) \in \operatorname{graph}(\mathfrak{M}), \ \forall P \in Q.$
(11)

with moment-based uncertainty set

$$Q = \{P \in \mathscr{P} : \mathbb{E}_P[f(\xi)] \le 0\},\tag{12}$$

 \mathscr{P} to denote the set of all probability measures on (Ξ, \mathscr{B}) , where Ξ is the support set of Ω , assumed to be compact, \mathscr{B} is the Borel sigma algebra on Ξ . $f: \Xi \to R^a$ is a continuous vector-valued functional and *a* is a positive integer.

Discrete approximation

Consider a discrete approximation to the set of probability distributions as

$$Q_N := \left\{ P \in \mathscr{P} : f_P(\cdot) = \sum_{i=1}^N p_i \delta_{\xi^i}(\cdot), \sum_{i=1}^N p_i f(\xi^i) \le 0, \sum_{i=1}^N p_i = 1, p_i \ge 0, i = 1, \dots, N \right\},$$

where $f_P(\cdot)$ is the probability mass functions of measure *P*.

• Obviously $Q_N \subset Q$.

We can now construct an approximation to problem (11) as follows:

$$\min_{z \in Z_0} \phi(H(z,\xi))$$
s.t.
$$\sup_{(c,\eta) \in graph(\mathfrak{M})} \sup_{P \in Q_N} \mathbb{E}_P[h(z,c,\eta,\xi)] \le 0,$$
(13)

where $h(z, c, \eta, \xi) = (\eta - \theta(c, G(z, \xi)))_+ - (\eta - \theta(c, Y(\xi))).$

Stability?

Define

$$v_N(z) := \sup_{\substack{(c,\eta) \in \text{graph}(\mathfrak{M}) \ P \in Q_N}} \sup_{P \in Q_N} \mathbb{E}_P[h(z, c, \eta, \xi)],$$
$$v(z) := \sup_{\substack{(c,\eta) \in \text{graph}(\mathfrak{M}) \ P \in Q}} \sup_{P \in Q} \mathbb{E}_P[h(z, c, \eta, \xi)].$$

• Obviously, $v_N(z) \le v(z)$ as $Q_N \subset Q$.

Convergence of v_N when $N \to \infty$? Properties of v(z)?

- Feasible regions: $\mathcal{F} := \{z \in Z_0 : v(z) \le 0\} \text{ and } \mathcal{F}_N := \{z \in Z_0 : v_N(z) \le 0\}.$
- Optimal solution set: $\vartheta := \min\{\phi(H(z,\xi)) : z \in \mathcal{F}\} \text{ and } S := \{z \in \mathcal{F} : \vartheta = \phi(H(z,\xi))\}.$ $\vartheta_N := \min\{\phi(H(z,\xi)) : z \in \mathcal{F}_N\} \text{ and } S_N := \{z \in \mathcal{F}_N : \vartheta_N = \phi(H(z,\xi))\}.$

Convergence of \mathcal{F}_N and ϑ_N when $N \to \infty$?

Assumption

Assumption (Assumption 5)

- There exists a probability measure $P^* \in \mathscr{P}$ such that $\mathbb{E}_{P^*}[f(\xi)] < 0$;
- O and ξ ∈ Ξ, there exists an index N' ∈ {1, · · · , N} such that ||ξ − ξ^{N'}|| ≤ ε;
- For each $\xi \in \Xi$, every component of $G(z, \xi)$, i.e., $G_i(z, \xi)$, $i = 1, \dots, m$, is Lipschitz continuous with the Lipschitz modulus being $\kappa(\xi)$, i.e., $|G_i(z_1, \xi) G_i(z_2, \xi)| \le \kappa_i(\xi) ||z_1 z_2||_2$, and $\kappa := \sup_{\xi \in \Xi} \sum_{i=1}^m \kappa_i(\xi)$ is finite;
- **③** $C_0 := \sup_{c \in \tilde{C}} C_0(c)$ is finite, where $C_0(c) = \max_{1 \le t \le T} ||a_t(c)||_1$;
- $A_0 := \sup_{x \in \mathbb{G}} A_0(x)$ is finite, where $\mathbb{G} := G(Z_0 \times \Xi) \cup Y(\Xi)$ and $A_0(x) = \max_{1 \le t \le T} ||d_t(x)||_2$.

Qualitative stability results

Lemma

For fixed x, $|\theta(c_1, x) - \theta(c_2, x)| \le A_0(x) \cdot ||c_1 - c_2||_2$.

Theorem

Given Assumptions 1-5 and \mathcal{F} is nonempty,

$$\lim_{N\to\infty} \mathbb{H}(\mathcal{F}_N,\mathcal{F}) = 0;$$

$$im_{N\to\infty} \vartheta_N = \vartheta \ (converges \ uniformly);$$

$$Iim_{N\to\infty} \mathbb{D}(S_N, S) = 0.$$

 $\mathbb{H}(A, B) := \max\{\mathbb{D}(A, B), \mathbb{D}(A, B)\}\$ is the Hausdorff distance between A and B, where $\mathbb{D}(A, B) := \sup_{x \in A} \inf_{y \in B}(x, y)$ is the deviation of A from B

Proposition

 $v(\cdot)$ is Lipschitz continuous on Z_0 with the Lipschitz modulus being $C_{0\kappa}$.

Conclusions

- We study multivariate robust SSD
- Use min-biaffine scalarization
- Analyze optimality conditions
- Examine stability of discrete approximation

Further work

- Multivariate robust SSD is still very hard problem. Many conditions. Efficiently solution method?
- How about date-driven uncertainty set?

Thank you!

ヨト・ヨト・

æ