Distributionally Robust Stochastic Optimization with Fixed Marginals

Rui Gao
CMS 2017, Bergamo, Italy

Joint work with Anton Kleywegt

School of Industrial and Systems Engineering
Georgia Tech
Multi-dimensional data vs. one-dimensional marginals

- Many stochastic optimization problems involve multidimensional, dependent random variables
  - Financial portfolio selection
  - Distributed inventory systems design
  - Joint distribution is hard to obtain exactly, especially for high-dimensional problem ($N < K$)
Multi-dimensional data vs. one-dimensional marginals

- Many stochastic optimization problems involve multidimensional, dependent random variables
  - Financial portfolio selection
  - Distributed inventory systems design
  - Joint distribution is hard to obtain exactly, especially for high-dimensional problem \((N < K)\)
- One-dimensional marginal is easier to estimate
Multi-dimensional data vs. one-dimensional marginals

▶ Many stochastic optimization problems involve multidimensional, dependent random variables
  ▶ Financial portfolio selection
  ▶ Distributed inventory systems design
  ▶ Joint distribution is hard to obtain exactly, especially for high-dimensional problem ($N < K$)

▶ One-dimensional marginal is easier to estimate
  (I) Requires fewer data, more statistical tools
Many stochastic optimization problems involve multidimensional, dependent random variables. Examples include:

- Financial portfolio selection
- Distributed inventory systems design
- Joint distribution is hard to obtain exactly, especially for high-dimensional problems ($N < K$)

One-dimensional marginal is easier to estimate:

1. Requires fewer data, more statistical tools
2. Data streams of different random variables are collected with different frequencies
   - Extra marginal data information

Q: How to make robust decisions when joint data are not known exactly but estimates of marginals are accurate?
Multi-dimensional data vs. one-dimensional marginals

Depending on seasonal demand, United Airlines operates two nonstop Los Angeles–Sydney flights, UA827 and UA839, both scheduled to arrive in the morning. Flight 827 operates only on Mondays, Wednesdays and Saturdays. On the other hand, 839 operates daily; it is scheduled to arrive at the same time each day, and about two hours after 827 when the latter operates. Occasionally, however, 827 is so late that it arrives after 839.

Below are two-vectors indicating the numbers of minutes late for flights 827 and 839 respectively, recorded on a sequence of Mondays, Wednesdays and Saturdays. A negative value indicates that that flight, on that day, arrived early, and a zero value indicates that the flight was on time, to the nearest minute:

\[(30, 4), (865, 116), (-1, 0), (-5, 7), (12, 13), (10, 0), (-5, 20), (0, 15), (32, 58), (15, 85),
(30, 45), (26, 30), (6, 23), (40, 55), (3, 40), (0, -8), (11, 12), (7, 13), (-5, 9), (-11, 6),
(-10, -20).\]

Joint data

The numbers of minutes by which flight 839 was late, on Sundays, Tuesdays, Thursdays and Fridays during the same period, were

\[20, 4, 5, 48, -30, -10, -22, -3, 80, -23, 0, 26, 10, 90, 90, 24, 30, 45, 17, 35, -10,
-1, 30, 5, 18, 0, 40, 16, 6.\]

Extra marginal data

---

Multi-dimensional data vs. one-dimensional marginals

- Many stochastic optimization problems involve multidimensional, dependent random variables
  - Financial portfolio selection
  - Distributed inventory systems design
  - Joint distribution is hard to obtain exactly, especially for high-dimensional problem \((N < K)\)

- One-dimensional marginal is easier to estimate
  1. Requires fewer data, more statistical tools
  2. Data streams of different random variables are collected with different frequencies
     - Extra marginal data information
  3. Limited data sharing
     - No joint distribution information
Multi-dimensional data vs. one-dimensional marginals

- Many stochastic optimization problems involve multidimensional, dependent random variables
  - Financial portfolio selection
  - Distributed inventory systems design
  - Joint distribution is hard to obtain exactly, especially for high-dimensional problem ($N < K$)

- One-dimensional marginal is easier to estimate
  1. Requires fewer data, more statistical tools
  2. Data streams of different random variables are collected with different frequencies
    - Extra marginal data information
  3. Limited data sharing
    - No joint distribution information

- Q: How to make robust decisions when joint data are not known exactly but estimates of marginals are accurate?
Outline

- Literature review
  - Fréchet class and Copula theory
  - Distributionally robust formulation with K-L divergence

- Applications
  - (I) Portfolio selection
  - (II) Copula density estimation
  - (III) Connection with Spearman’s rank correlation
Outline

- Literature review
  - Fréchet class and Copula theory
  - Distributionally robust formulation with K-L divergence
- Distributionally robust formulation with Wasserstein metric
  - Tractable and scalable dual reformulation
Outline

- Literature review
  - Fréchet class and Copula theory
  - Distributionally robust formulation with K-L divergence
- Distributionally robust formulation with Wasserstein metric
  - Tractable and scalable dual reformulation
- Applications
  1. Portfolio selection
  2. Copula density estimation
  3. Connection with Spearman’s rank correlation
Fréchet class and Copula theory

- Fréchet class: all joint distributions with fixed marginals


\[
\min_{x \in X} \sup_{\mu \in P(\Xi)} \left\{ E_{\mu}[\Psi(x, \xi)] : \mu \text{ has marginals } \{ F_k \}_{k=1}^K \right\}
\]

- Copula \( C_\mu \): a distribution function on \([0, 1]^K\) with uniform marginals [Sklar]

\[
F_\mu(\xi_1, \ldots, \xi_K) = C_\mu(F_1(\xi_1), \ldots, F_K(\xi_K)), \quad \forall \xi \in \Xi.
\]
Fréchet class and Copula theory

- Fréchet class: all joint distributions with fixed marginals
- Minimax worst-case approach

\[
\min_{x \in X} \sup_{\mu \in \mathcal{P}(\Xi)} \left\{ \mathbb{E}_{\mu}[\Psi(x, \xi)] : \mu \text{ has marginals } \{F_k\}_{k=1}^K \right\}
\]
Fréchet class and Copula theory

- Fréchet class: all joint distributions with fixed marginals
- Minimax worst-case approach

$$\min_{x \in X} \sup_{\mu \in \mathcal{P}(\Xi)} \{ \mathbb{E}_\mu[\Psi(x, \xi)] : \mu \text{ has marginals } \{F_k\}_{k=1}^K \}$$

- Copula $C^\mu$: a distribution function on $[0, 1]^K$ with uniform marginals

[Sklar]  
$$F^\mu(\xi_1, \ldots, \xi_K) = C^\mu(F_1(\xi_1), \ldots, F_K(\xi_K)), \forall \xi \in \Xi.$$
Fréchet class and Copula theory

For finite-supported data, subcopula is the distribution of component-wise rankings of the data.

\[ (-0.3333, -0.1667) \Rightarrow \left( \frac{4}{20+29}, \frac{2}{20} \right) \]
Fréchet class and Copula theory

- Fréchet class: all joint distributions with fixed marginals
- Minimax worst-case approach

$$\min_{x \in X} \sup_{\mu \in \mathcal{P}(\Xi)} \left\{ \mathbb{E}_\mu[\Psi(x, \xi)] : \mu \text{ has marginals } \{F_k\}_{k=1}^K \right\}$$

- Copula $C^\mu$: a distribution function on $[0, 1]^K$ with uniform marginals

[Sklar] \[ F^\mu(\xi_1, \ldots, \xi_K) = C^\mu(F_1(\xi_1), \ldots, F_K(\xi_K)), \forall \xi \in \Xi. \]

- Minimax over subcopulas

$$\min_{x \in X} \sup_{C} \left\{ \mathbb{E}_\mu[\Psi(x, \xi)] : \mu \text{ has marginals } \{F_k\}_{k=1}^K \text{ and subcopula } C \right\}$$
Worst-case distribution in a Fréchet class may be too extreme

Example. A life insurance model

- $K$ individuals: survival probability $p_1 \leq \cdots \leq p_K$, value of claim $\alpha_k > 0$
Example. A life insurance model

- $K$ individuals: survival probability $p_1 \leq \cdots \leq p_K$, value of claim $\alpha_k > 0$
- Risk $\xi_k$: $\mathbb{P}(\xi_k = 0) = p_k$, $\mathbb{P}(\xi_k = \alpha_k) = 1 - p_k$
Worst-case distribution in a Fréchet class may be too extreme

Example. A life insurance model

- $K$ individuals: survival probability $p_1 \leq \cdots \leq p_K$, value of claim $\alpha_k > 0$
- Risk $\xi_k$: $\mathbb{P}(\xi_k = 0) = p_k$, $\mathbb{P}(\xi_k = \alpha_k) = 1 - p_k$
- $\Psi$ supermodular, e.g., stop-loss $\Psi(\xi) = \max(0, \sum_{i=1}^{K} \xi_k - t)$
Worst-case distribution in a Fréchet class may be too extreme

Example. A life insurance model

- $K$ individuals: survival probability $p_1 \leq \cdots \leq p_K$, value of claim $\alpha_k > 0$
- Risk $\xi_k$: $\mathbb{P}(\xi_k = 0) = p_k$, $\mathbb{P}(\xi_k = \alpha_k) = 1 - p_k$
- $\Psi$ supermodular, e.g., stop-loss $\Psi(\xi) = \max(0, \sum_{i=1}^{K} \xi_k - t)$
- Worst-case distribution is comonotonic

\[
F^{\mu^*}(\xi_1, \ldots, \xi_K) = \min_{1 \leq k \leq K} F_k^{\mu^*}(\xi_k)
\]

$\Leftrightarrow \mathbb{P}^{\mu^*}[\xi_k = \alpha_k \mid \xi_{k+1} = \alpha_{k+1}] = 1.$
Worst-case distribution in a Fréchet class may be too extreme

Example. A life insurance model

- $K$ individuals: survival probability $p_1 \leq \cdots \leq p_K$, value of claim $\alpha_k > 0$
- Risk $\xi_k$: $\mathbb{P}(\xi_k = 0) = p_k, \mathbb{P}(\xi_k = \alpha_k) = 1 - p_k$
- $\Psi$ supermodular, e.g., stop-loss $\Psi(\xi) = \max(0, \sum_{i=1}^{K} \xi_i - t)$
- Worst-case distribution is comonotonic

$$F^\mu_\ast(\xi_1, \ldots, \xi_K) = \min_{1 \leq k \leq K} F^\mu_\ast(\xi_k)$$

$$\Leftrightarrow \mathbb{P}^\mu_\ast[\xi_k = \alpha_k \mid \xi_{k+1} = \alpha_{k+1}] = 1.$$ 

- Particularly, when $p_1 = \cdots = p_K$, either all individuals are alive or they all die
Restrict $\mu$ to a smaller set

$$\{\mu \in \mathcal{P}(\Xi) : \mu \text{ has marginals } \{F_k\}_{k=1}^K, KL(C^\mu, C^\nu) \leq \rho\}$$
Distributionally robust formulation with K-L divergence

- Restrict \( \mu \) to a smaller set

\[
\{ \mu \in \mathcal{P}(\Xi) : \mu \text{ has marginals } \{F_k\}_{k=1}^K, KL(C^\mu, C^\nu) \leq \rho \}
\]

- However, limitations due to \( \text{supp } C^\mu \subset \text{supp } C^\nu \)
Distributionally robust formulation with K-L divergence

- Restrict $\mu$ to a smaller set

$$\left\{ \mu \in \mathcal{P}(\Xi) : \mu \text{ has marginals } \{F_k\}_{k=1}^{K}, \ KL(C^\mu, C^\nu) \leq \rho \right\}$$

- However, limitations due to $\text{supp } C^\mu \subset \text{supp } C^\nu$
  - For data-driven problems with continuous density, w.p.1, K-L ball is a singleton that only contains $C^\nu$

Empirical (sub)copula looks like a Latin hypercube
Distributionally robust formulation with K-L divergence

- Restrict \( \mu \) to a smaller set

\[
\{ \mu \in \mathcal{P}(\Xi) : \mu \text{ has marginals } \{F_k\}_{k=1}^K, \ KL(C^\mu, C^\nu) \leq \rho \}
\]

- However, limitations due to \( \text{supp } C^\mu \subset \text{supp } C^\nu \)

  - For data-driven problems with continuous density, w.p.1, K-L ball is a singleton that only contains \( C^\nu \)

  Empirical (sub)copula looks like a Latin hypercube

  ![Empirical (sub)copula](image)

- Remedy: partition the space into bins, but problematic for high-dimensional problems
Distributionally robust formulation with Wasserstein metric
Wasserstein metric

Let $d$ be a metric on $[0,1]^K$, $p \geq 1$.

$$W^p_p(\mathcal{C}^\mu, \mathcal{C}^\nu) := \min_{\gamma} \left\{ \int_{[0,1]^{2K}} d^p(u,v)\gamma(du,dv) : \gamma \text{ has marginals } \mathcal{C}^\mu, \mathcal{C}^\nu \right\}$$

Example (Transportation problem)

Let $\mathcal{C}^\nu = \sum_{i=1}^N \nu_i \delta_{\nu_i}$, $\mathcal{C}^\mu = \sum_{j=1}^M \mu_j \delta_{\mu_j}$,

$$W^p_p(\mathcal{C}^\mu, \mathcal{C}^\nu) = \min_{\gamma_{i,j} \geq 0} \left\{ \sum_{i,j} \gamma_{i,j} d^p(u^i, v^j) : \sum_j \gamma_{i,j} = \nu_i, \forall i, \sum_i \gamma_{i,j} = \mu_j, \forall j \right\}.$$
Wasserstein metric

Let \( d \) be a metric on \([0, 1]^K\), \( p \geq 1 \).

\[
W_p^p(\mathcal{C}_\mu, \mathcal{C}_\nu) := \min_\gamma \left\{ \int_{[0,1]^2K} d^p(u,v)\gamma(du, dv) : \gamma \text{ has marginals } \mathcal{C}_\mu, \mathcal{C}_\nu \right\}
\]

Example (Transportation problem)

Let \( \mathcal{C}_\nu = \sum_{i=1}^N \nu_i \delta_{v_i} \), \( \mathcal{C}_\mu = \sum_{j=1}^M \mu_j \delta_{u_j} \),

\[
W_p^p(\mathcal{C}_\mu, \mathcal{C}_\nu) = \min_{\gamma_{ij} \geq 0} \left\{ \sum_{i,j} \gamma_{ij} d^p(u_i, v_j) : \sum_j \gamma_{ij} = \nu_i, \forall i, \sum_i \gamma_{ij} = \mu_j, \forall j \right\}
\]

▶ Equivalent expression on \( \Xi \times \Xi \) using pre-metric \( d_F \)

\[
W_p^p(\mathcal{C}_\mu, \mathcal{C}_\nu) = \min_{\gamma \in \mathcal{P}(\Xi \times \Xi)} \left\{ \int_{\Xi^2} d_F^p(\xi, \zeta)\gamma(d\xi, d\zeta) : \gamma \text{ has marginals } \mu, \nu \right\}
\]
Wasserstein metric: advantages over K-L

- No restriction on the support (other than marginal constraints)
Wasserstein metric: advantages over K-L

- No restriction on the support (other than marginal constraints)

- More reasonable quantitative relation for highly correlated copulas

  Example. Three bivariate Gaussians with correlation matrix

  \[
  R_1 = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}, \quad R_2 = \begin{bmatrix} 1 & 0.99 \\ 0.99 & 1 \end{bmatrix}, \quad R_3 = \begin{bmatrix} 1 & 0.9999 \\ 0.9999 & 1 \end{bmatrix}
  \]

<table>
<thead>
<tr>
<th>Distances</th>
<th>Fisher-Rao</th>
<th>K-L</th>
<th>Burg</th>
<th>Hellinger</th>
<th>TV</th>
<th>( W_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( C^{\mu_1}, C^{\mu_2} )</td>
<td>2.77</td>
<td>22.56</td>
<td>1.48</td>
<td>0.69</td>
<td>2.45</td>
<td>0.15</td>
</tr>
<tr>
<td>( C^{\mu_2}, C^{\mu_3} )</td>
<td>3.26</td>
<td>47.20</td>
<td>1.81</td>
<td>0.75</td>
<td>4.42</td>
<td>0.03</td>
</tr>
</tbody>
</table>
Main result: dual reformulation

$$\sup_{\mu \in \mathcal{M}} \{ \mathbb{E}_\mu[\Psi(\xi)] : \mu \text{ has marginals } F_k, \ W_p(\mathcal{C}^\mu, \mathcal{C}^\nu) \leq \rho \}$$

$$= \inf_{\lambda \geq 0} \left\{ \lambda \rho^p + \sum_{k=1}^K \int_{\Xi_k} f_k(t)F_k(dt) + \int_{\Xi} \sup_{\xi \in \Xi} \left[ \Psi(\xi) - \sum_{k=1}^K f_k(\xi_k) - \lambda d_F^p(\xi, \zeta) \right] \nu(d\zeta) \right\}$$

- \(\lambda\): Lagrangian multiplier for \(W_p(\mathcal{C}^\mu, \mathcal{C}^\nu) \leq \rho\)
- \(f_k(t)\): Lagrangian multiplier for \(k\)-th marginal constraint, where \(\Xi_k\) is \(k\)-th projection of \(\Xi\)
- \(d_F\): equivalent (pre)-metric on \(\Xi\), distance between “rankings”

Proof is based on a refined constructive approach, combining ideas in multi-marginal optimal transport theory and variational analysis.
Size reduction

- **Data-driven problem:** \( \nu = \frac{1}{N} \sum_{i=1}^{N} \delta \hat{\xi}_i \), thus
  \[ \Xi_k = \{ \hat{\xi}_i^k, 1 \leq i \leq N \} \]
- **Piecewise-linear convex objective:**
  \[ \Psi(\xi) = \max_{1 \leq m \leq M} a^m \top \xi + b^m \]
- **Separable metric:**
  \[ d^p_F((\hat{\xi}_{j1}, \ldots, \hat{\xi}_{jK}), \hat{\xi}_i) = \sum_{k=1}^{K} d_{F,k}(\hat{\xi}_{jk}^i, \hat{\xi}_{ik}) \]
- **Distance between rankings:**
  \[ d_{F,k}(\hat{\xi}_{jk}^i, \hat{\xi}_{ik}) \]: distance between rankings of \( \hat{\xi}_{jk}^i \) and \( \hat{\xi}_{ik} \)
Size reduction

- Data-driven problem: \( \nu = \frac{1}{N} \sum_{i=1}^{N} \delta \hat{\xi}^i \), thus
  \( \Xi_k = \{ \hat{\xi}_k^i, 1 \leq i \leq N \} \)

- Piecewise-linear convex objective: \( \Psi(\xi) = \max_{1 \leq m \leq M} a^m \top \xi + b^m \)

- Separable metric: \( d_F^p((\hat{\xi}_1^j, \ldots, \hat{\xi}_K^j), \hat{\xi}^i) = \sum_{k=1}^{K} d_{F,k}(\hat{\xi}_k^j, \hat{\xi}_k^i) \)

- \( d_{F,k}(\hat{\xi}_k^j, \hat{\xi}_k^i) \): distance between rankings of \( \hat{\xi}_k^j \) and \( \hat{\xi}_k^i \)

\[
\min_{\lambda \geq 0, f^i_k, y^i, z^{im}_k \in \mathbb{R}} \lambda \rho^p + \frac{1}{N} \sum_{k=1}^{K} \sum_{i=1}^{N} f^i_k + \frac{1}{N} \sum_{i=1}^{N} y^i
\]

\[
y^i \geq b^m + \sum_{k=1}^{K} z^{im}_k, \forall i, m,
\]

\[
z^{im}_k \geq a^m \top \xi_k^j - f^i_k - \lambda d_{F,k}(\hat{\xi}_k^j, \hat{\xi}_k^i), \forall i, m, j, k.
\]

- \( MN(K + 1) \) constraints, while the support is chosen from \( O(\exp(K)) \) potential points
Applications
Portfolio selection

\[
\min_{x \in X} \max_{\mu \in \mathcal{M}} \mathbb{E}_{\mu}[-x^\top \xi] + c \cdot \text{CVaR}_{\mu}^{\alpha}[-x^\top \xi],
\]

▶ Fama-French three-factor model

\[
\xi_k = b_{k1} f_1 + b_{k2} f_2 + b_{k3} f_3 + \epsilon_k
\]

▶ Distribution parameters are estimated using historical three-year daily data of 30 Industry Portfolios

▶ Simulate \( N = 50 \) samples
Portfolio selection

$$\min \max_{x \in X, \mu \in \mathcal{M}} \mathbb{E}_\mu[-x^\top \xi] + c \cdot \text{CVaR}_\mu^\alpha[-x^\top \xi],$$

- Fama-French three-factor model
  \[ \xi_k = b_{k1} f_1 + b_{k2} f_2 + b_{k3} f_3 + \epsilon_k \]

- Distribution parameters are estimated using historical three-year daily data of 30 Industry Portfolios

- Simulate $N = 50$ samples

- Better out-of-sample performance especially when $K > N = 50$
Non-parametric copula density estimation

- \( X = \{ x \in \mathbb{R}_+^{M \times M} : \sum_{k_1} x_{k_1k_2} = \sum_{k_2} x_{k_1k_2} = \frac{1}{M} \} \)

  p.m.f of subcopula on \( M \times M \) grids in \([0, 1]^2\)

- Dataset: extra marginal data, \( M \times M \gg N \)

- Maximum likelihood estimation

  \[
  \min_{x \in X} \mathbb{E}_{C_0}[-\log(x(\xi))]
  \]

- Benchmark\(^a\): total variation penalized likelihood

  \[
  \min_{x \in X} \mathbb{E}_{C_0}[-\log(x(\xi))] + \lambda \sum_{k_1,k_2=1}^M \sqrt{(x_{k_1+1,k_2} - x_{k_1,k_2})^2 + (x_{k_1,k_2+1} - x_{k_1,k_2})^2}
  \]

- DRSO: \( \min_{x \in X} \max_{C \in \mathcal{M}} \mathbb{E}_C[-\log(x(\xi))] \)

\(^a\)Qu, L., Yin, W.: Copula density estimation by total variation penalized likelihood with linear equality constraints. Computational Statistics & Data Analysis 56(2), 384398 (2012)
Non-parametric copula density estimation

Benchmark
Connection with Spearman’s rank correlation

Only marginal information

\[ \omega(\mu) := W_p(C^\mu, \Pi^\mu) \]

where \( \Pi^\mu \) is the independent subcopula

- When \( K = 2 \), \( \omega(\mu) \) satisfies Rényi’s axioms on measure of dependence
Connection with Spearman’s rank correlation

Only marginal information

\[ \omega(\mu) := W_p(C^\mu, \Pi^\mu) \]

where \( \Pi^\mu \) is the independent subcopula

- When \( K = 2 \), \( \omega(\mu) \) satisfies Rényi’s axioms on measure of dependence
- When \( d(u, v) = \|u - v\|_1 \), \( \max_\mu W_p(C^\mu, \Pi^\mu) \) is attained at comonotonic / counter-monotonic distribution
Connection with Spearman’s rank correlation

Only marginal information

$$\omega(\mu) := W_p(C^{\mu}, \Pi^{\mu})$$

where $\Pi^{\mu}$ is the independent subcopula

- When $K = 2$, $\omega(\mu)$ satisfies Rényi’s axioms on measure of dependence
- When $d(u, v) = \|u - v\|_1$, $\max_{\mu} W_p(C^{\mu}, \Pi^{\mu})$ is attained at comonotonic / counter-monotonic distribution
- When $d((u_1, u_2), (v_1, v_2)) = \begin{cases} |u_1 - v_1|, & \text{if } u_2 = v_2, \\ +\infty, & \text{o.w.} \end{cases}$, $W_p(C^{\mu}, \Pi^{\mu})$ equals Spearman’s rank correlation
Connection with Spearman’s rank correlation

Only marginal information

\[ \omega(\mu) := W_p(C^\mu, \Pi^\mu) \]

where \( \Pi^\mu \) is the independent subcopula

- When \( K = 2 \), \( \omega(\mu) \) satisfies Rényi’s axioms on measure of dependence
- When \( d(u,v) = \|u - v\|_1 \), \( \max_\mu W_p(C^\mu, \Pi^\mu) \) is attained at comonotonic / counter-monotonic distribution
- When \( d((u_1,u_2),(v_1,v_2)) = \begin{cases} |u_1 - v_1|, & \text{if } u_2 = v_2, \\ +\infty, & \text{o.w.} \end{cases} \), \( W_p(C^\mu, \Pi^\mu) \) equals Spearman’s rank correlation
- Natural generalization to multivariate distribution
1. A new framework of providing robust yet less conservative decisions for stochastic optimization with fixed marginals
Summary

1. A new framework of providing robust yet less conservative decisions for stochastic optimization with fixed marginals

2. Measure the similarity between dependence structures using Wasserstein metric between subcopulas
Summary

1. A new framework of providing robust yet less conservative decisions for stochastic optimization with fixed marginals

2. Measure the similarity between dependence structures using Wasserstein metric between subcopulas

3. Scalable reformulation using a novel constructive proof, combining ideas from optimal transport theory
Summary

1. A new framework of providing robust yet less conservative decisions for stochastic optimization with fixed marginals

2. Measure the similarity between dependence structures using Wasserstein metric between subcopulas

3. Scalable reformulation using a novel constructive proof, combining ideas from optimal transport theory