The Directional Optimal Transport and its Applications

Ruodu Wang

http://sas.uwaterloo.ca/~wang

Department of Statistics and Actuarial Science

University of Waterloo

School of Management, University of Science and Technology of China

June 2021

Agenda

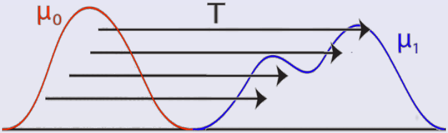
Agenda

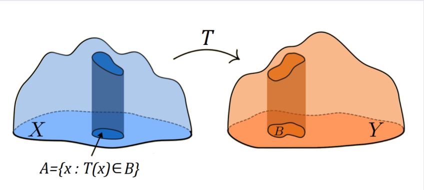
Based on

- Nutz-W., The directional optimal transport. arXiv:2002.08717, 2021, Annals of Applied Probability
- Chen-Lin-W., Ordered risk aggregation under dependence uncertainty. arXiv:2104.07718, 2021

Transport theory

- Pure mathematics theory
- Important applications
  - economics
  - decision theory
  - finance
  - engineering
  - operations research
  - physics
- 1 Nobel Prize laureate
- 2 Fields medalists





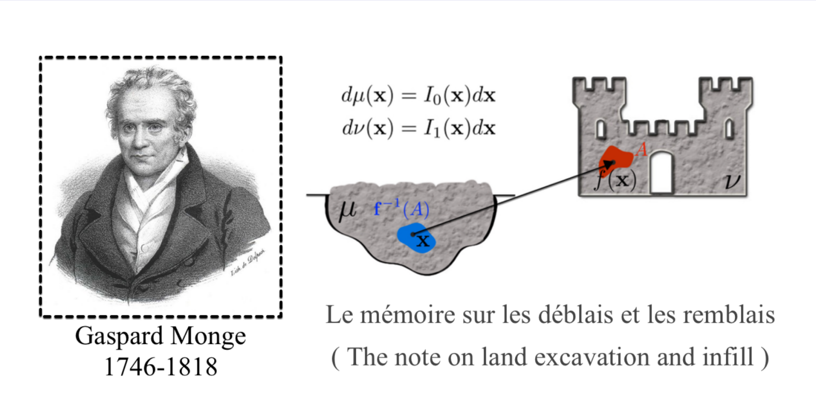
Monge's formulation

- X and Y are two Radon spaces (main example: $\mathbb{R}^d$)
- Cost function $c : X \times Y \to [0, \infty]$ or $(-\infty, \infty]$
- Given probability measures μ on X and ν on Y
- Monge's problem: find a transport map $T : X \to Y$ that attains

$$\inf \left\{ \int_X c(x, T(x)) \, \mathrm{d}\mu(x) \;\middle|\; T_*(\mu) = \nu \right\},$$

  where $T_*(\mu)$ is the push forward of μ by T
- Such T is an optimal transport map

Monge's formulation



Kantorovich's formulation

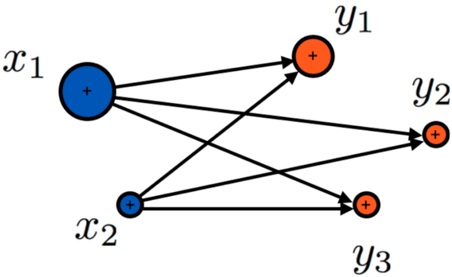
- Monge's formulation may be ill-posed (e.g., point masses)
- Kantorovich's problem: find a probability measure P on $X \times Y$ that attains

$$\inf\left\{\int_{X\times Y} c(x,y)\,\mathrm{d}P(x,y) \mid P \in \Gamma(\mu,\nu)\right\},$$

  where $\Gamma(\mu,\nu)$ is the set of probability measures on $X \times Y$ with marginals μ and ν.
- $X \times Y = \mathbb{R} \times \mathbb{R}$: copulas and dependence
- Discrete version: linear programming

Kantorovich’s formulation

Leonid Kantorovich
1912-1986



Resource allocation

Transport duality

If c is non-negative and lower semi-continuous, then the minimum of the Kantorovich problem is equal to

$$\sup\left(\int_X \varphi(x)\,\mathrm{d}\mu(x) + \int_Y \psi(y)\,\mathrm{d}\nu(y)\right),$$

where the supremum runs over all pairs of bounded and continuous functions $\varphi : X \to \mathbb{R}$ and $\psi : Y \to \mathbb{R}$ such that

$$\varphi(x) + \psi(y) \leq c(x, y).$$

Economic interpretation

- $x \in X$: the vector of characteristics of a worker
- $y \in Y$: the vector of characteristics of a firm
- $\Phi(x, y)$ the economic output generated by worker x matched with firm y
- Social economic-output maximization problem

$$\sup \left\{ \int_{X \times Y} \Phi(x, y) \, \mathrm{d}P(x, y) \mid P \in \Gamma(\mu, \nu) \right\}$$

On the cost function

Assume $X = Y = \mathbb{R}$.

- If c is submodular, i.e.,

$$c(x,y) + c(x',y') \le c(x,y') + c(x',y) \quad \text{for } x \le x' \text{ and } y \le y',$$

  the optimal transport is comonotone. Examples:
  - $c(x,y) = (y-x)^2$
  - $c(x,y) = -\mathbb{1}_{\{(x,y)\le(x_0,y_0)\}}$
  - $c(x,y) = f(x) + g(y) + h(y-x)$ where h is convex
- If c is supermodular, the optimal transport is antitone (counter-monotonic).
- $c(x,y) = \mathbb{1}_{\{y-x>d_0\}}$: probability of transport distance $> d_0$

Probabilistic formulation

New notation: $X \sim \mu$ and $Y \sim \nu$ stand for random variables

- Classic optimal transport (OT)

$$\inf_{X\sim\mu, Y\sim\nu} \mathbb{E}[c(X,Y)]$$

- Martingale optimal transport (MOT) require: $\mu \preceq_{\mathrm{cx}} \nu$

$$\inf_{X\sim\mu, Y\sim\nu} \mathbb{E}[c(X,Y)] : \; X = \mathbb{E}[Y|X]$$

- Supermartingale optimal transport (SMOT) require: $\mu \succeq_{\mathrm{ssd}} \nu$

$$\inf_{X\sim\mu, Y\sim\nu} \mathbb{E}[c(X,Y)] : \; X \geq \mathbb{E}[Y|X]$$

- Directional optimal transport (DOT) require: $\mu \preceq_{\mathrm{st}} \nu$

$$\inf_{X\sim\mu, Y\sim\nu} \mathbb{E}[c(X,Y)] : \; X \leq Y$$

MOT: Beiglböck/Henry-Labordère/Penkner'13 F&S; Beiglböck/Juillet'16 AoP

SMOT: Nutz/Stebegg'18 AoP

Directional transport

- Given $\mu \preceq_{\mathrm{st}} \nu$ on $\mathbb{R}$ and cost $c : \mathbb{R}^2 \to \mathbb{R}$
- $\mathbb{H} = \{(x, y) \in \mathbb{R}^2 : x \le y\}$

Three formulations:

- Monge:

$$\inf \left\{ \int_{\mathbb{R}} c(x, T(x))\mathrm{d}\mu(x) : T_*(\mu) = \nu,\ \ T(x) \ge x \ \forall x \right\}$$

- Kantorovich:

$$\inf \left\{ \int_{\mathbb{R}\times\mathbb{R}} c(x, y)\mathrm{d}P(x, y) : P \in \Gamma(\mu, \nu),\ \ P(\mathbb{H}) = 1 \right\}$$

- Probabilistic:

$$\inf \{\mathbb{E}[c(X, Y)] : X \sim \mu,\ \ Y \sim \nu,\ \ X \le Y\}$$

Smith'83 Comm. Stat.; Rogers'99 JLMS; Arnold/Molchanov/Ziegel'20 JMVA

Directional transport

- If c is submodular, then OT is attained by comonotonicity. Moreover,

$$(X,Y) \text{ is comonotone and } X \preceq_{\mathrm{st}} Y \;\Rightarrow\; X \le Y$$

$\Rightarrow$ the directional constraint is not binding, $\mathrm{OT} = \mathrm{DOT}$

- If c is supermodular, then it is unclear:

$$(X,Y) \text{ is antitone and } X \preceq_{\mathrm{st}} Y \;\not\Rightarrow\; X \le Y$$

$\Rightarrow$ the directional constraint may be binding, $\mathrm{OT} \neq \mathrm{DOT}$

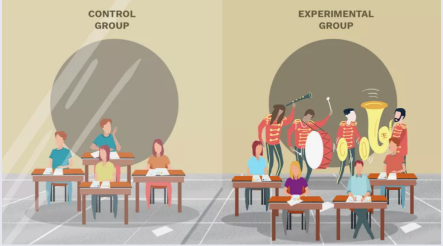
$X \preceq_{\mathrm{st}} Y$ (or $\mu \preceq_{\mathrm{st}} \nu$) means $\mathbb{P}(X \le t) \ge \mathbb{P}(Y \le t)$ for all $t \in \mathbb{R}$.

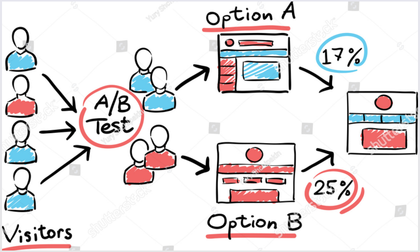
Treatment effect analysis

score X (control)

score Y (experimental)

- Marginals of (X, Y): ✓
- Effect measurement $\mathbb{E}[Y - X]$: ✓
- Dependence of (X, Y): unidentifiable (Neyman'1923)





Treatment effect analysis

- Marginals of (X, Y): ✓
- Effect measurement $\mathbb{E}[Y - X]$: ✓
- Monotone response assumption: $Y \geq X$
  - e.g., Manski'97 ECMA
- Requires $\mathrm{Var}(Y - X)$ for uncertainty quantification
  - No unbiased or consistent estimators
  - Or one may want to know $\mathbb{P}(Y - X > t)$

For $\mathrm{Var}(Y - X)$:

$$\sup \mathrm{Var}(Y - X) \iff \sup \mathbb{E}[(Y - X)^2] \iff \inf \mathbb{E}[XY]$$

$\Longrightarrow$: sup of submodular cost (or inf of supermodular cost)

Goal

- A coupling (transport) P of μ and ν is an element of $\Gamma(\mu,\nu)$
- We will focus on the optimal coupling P_* which maximizes a submodular function c

$$P_* = \operatorname*{arg\,max}_{P \in \Gamma(\mu,\nu)} \left\{ \int_{\mathbb{R}\times\mathbb{R}} c(x,y)\mathrm{d}P(x,y) : P(\mathbb{H}) = 1 \right\}$$

- Primary example: $c(x,y) = (y-x)^2$

Solution in the discrete setting

Simple setting

- $\mu = \frac{1}{n}\sum_{i=1}^{n}\delta_{x_i}$ and $\nu = \frac{1}{n}\sum_{i=1}^{n}\delta_{y_i}$
- $x_1 > \cdots > x_n$
- $y_1, \ldots, y_n$ distinct, $S_1 = \{y_1, \ldots, y_n\}$

The coupling P_* and map T:

- Iterate for $k = 1, \ldots, n$:
  - (i) $T(x_k) := \min\{y \in S_k : y \geq x_k\}$,
  - (ii) $S_{k+1} := S_k \setminus \{T(x_k)\}$.
- The antitone coupling: omit the inequality in (i)
- Such P_* is unique and universal over c

Solution in the discrete setting

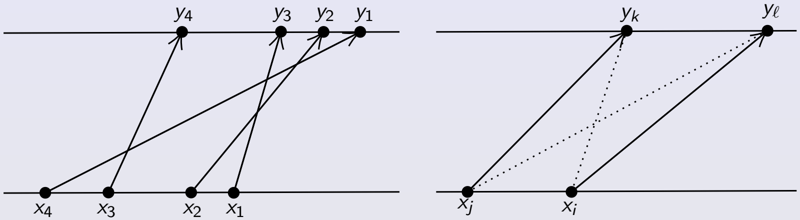


Figure: Left panel: An example of P_*, with the y-axis shown at the top. Right panel: If atomes are not coupled by P_*, then a rearrangement will improve.

Solution in the discrete setting

$$\cdots \underset{7}{\overset{x}{\Bigg(}} \cdots \underset{6}{\overset{x}{\Big(}} \cdots \underset{5}{\overset{x}{(}} \cdots \underset{5}{\overset{y}{)}} \cdots \underset{4}{\overset{x}{(}} \cdots \underset{3}{\overset{x}{(}} \cdots \underset{3}{\overset{y}{)}} \cdots \underset{2}{\overset{x}{(}} \cdots \underset{2}{\overset{y}{)}} \cdots \underset{4}{\overset{y}{)}} \cdots \underset{6}{\overset{y}{\Big)}} \cdots \underset{1}{\overset{x}{(}} \cdots \underset{1}{\overset{y}{)}} \cdots \underset{7}{\overset{y}{\Bigg)}} \cdots$$

- At each position x_i place "(" and at each position y_i place ")"
- P_* couples each "(" to its ")" via the standard rule of algebraic operations
- If one types several "{" and "}" in LaTeX, then P_* describes the way LaTeX processes these curly brackets

Notation

Notation

- μ and ν are probabilities on $\mathbb{R}$ with cdfs F_μ and F_ν
- A coupling P of μ and ν is directional if $P(\mathbb{H}) = 1$
- Denoting by $\mathcal{D} = \mathcal{D}(\mu, \nu)$ the set of all directional couplings
  - $\mathcal{D} \neq \emptyset \iff \mu \preceq_{\rm st} \nu \iff P_{\rm como} \in \mathcal{D}$
  - $\mathcal{D} = \Gamma(\mu, \nu) \iff \sup(\mathrm{supp}\mu) \leq \inf(\mathrm{supp}\nu) \iff P_{\rm anti} \in \mathcal{D}$
- For subprobabilities θ_1, θ_2,
  - $\theta_1 \preceq_{\rm st} \theta_2$ means $\theta_1(\mathbb{R}) = \theta_2(\mathbb{R})$ and $F_{\theta_1} \geq F_{\theta_2}$
  - $\theta_1 \leq \theta_2$ means $\theta_1(A) \leq \theta_2(A)$ for all $A \in \mathcal{B}(\mathbb{R})$.

Assume $\mu \preceq_{\rm st} \nu$ from now on

Characterizing the optimal coupling

Theorem 1

There exists a unique directional coupling $P_* = P_*(\mu, \nu)$ *which couples* $\mu|_{(x,\infty)}$ *to* ν_x *for all* $x \in \mathbb{R}$*, where the subprobability* ν_x *is defined by its cdf*

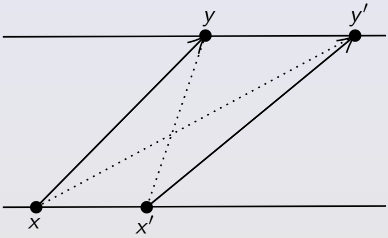
$$F_{\nu_x} = \sup_{\theta \in S_x} F_\theta \quad \text{for} \quad S_x = \{\theta : \mu|_{(x,\infty)} \preceq_{\mathrm{st}} \theta \le \nu\}.$$

The measure ν_x *is the unique minimal element of* S_x *for the order* $\preceq_{\mathrm{st}}$*.*

- Theorem 1 formalizes the intuition in the discrete setting
- P_* is called the directional lower (DL) coupling

Characterizing the optimal coupling

- A pair $((x, y), (x', y')) \in \mathbb{H}^2$ is improvable if $x < x' \leq y < y'$
- (x, y) and (x', y') do not cross, but could be rearranged into $((x, y'), (x', y)) \in \mathbb{H}^2$ which forms a cross



Characterizing the optimal coupling

- c is (μ,ν)-integrable if $|c(x,y)| \le \phi(x)+\psi(y)$ for some $\phi \in L^1(\mu)$ and $\psi \in L^1(\nu)$, implying

$$\sup_{P\in\mathcal{D}} \int c\mathrm{d}P \le \int \phi \mathrm{d}\mu + \int \psi \mathrm{d}\nu < \infty.$$

- For any strictly submodular c,

$$c(x,y)+c(x',y') < c(x,y')+c(x',y)$$

if $((x,y),(x',y'))$ is improvable

Characterizing the optimal coupling

Theorem 2

For a coupling $P \in \mathcal{D}(\mu,\nu)$, the following are equivalent.

(i) $F_P \le F_Q$ *on* $\mathbb{R}^2$ *for all* $Q \in \mathcal{D}(\mu,\nu)$*, where* F_Q *is the cdf of* Q.

(ii) P *is optimal for all* (μ,ν)*-integrable and submodular* c.

(iii) P *is optimal for some* (μ,ν)*-integrable and strictly submodular* c.

(iv) P *is supported by a set* $A \subseteq \mathbb{H}$ *with no improvable pairs.*

(v) $P = P_*$.

Invariance

Corollary 1

Let $\phi : \mathbb{R} \to \mathbb{R}$ be a strictly increasing function. Then

$$P_*(\mu, \nu) = P_*(\mu \circ \phi^{-1}, \nu \circ \phi^{-1}) \circ (\phi, \phi).$$

- Copulas of $P_*(\mu, \nu)$ are precisely those of $P_*(\mu \circ \phi^{-1}, \nu \circ \phi^{-1})$
- For this invariance property, the same transformation must be applied to both axes, in contrast to the classic OT setting

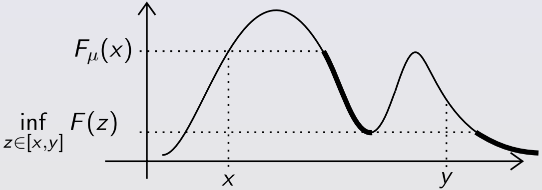
Distribution function

Corollary 2

The cdf of P_* is given by

$$F_*(x,y) = \begin{cases} F_\nu(y) & \text{if } y \le x, \\ F_\mu(x) - \inf_{z\in[x,y]} (F_\mu(z) - F_\nu(z)) & \text{if } y > x. \end{cases}$$

$$F(x) := F_\mu(x) - F_\nu(x)$$



cf. Arnold/Molchanov/Ziegel'20 JMVA

Continuity

Corollary 3

Consider marginals $\mu_n \preceq_{\mathrm{st}} \nu_n$, $n \geq 1$ *with* $\mu_n \xrightarrow{\mathrm{w}} \mu$ *and* $\nu_n \xrightarrow{\mathrm{w}} \nu$, *and suppose that* μ *and* ν *are atomless. Then* $P_*(\mu_n, \nu_n) \xrightarrow{\mathrm{w}} P_*(\mu, \nu)$.

Example 1 (Continuity fails in the presence of atoms)

For $n \in \mathbb{N}$, let μ_n and ν_n be such that $\mu_n\{0\} = \mu_n\{1\} = 1/2$ and $\nu_n\{1 - 1/n\} = \nu_n\{2\} = 1/2$. Then $\mu_n \preceq_{\mathrm{st}} \nu_n$ and $\nu_n \xrightarrow{\mathrm{w}} \nu$ with $\nu\{1\} = \nu\{2\} = 1/2$, and $\mu_n \equiv \mu$ is constant. We see that $P_*(\mu_n, \nu_n)$ is the comonotone coupling, $P_*(\mu, \nu)$ is the antitone coupling, and $P_*(\mu_n, \nu_n) \overset{\mathrm{w}}{\not\to} P_*(\mu, \nu)$.

Decomposition

- The common part $\mu \wedge \nu$ of μ and ν is given by

$$\frac{\mathrm{d}(\mu \wedge \nu)}{\mathrm{d}(\mu + \nu)} := \frac{\mathrm{d}\mu}{\mathrm{d}(\mu + \nu)} \wedge \frac{\mathrm{d}\nu}{\mathrm{d}(\mu + \nu)}.$$

- The mutually singular parts of μ and ν are

$$\mu' = \mu - \mu \wedge \nu \quad \text{and} \quad \nu' = \nu - \mu \wedge \nu$$

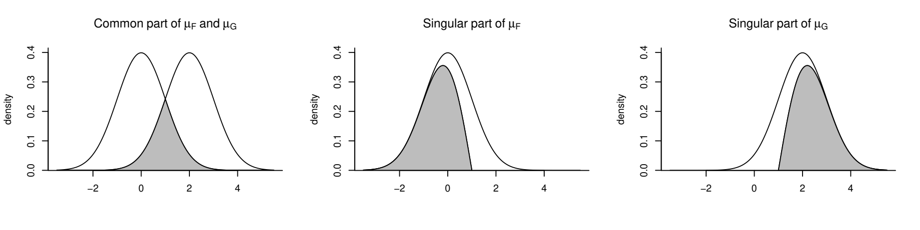
Proposition 1

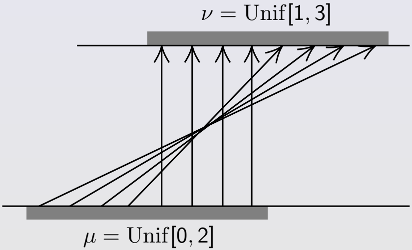
The optimal coupling $P_*(\mu, \nu)$ satisfies

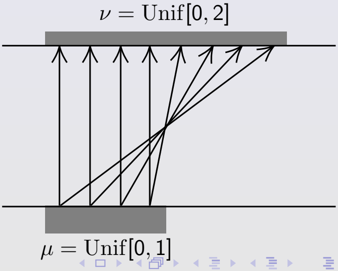
$$P_*(\mu, \nu) = \mathbb{I}(\mu \wedge \nu) + P_*(\mu', \nu')$$

where $\mathbb{I}(\cdot) \in \Gamma(\cdot, \cdot)$ is the identical coupling.

Decomposition







Transport maps

- A coupling P is of Monge-type if $P(Y|X) = T(X)$ is a deterministic function T of X which is then called a Monge map (transport map) of P
- P_* may be randomized (not of Monge-type) even in the absence of atoms, in contrast to the classic OT setting

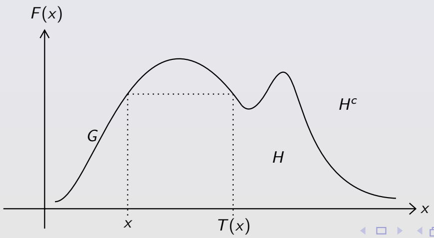
Transport maps

Theorem 3

Let μ, ν be atomless and $\mu \wedge \nu = 0$. Then P_* is of Monge-type with transport map T given by

$$T(x) = \inf\{y \geq x : (y, F(x)) \notin H\}$$

where $F = F_\mu - F_\nu$ and $H = \{(x, z) : z \leq F(x)\}$.



Transport maps

Corollary 4

Let μ, ν be atomless. Then

$$P_*(\mu, \nu) = (\mu \wedge \nu) \otimes_x \delta_x + \mu' \otimes_x \delta_{T(x)}$$

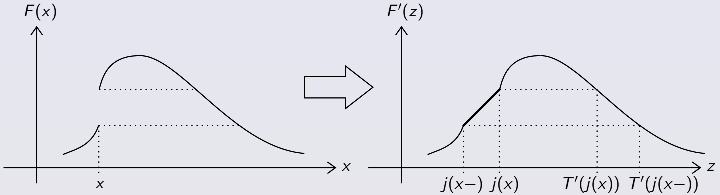
where $\mu' = \mu - \mu \wedge \nu$. In particular, P_* is of Monge-type if and only if μ' and $\mu \wedge \nu$ are mutually singular.

- The "coin flip" is the only source of randomization in P_*

Transport maps

Remarks.

- Previous results hold if ν' has atoms
- If μ' has atoms, it can be addressed by "stretching"



- If T is the map of $P_*(\mu, \nu)$, then $T^\phi := \phi \circ T \circ \phi^{-1}$ is that of $P_*(\mu \circ \phi^{-1}, \nu \circ \phi^{-1})$; in other words, T^ϕ transports $\phi(x)$ to $\phi(y)$ whenever T transports x to y

Distribution function

Corollary 5

We have $H^{\wedge} \leq F_* \leq H^{\vee}$ *where*

$$H^{\wedge}(x,y) = F_\nu(y) - [(F_\mu(y) - F_\mu(x)) \wedge (F_\nu(y) - F_\nu(x))]_+,$$
$$H^{\vee}(x,y) = F_\mu(x) \wedge F_\nu(y).$$

Moreover,

(i) $F_* = H^{\wedge}$ *if and only if* $F = F_\mu - F_\nu$ *is unimodal.*

(ii) $F_* = H^{\vee}$ *if and only if* $\mathcal{D}(\mu,\nu)$ *is a singleton. If, in addition,* F *is continuous, these conditions are further equivalent to* $\mu = \nu$*.*

Decomposition

Proposition 2

Let μ, ν satisfy $\mu \wedge \nu = 0$. Then P_* is the sum of countably many antitone couplings.

Example 2 (Multiple-crossing densities)

Assume that μ and ν are atomless and that $F = F_\mu - F_\nu$ is piecewise monotone (with finitely many pieces). Then P_* is the sum of the identical coupling of $\mu \wedge \nu$ and finitely many antitone couplings between pairs of disjoint intervals.

Other properties

Example 3 (Absence of antitone intervals)

Let μ be the Cantor distribution on $[0,1]$ (i.e. uniform on the Cantor ternary set C) and ν be uniform on $[0,2]$. Clearly $\mu \wedge \nu = 0$.

- Each $x \in C$ can be represented in base 3 as $x = 2\sum_{n=1}^{\infty} x_n 3^{-n}$ where $x_n \in \{0,1\}$. The comonotone transport T_C given by $T_C(x) = 2\sum_{n=1}^{\infty} x_n 2^{-n}$ is directional from μ to ν. Hence, $\mu \preceq_{\rm st} \nu$.
- Assume for contradiction that there exists an interval $[a,b] \subseteq [0,1]$ such that $\mu([a,b]) > 0$ and $T|_{[a,b]}$ is antitone between $\mu|_{[a,b]}$ and its image. There exists c such that $\mu([a,c]) > 0$ and T transports $\mu|_{[a,c]}$ to a distribution on (c,∞). By Theorem 1, T transports $\mu|_{(a,\infty)}$ to a distribution ν_a with $\nu_a([a,c]) > 0$, a contradiction.

Classic OT

- Consider $\mu \preceq_{\mathrm{st}} \nu$ and the classic OT problem

$$\inf_{P \in \Gamma(\mu,\nu)} \int c(|y-x|)\, P(\mathrm{d}x, \mathrm{d}y) \tag{1}$$

- Let $c : \mathbb{R} \to \mathbb{R}_+$ be increasing and concave
  $\Rightarrow c(|y-x|)$ is supermodular on $\mathbb{H}$ but typically not on $\mathbb{R}^2$

Proposition 3

If $F = F_\mu - F_\nu$ is unimodal, then $P_*(\mu,\nu)$ is an optimal coupling for the unconstrained problem (1). If c is strictly concave, the optimizer is unique.

Ordered risk aggregation

- $\mathcal{M}$ is the set of cdfs on $\mathbb{R}$
- $\mathcal{X}$ is a set of random variables in $(\Omega, \mathcal{A}, \mathbb{P})$
- For $F \preceq_{\rm st} G$, let

$$\mathcal{F}_2^o(F,G) = \{(X,Y) : X \sim F,\ \ Y \sim G,\ \ X \le Y\}$$

- $\rho : \mathcal{X} \to \mathbb{R}$ or $\rho : \mathcal{M} \to \mathbb{R}$

Goal: compute

$$\overline{\rho}(\mathcal{F}_2^o(F,G)) := \sup\{\rho(X+Y) : (X,Y) \in \mathcal{F}_2^o(F,G)\}$$

$$\underline{\rho}(\mathcal{F}_2^o(F,G)) := \inf\{\rho(X+Y) : (X,Y) \in \mathcal{F}_2^o(F,G)\}$$

Risk aggregation (OT): W'/Peng/Yang'13 F&S; Embrechts/Wang/W.'15 F&S

Risk aggregation

Lemma 1

For $(X, Y), (X^c, Y^c), (X', Y') \in \mathcal{F}_2^o(F, G)$ *such that* (X^c, Y^c) *is comonotone and* (X', Y') *is DL-coupled, we have*

$$X^c + Y^c \preceq_{\mathrm{cv}} X + Y \preceq_{\mathrm{cv}} X' + Y',$$

where $\preceq_{\mathrm{cv}}$ *is the concave order (i.e.,* $\succeq_{\mathrm{cx}}$*).*

- If ρ is increasing in $\preceq_{\mathrm{cx}}$, then ρ is minimized by the DL coupling and maximized by comonotonicity
  - law-invariant convex risk measures
  - consistent risk measures (Mao/W.'20 SIFIN)
  - $\mathrm{Var}(X + Y)$ and $\mathrm{Var}(Y - X)$

Risk aggregation

- The important cases of $\mathrm{VaR}_p(X+Y)$, $\mathrm{VaR}_p(Y-X)$, $\mathbb{P}(X+Y>t)$ and $\mathbb{P}(Y-X>t)$ are not included
- Goal: compute

$$\overline{\mathrm{VaR}}_p(\mathcal{F}_2^o(F,G)) \quad \text{and} \quad \underline{\mathrm{VaR}}_p(\mathcal{F}_2^o(F,G))$$

where

$$\mathrm{VaR}_p(F) = F^{-1}(p+) = \inf\{t \in \mathbb{R} : F(t) > p\}$$

- More generally,

$$\overline{\rho}(\mathcal{F}_2^o(F,G)) \quad \text{and} \quad \underline{\rho}(\mathcal{F}_2^o(F,G))$$

for a tail risk measure ρ (e.g., VaR, ES, RVaR)

Risk aggregation

- Idea: let $F^{[p,1]}$ be the upper p-tail distribution of F, namely

$$F^{[p,1]}(x) = \frac{(F(x)-p)_+}{1-p}, \quad x \in \mathbb{R}$$

- Let ρ^* be the generator of ρ, i.e., $\rho(F) = \rho^*(F^{[p,1]})$
  - Generator of VaR_p: ess-inf; that of ES_p: $\mathbb{E}$
  - Generators of VaR, ES and RVaR are all increasing in $\preceq_{\mathrm{cv}}$
- Guess: if ρ is monotone, then

$$\overline{\rho}(\mathcal{F}_2^o(F,G)) = \overline{\rho^*}(\mathcal{F}_2^o(F^{[p,1]}, G^{[p,1]}))$$

True in classic OT: Liu/W.'21 MOR

Risk aggregation

- To show

$$\overline{\rho}(\mathcal{F}_2^o(F,G)) = \overline{\rho^*}(\mathcal{F}_2^o(F^{[p,1]},G^{[p,1]}))$$

we need to prove, for any $(X,Y)\in\mathcal{F}_2^o(F,G)$, there exists $(X',Y')\in\mathcal{F}_2^o(F^{[p,1]},G^{[p,1]})$ s.t. $\rho(X+Y)\le\rho^*(X'+Y')$

- Let A be a p-tail event of $X+Y$ (W'/Zitikis'21 MS), then $\rho(X+Y)=\rho^*(\hat{X}+\hat{Y})$ for some $\hat{X}\overset{\mathrm{d}}{=}X|A$ and $\hat{Y}\overset{\mathrm{d}}{=}Y|A$
- We can show $\hat{X}\preceq_{\mathrm{st}}F^{[p,1]}$ and $\hat{Y}\preceq_{\mathrm{st}}G^{[p,1]}$
- In OT: take (X',Y') with the same dependence as $(\hat{X},\hat{Y})$
- In DOT: cannot guarantee $(X',Y')\in\mathcal{F}_2^o(F^{[p,1]},G^{[p,1]})$

The monotone embedding theorem

- For $F, G \in \mathcal{M}$, we say F is smaller than G in strong stochastic order if $G(y) - G(x) \geq F(y) - F(x)$ for all $y \geq x \geq G^{-1}(0)$, denoted by $F \preceq_{\mathrm{ss}} G$.

Proposition 4

The strong stochastic order satisfies the following properties:

(i) *If* $F \leq_{\mathrm{ss}} G$ *then* $F \leq_{\mathrm{st}} G$;

(ii) *Assuming* $F^{-1}(0) = G^{-1}(0)$, $F \leq_{\mathrm{ss}} G$ *if and only if* $F = G$;

(iii) *If* $G^{-1}(0) = -\infty$, *then* $F \leq_{\mathrm{ss}} G$ *means* $F = G$;

(iv) *The relation* $\leq_{\mathrm{ss}}$ *is a partial order.*

The monotone embedding theorem

The problem of monotone embedding: Suppose that $F \le_{\rm st} F' \le_{\rm st} G$ and $(X,Y) \in \mathcal{F}_2^o(F,G)$. The question is whether there exists $X' \sim F'$ such that $X \le X' \le Y$ holds.

Example 4

Let G be the Bernoulli(1/2) distribution. Take $Y \sim G$, let $X = -Y$, and F be the distribution of X. Clearly, (X,Y) is countermonotonic, and hence (X,Y) is DL-coupled. Take another random variable $X' \sim F' = \mathrm{U}[-1,1]$. It is easy to see that $F \le_{\rm st} F' \le_{\rm st} G$. Since $\mathbb{P}(X=Y) = 1/2$ but $\mathbb{P}(X'=Y)=0$, we know that $X \le X' \le Y$ cannot hold for any $X' \sim F'$.

The monotone embedding theorem

- Let $D_*^{F,G}$ be the DL coupling of F and G with transport map $T^{F,G}$

Theorem 4 (Monotone embedding)

Suppose that $F \le_{\mathrm{ss}} F' \le_{\mathrm{st}} G$, *and* $(X, Y) \sim D_*^{F,G}$. *Then there exists* $X' \sim F'$ *such that* $X \le X' \le Y$ *almost surely and* (X', Y) *is DL-coupled.*

Bounds on tail risk measures

Theorem 5

Suppose that $F \le_{\mathrm{st}} G$, $p \in (0,1)$, (ρ, ρ^*) *is a p-tail pair of risk measures, and* ρ^* *is monotone and* $\preceq_{\mathrm{cv}}$*-increasing. We have*

$$\overline{\rho}(\mathcal{F}_2^o(F,G)) = \overline{\rho^*}\left(\mathcal{F}_2^o\left(F^{[p,1]}, G^{[p,1]}\right)\right) = \rho^*(X+Y), \quad (2)$$

where $(X,Y) \sim D_*^{F^{[p,1]}, G^{[p,1]}}$.

VaR bounds

Proposition 5

For continuous distributions F and G such that $F \leq_{\mathrm{st}} G$ and $p \in (0,1)$, we have

$$\overline{\mathrm{VaR}}_p(\mathcal{F}_2^o(F,G)) = \min\left\{ \inf_{x\in[a,b]} \left\{ T^{F^{[p,1]},G^{[p,1]}}(x) + x \right\}, 2b \right\},$$

where $a = F^{-1}(p+)$ and $b = G^{-1}(p+)$.

- The lower bound is also available
- Bounds on left and right VaR are the same if the marginal distributions are continuous and strictly increasing
- Invert bounds on VaR to get bounds on $\mathbb{P}(X+Y>t)$

An application

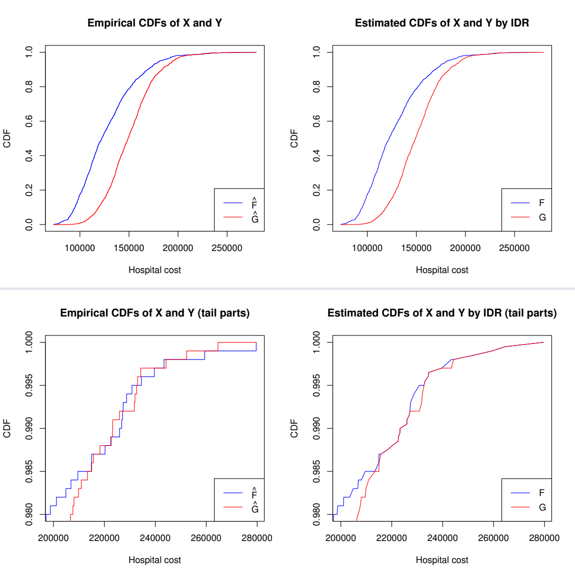
- A health insurance portfolio
- The aggregate loss: $S = X + Y$ where $X \sim F$ and $Y \sim G$ represent the losses caused by females and males, respectively, from a portfolio of 50 males and 50 females
- $F \preceq_{\mathrm{st}} G$ can be verified
- $X \leq Y$ is reasonable due to many common risk factors

An application

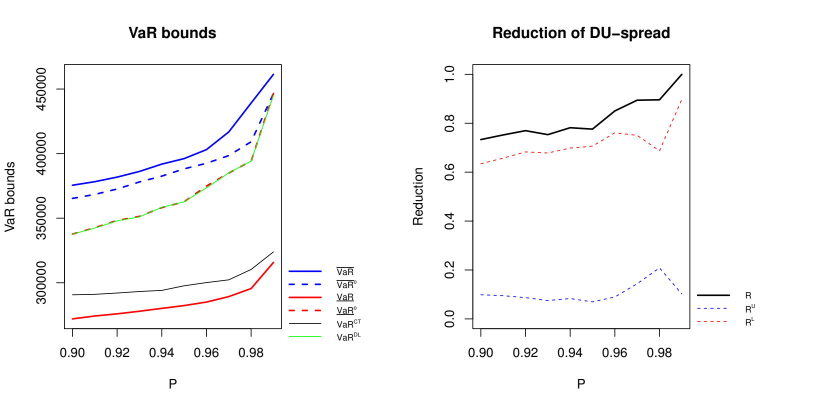
- Data from the Nationwide Inpatient Sample of the Healthcare Cost and Utilization Project (NIS-HCUP)
- 500 observations with 244 males and 256 females
- Testing the bootstrap sample $\hat{F}$ and $\hat{G}$ cannot reject $\hat{F} \preceq_{\rm st} \hat{G}$
- Use the isotonic distributional regression to get F and G with $F \preceq_{\rm st} G$ (Henzi/Ziegel/Gneiting'21 JRSSB)

Data: Frees'09 Regression Modeling with Actuarial and Financial Applications

An application



An application



Future directions

- Higher dimension
- Higher number of marginals
- Statistical inference on variance bounds
    - Aronow/Green/Lee'14 AoS for classic OT (Fréchet-Hoeffding)
- Bounds on $\mathbb{P}(Y - X > t)$

Future directions

Higher dimension.

$$\max_{P \in \mathcal{D}(\mu,\nu)} \int c \mathrm{d}P \quad \text{and} \quad \min_{P \in \mathcal{D}(\mu,\nu)} \int c \mathrm{d}P$$

where μ and ν are probability measures on $\mathbb{R}^d$ and $g : \mathbb{R}^{d+d} \to \mathbb{R}$ is submodular. Possible formulations for $\mathcal{D}$:

- $x \le T(x)$ component-wise
- $T(x) - x \in K$ for some set K (e.g., specific directions)

Future directions

Higher number of marginals.

$$\max_{P \in \mathcal{D}(\mu_1,\dots,\mu_T)} \int c \mathrm{d}P \quad \text{and} \quad \min_{P \in \mathcal{D}(\mu_1,\dots,\mu_T)} \int c \mathrm{d}P$$

where $\mathcal{D}(\mu_1, \dots, \mu_T)$ is the set of directional couplings of $\mu_1, \dots, \mu_T$, and $c : \mathbb{R}^d \to \mathbb{R}$ is submodular.

- Even the unconstrained OT with $T \geq 3$ is very difficult
  - $c(x) = -(x_1 + \cdots + x_T)^2 \Rightarrow$ joint mixability (Wang/W.'16)

Thank you

Thank you for your attention!