Measuring dependence with Wasserstein distances (in Bayesian Nonparametrics)

Hugo Lavenant

Bocconi University

Workshop “Emerging topics in applications of Optimal Transport”, Zurich (Switzerland), June 8, 2023

Joint work with:

Marta Catalano

Antonio Lijoi

Igor Prünster

Joint work with:

Marta Catalano

Antonio Lijoi

Igor Prünster

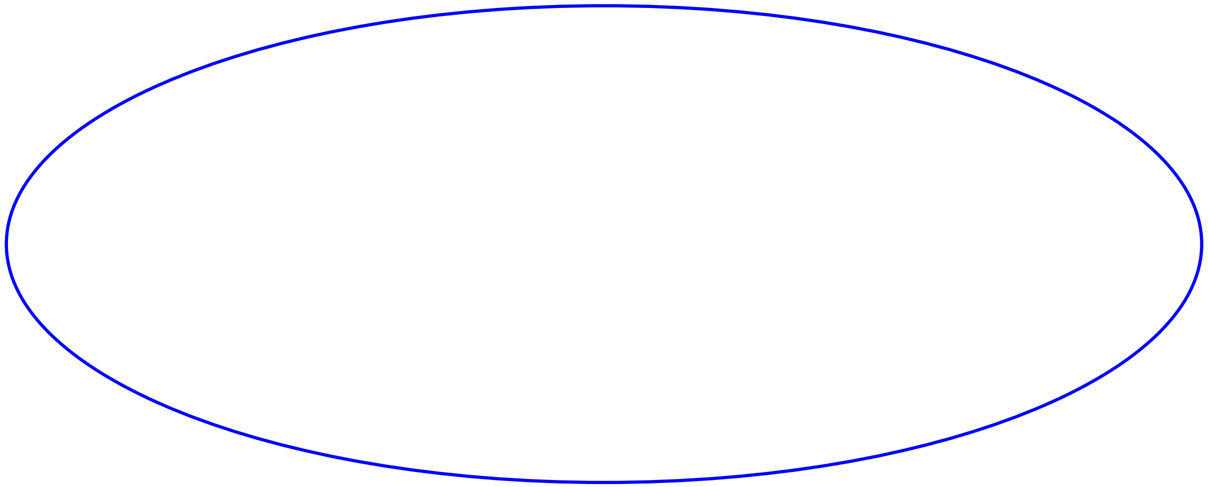
Disclaimers

I am not a (Bayesian) statistician.

My background: mathematical analysis, optimal transport.

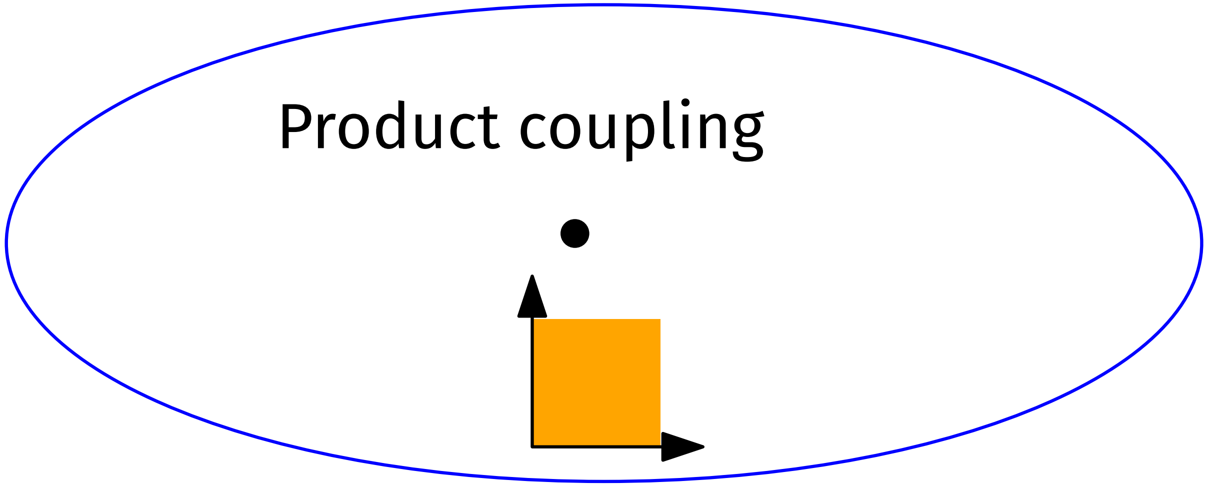
A problem I don't know how to solve

$\Pi(\mathrm{d}x, \mathrm{d}x) = \{$probability on $[0,1]^2$ with uniform marginals$\}$



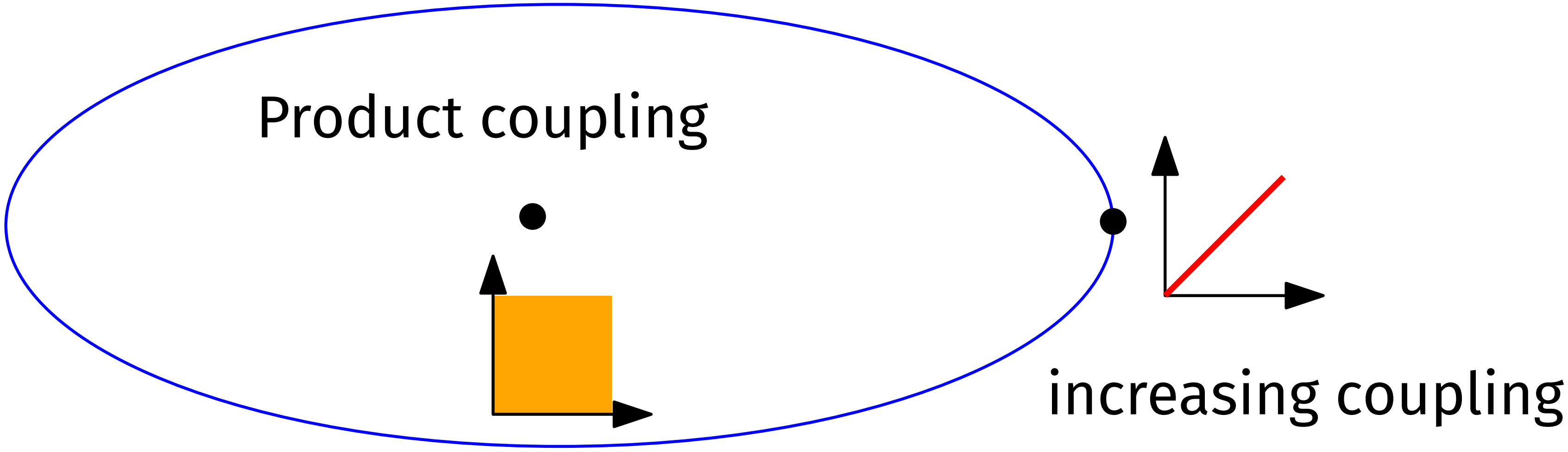
A problem I don't know how to solve

$\Pi(\mathrm{d}x, \mathrm{d}x) = \{$probability on $[0,1]^2$ with uniform marginals$\}$



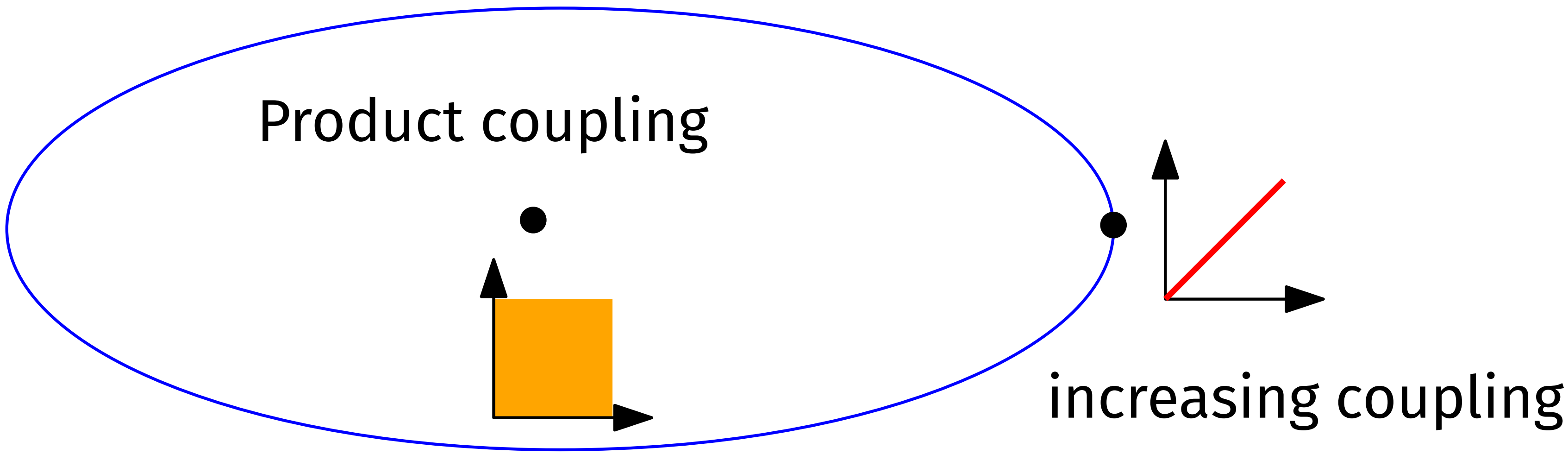
A problem I don't know how to solve

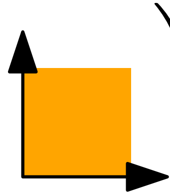
$\Pi(\mathrm{d}x, \mathrm{d}x) = \{$probability on $[0,1]^2$ with uniform marginals$\}$



A problem I don't know how to solve

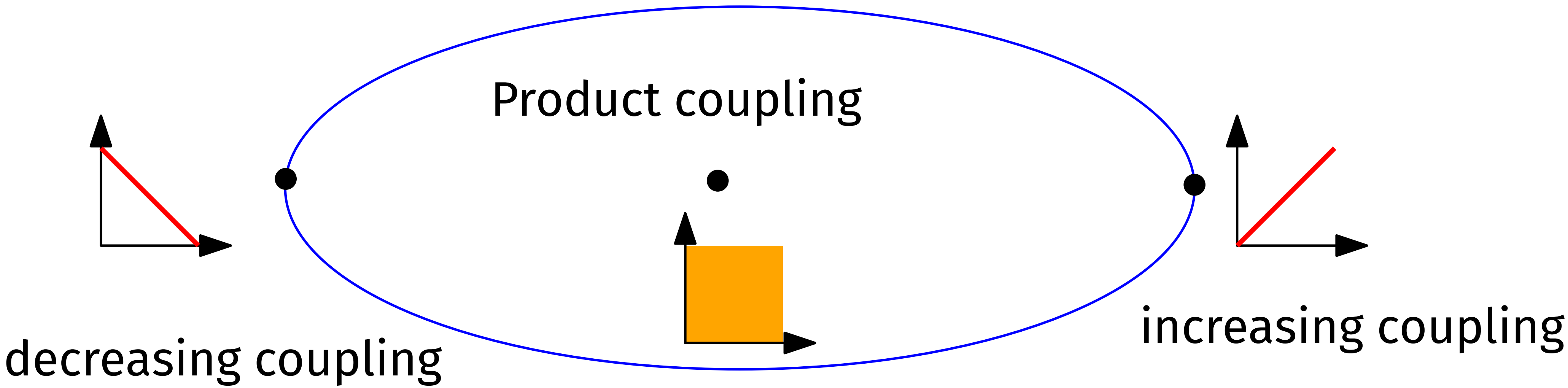
$\Pi(\mathrm{d}x,\mathrm{d}x) = \{$probability on $[0,1]^2$ with uniform marginals$\}$

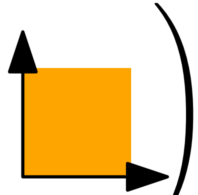


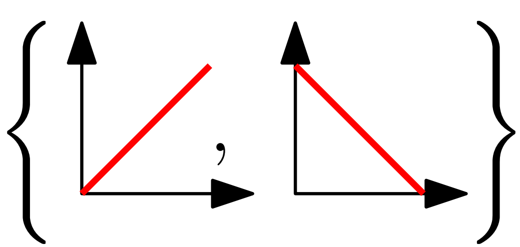
$$\underset{p\in\Pi(\mathrm{d}x,\mathrm{d}x)}{\operatorname{argmax}} W_2\left(p, \right.$$  $$\left.\right) =$$

A problem I don't know how to solve

$\Pi(\mathrm{d}x, \mathrm{d}x) = \{$probability on $[0,1]^2$ with uniform marginals$\}$

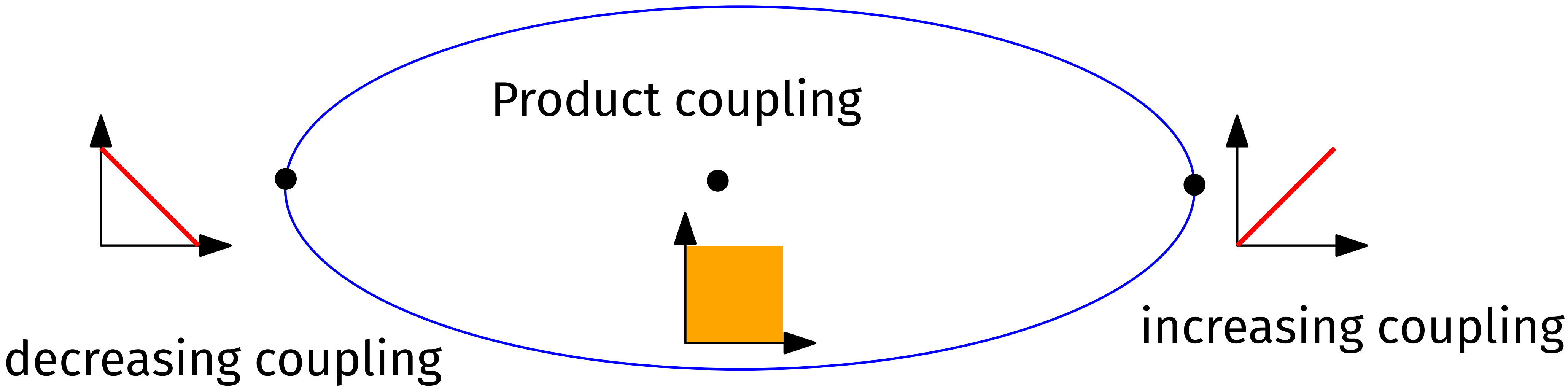


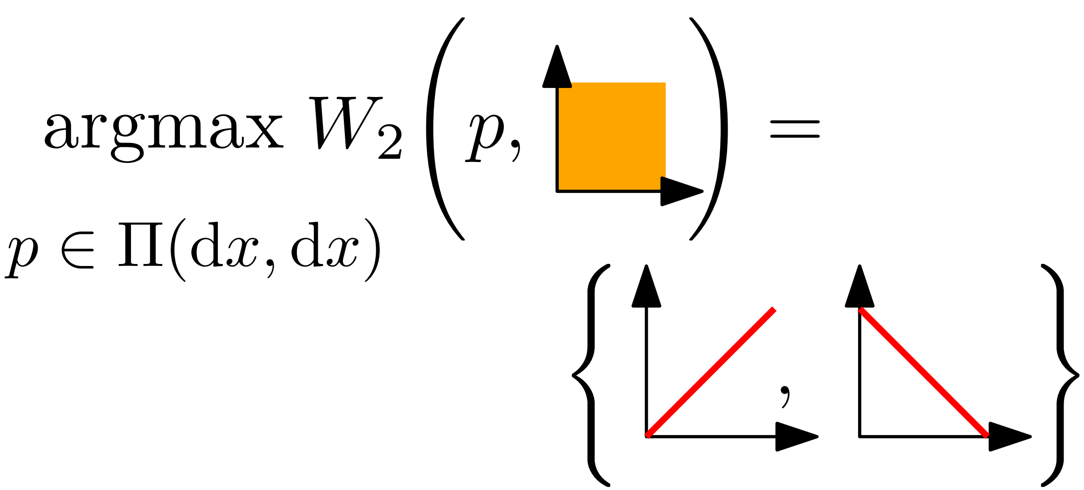
$$\underset{p \in \Pi(\mathrm{d}x, \mathrm{d}x)}{\operatorname{argmax}} W_2\Big(p,$$  $$\Big) =$$



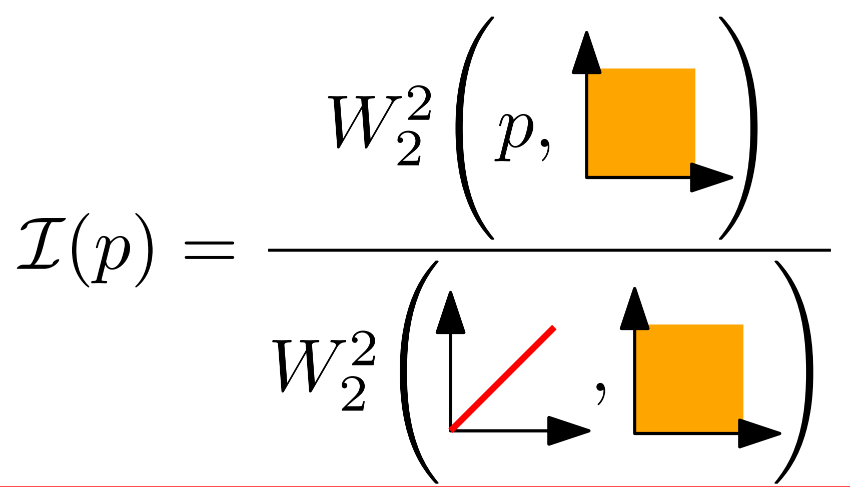
A problem I don't know how to solve

$\Pi(\mathrm{d}x, \mathrm{d}x) = \{$probability on $[0,1]^2$ with uniform marginals$\}$





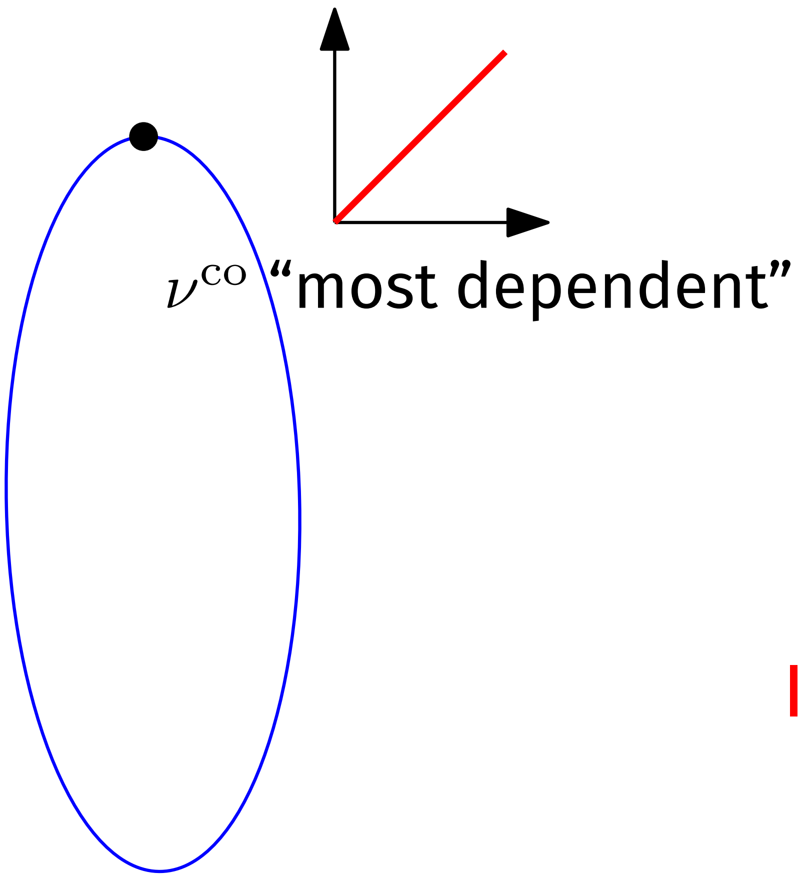
Why? Index of dependence

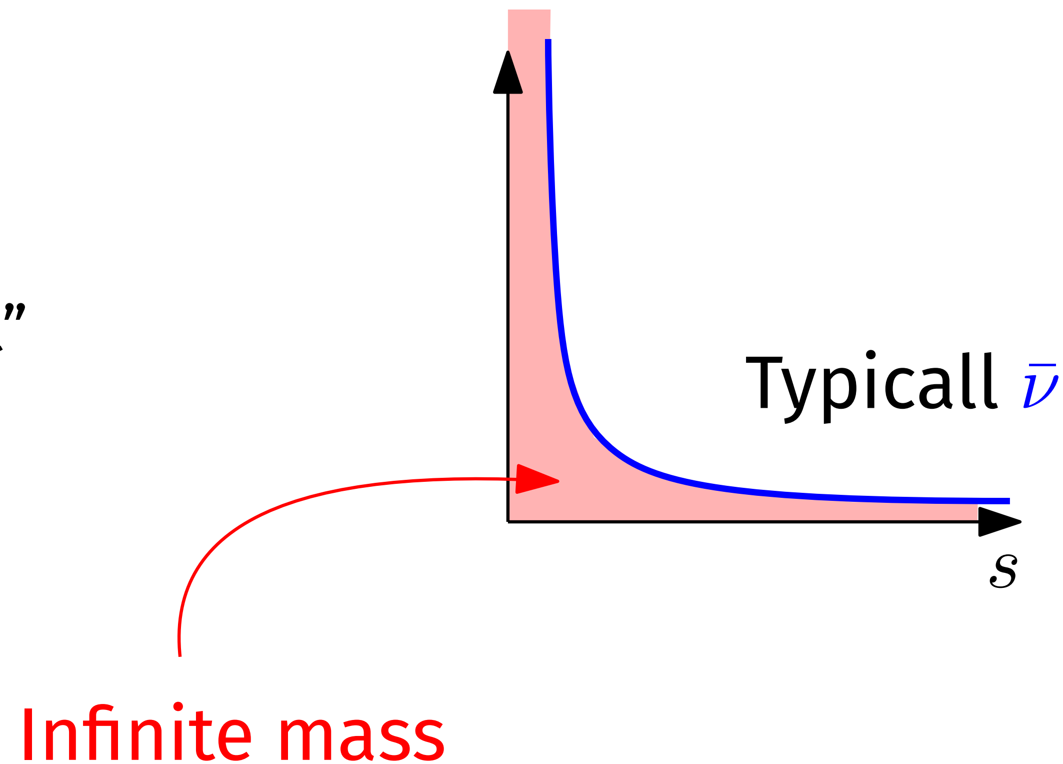


The problem I will solve

$\bar{\nu}$ measure on $(0, +\infty)$ with $\int s^2 \, \mathrm{d}\bar{\nu}(s) < +\infty$

$\Pi(\bar{\nu}, \cdots, \bar{\nu}) = \{$Measure on $[0, +\infty)^d \setminus \{0\}$ with marginals $\bar{\nu}\}$

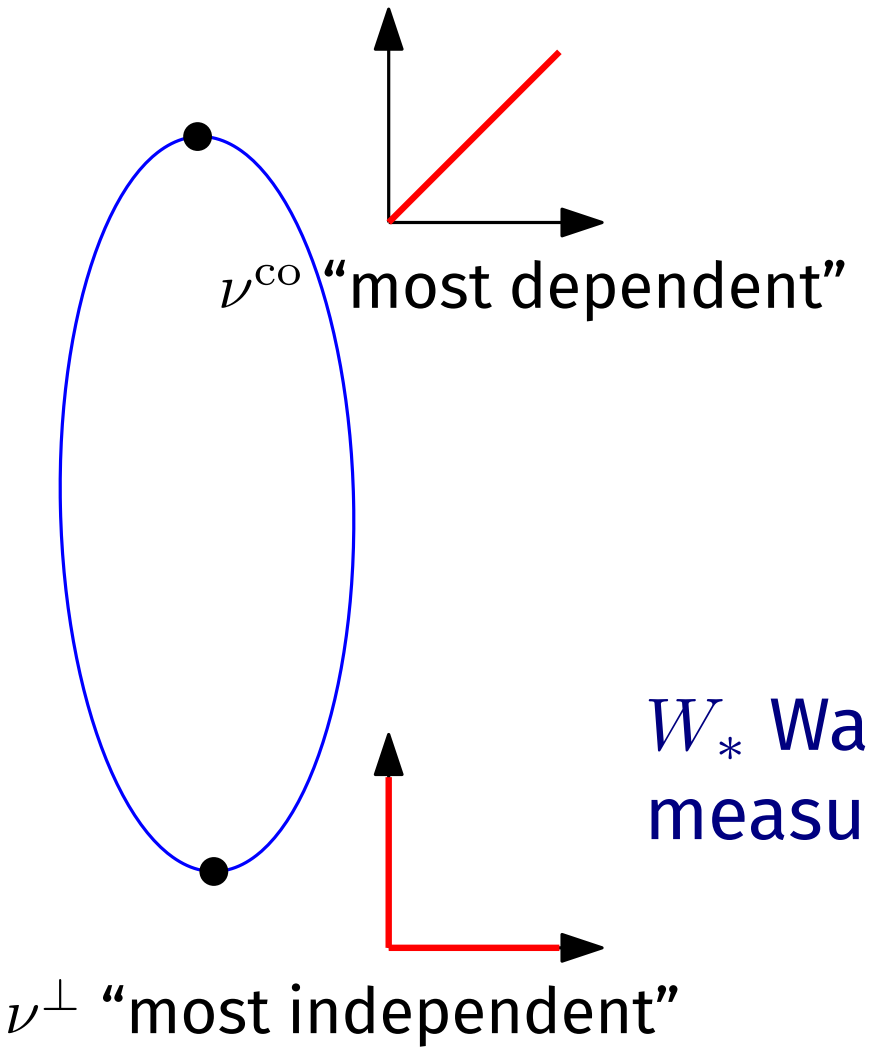


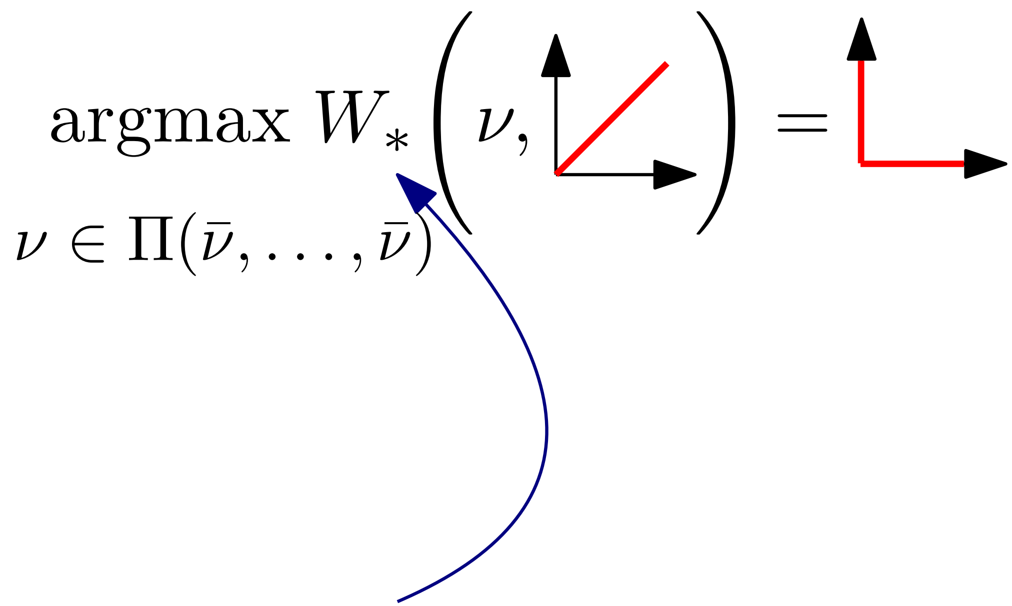


The problem I will solve

$\bar\nu$ measure on $(0,+\infty)$ with $\int s^2 \,\mathrm{d}\bar\nu(s) < +\infty$

$\Pi(\bar\nu,\cdots,\bar\nu) = \{$Measure on $[0,+\infty)^d \setminus \{0\}$ with marginals $\bar\nu\}$



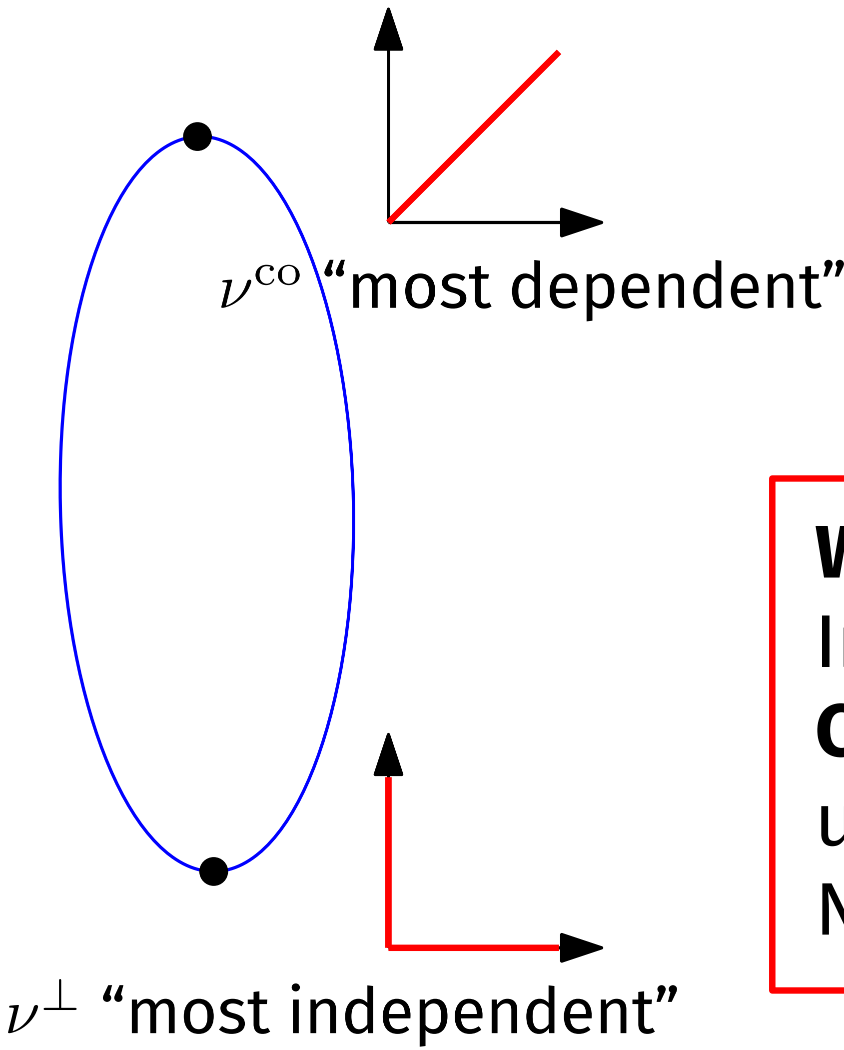


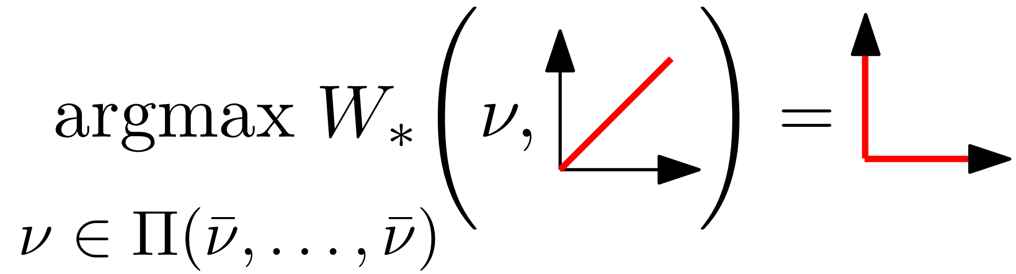
W_* Wasserstein distance between measures of infinite mass

The problem I will solve

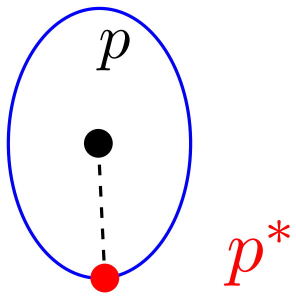
$\bar{\nu}$ measure on $(0,+\infty)$ with $\int s^2 \, \mathrm{d}\bar{\nu}(s) < +\infty$

$\Pi(\bar{\nu},\cdots,\bar{\nu}) = \{$Measure on $[0,+\infty)^d \setminus \{0\}$ with marginals $\bar{\nu}\}$



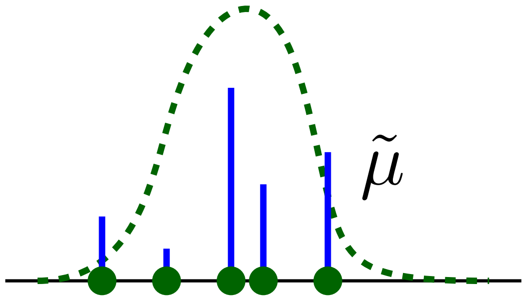


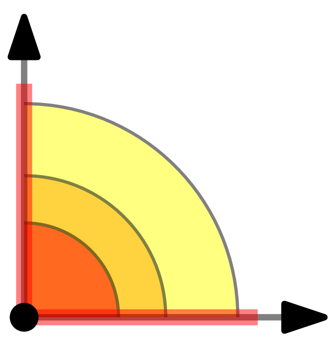
Why?
Index of dependence between **Completely Random Measures,** used as prior in Bayesian NonParametrics



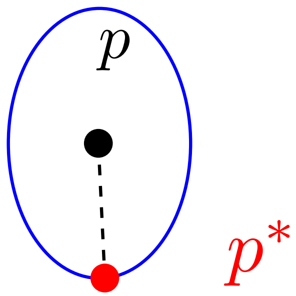
1 - Measuring dependence with Wasserstein distance

2 - Why look at Lévy intensities: link with completely random measures



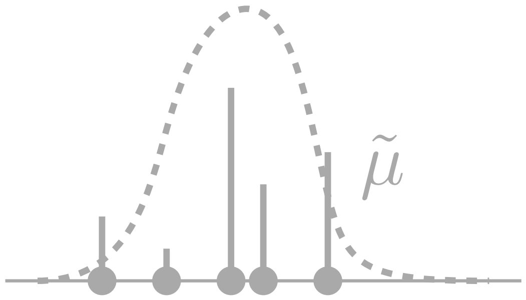


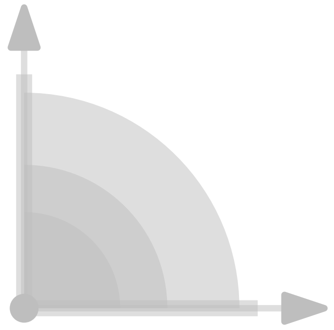
3 - Extended Wasserstein distance and index of dependence



1 - Measuring dependence with Wasserstein distance

2 - Why look at Lévy intensities: link with completely random measures



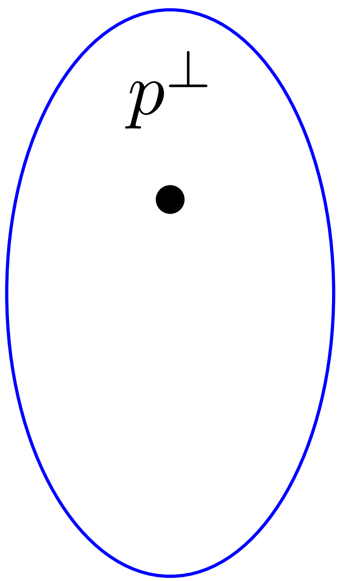


3 - Extended Wasserstein distance and index of dependence

Measuring dependence with a distance

Dependence structure Π

"most independent"



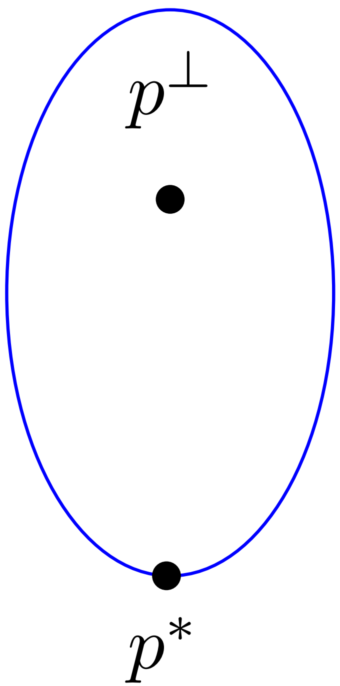
If $\mathcal{D}$ measure of discrepancy on Π then $\mathcal{D}(p, p^{\perp})$ measure of dependence of p.

Móri, and Székely (2020). The Earth Mover's correlation.
Nies, Staudt, and Munk (2021). Transport dependency: Optimal transport based dependency measures.
Mordant and Segers (2022). Measuring dependence between random vectors via optimal transport.
Wiesel (2022). Measuring association with Wasserstein distances.

Measuring dependence with a distance

Dependence structure Π

"most independent"



"most dependent"

If $\mathcal{D}$ measure of discrepancy on Π then $\mathcal{D}(p, p^{\perp})$ measure of dependence of p.

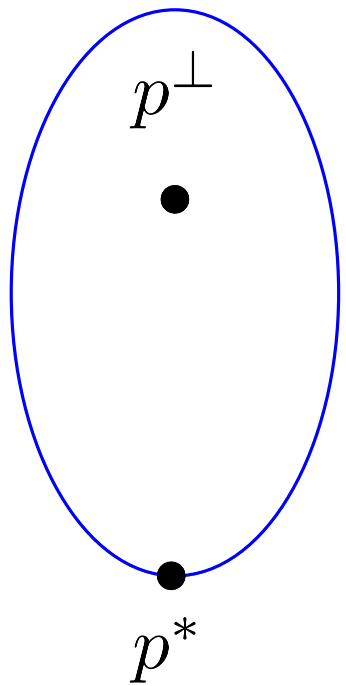
To solve: when is $\mathcal{D}(p, p^{\perp})$ maximized? say p^*

Móri, and Székely (2020). The Earth Mover's correlation.
Nies, Staudt, and Munk (2021). Transport dependency: Optimal transport based dependency measures.
Mordant and Segers (2022). Measuring dependence between random vectors via optimal transport.
Wiesel (2022). Measuring association with Wasserstein distances.

Measuring dependence with a distance

Dependence structure Π

"most independent"



"most dependent"

If $\mathcal{D}$ measure of discrepancy on Π then $\mathcal{D}(p, p^\perp)$ measure of dependence of p.

To solve: when is $\mathcal{D}(p, p^\perp)$ maximized? say p^*

Index of dependence

$$\mathcal{I}(p) = \frac{\mathcal{D}(p, p^\perp)}{\mathcal{D}(p^*, p^\perp)}$$

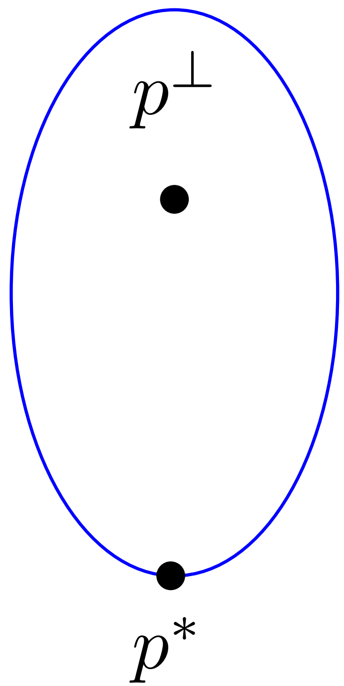
$\mathcal{I}(p) \in [0, 1]$ and equal to 0 (resp. 1) for $p^\perp$ (resp. p^*)

Móri, and Székely (2020). The Earth Mover's correlation.
Nies, Staudt, and Munk (2021). Transport dependency: Optimal transport based dependency measures.
Mordant and Segers (2022). Measuring dependence between random vectors via optimal transport.
Wiesel (2022). Measuring association with Wasserstein distances.

Measuring dependence with a distance

Dependence structure Π

"most independent"



"most dependent"

If $\mathcal{D}$ measure of discrepancy on Π then $\mathcal{D}(p, p^\perp)$ measure of dependence of p.

To solve: when is $\mathcal{D}(p, p^\perp)$ maximized? say p^*

Index of dependence

$$\mathcal{I}(p) = \frac{\mathcal{D}(p, p^\perp)}{\mathcal{D}(p^*, p^\perp)}$$

$\mathcal{I}(p) \in [0, 1]$ and equal to 0 (resp. 1) for $p^\perp$ (resp. p^*)

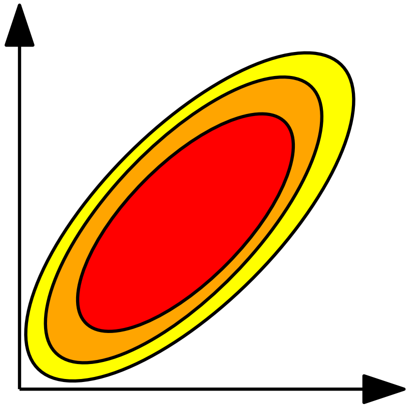
Alternative: if $\mathcal{D}(p^*, p)$ maximized at $p^\perp$ then

$$\mathcal{I}(p) = 1 - \frac{\mathcal{D}(p^*, p)}{\mathcal{D}(p^*, p^\perp)}.$$

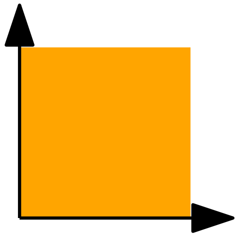
Móri, and Székely (2020). The Earth Mover's correlation.
Nies, Staudt, and Munk (2021). Transport dependency: Optimal transport based dependency measures.
Mordant and Segers (2022). Measuring dependence between random vectors via optimal transport.
Wiesel (2022). Measuring association with Wasserstein distances.

$\mathcal{D}$ is Wasserstein over the Manhattan distance

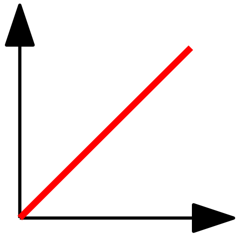
Over $[0,1]^2$



$p \in \Pi(\mathrm{d}x, \mathrm{d}x)$



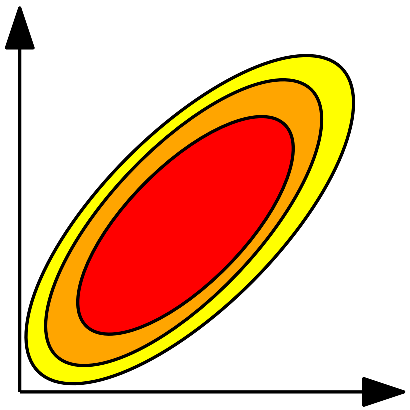
$p^{\perp}$



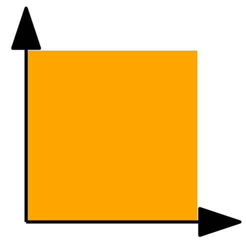
p^{*}

$\mathcal{D}$ is Wasserstein over the Manhattan distance

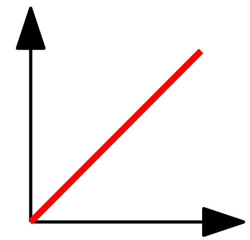
Over $[0,1]^2$



$p \in \Pi(\mathrm{d}x, \mathrm{d}x)$



$p^\perp$



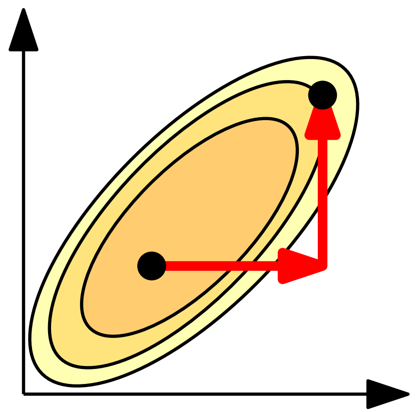
p^*

Theorem.

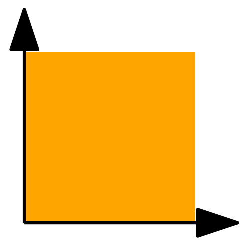
$W_2(p, p^\perp) \leq W_2(p^*, p^\perp)$ for any $p \in \Pi(\mathrm{d}x, \mathrm{d}x)$.

$\mathcal{D}$ is Wasserstein over the Manhattan distance

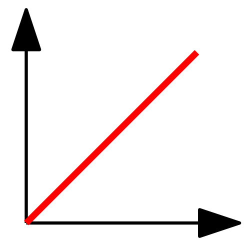
Over $[0,1]^2$



$p \in \Pi(\mathrm{d}x, \mathrm{d}x)$



$p^\perp$



p^*

Theorem. Choose

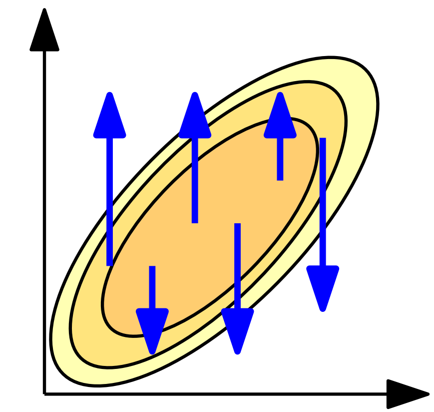
$$d((x_1, y_1), (x_2, y_2)) = |x_1 - x_2| + |y_1 - y_2|.$$

Then

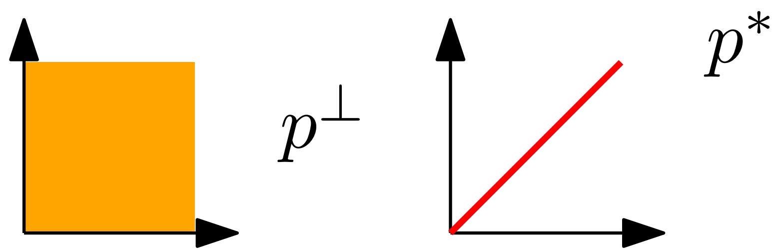
$W_2(p, p^\perp) \leq W_2(p^*, p^\perp)$ for any $p \in \Pi(\mathrm{d}x, \mathrm{d}x)$.

$\mathcal{D}$ is Wasserstein over the Manhattan distance

Over $[0,1]^2$



$p \in \Pi(\mathrm{d}x, \mathrm{d}x)$



Theorem. Choose

$$d((x_1, y_1), (x_2, y_2)) = |x_1 - x_2| + |y_1 - y_2|.$$

Then

$W_2(p, p^\perp) \leq W_2(p^*, p^\perp)$ for any $p \in \Pi(\mathrm{d}x, \mathrm{d}x)$.

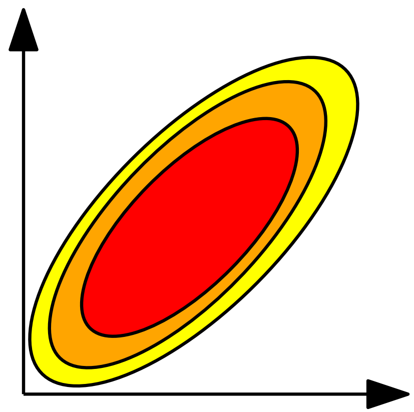
Proof. Send (x, y) onto (x, y') with y' independent of (x, y).

$$W_2(p, p^\perp)^2 \leq \iint (y - y')^2 \, \mathrm{d}y \mathrm{d}y' = W_2(p^*, p^\perp)^2.$$

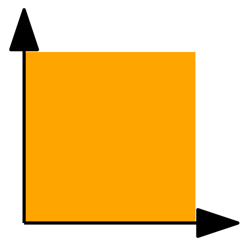
Explicit computation

$\mathcal{D}$ is Wasserstein over the Manhattan distance

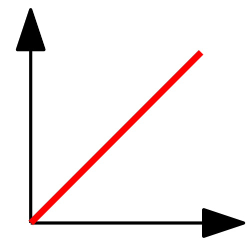
Over $[0,1]^2$



$p \in \Pi(\mathrm{d}x, \mathrm{d}x)$



$p^\perp$



p^* supported on graph 1-Lipschitz function

Theorem. Choose

$$d((x_1, y_1), (x_2, y_2)) = |x_1 - x_2| + |y_1 - y_2|.$$

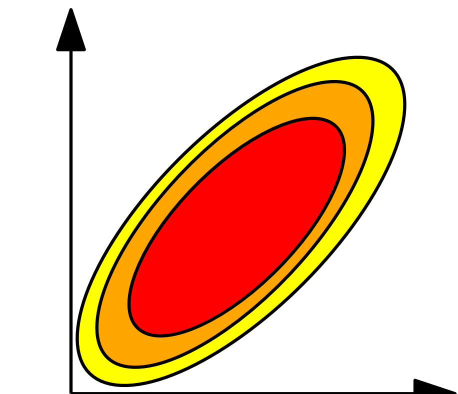
Then

$W_2(p, p^\perp) \leq W_2(p^*, p^\perp)$ for any $p \in \Pi(\mathrm{d}x, \mathrm{d}x)$.

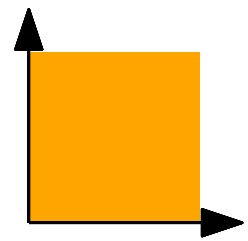
- Extension to metric spaces.
- Inherits properties of optimal transport.

$\mathcal{D}$ is Wasserstein over the Manhattan distance

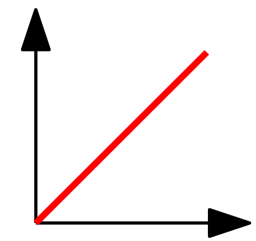
Over $[0,1]^2$



$p \in \Pi(\mathrm{d}x, \mathrm{d}x)$



$p^{\perp}$



p^* supported on graph 1-Lipschitz function

Theorem. Choose

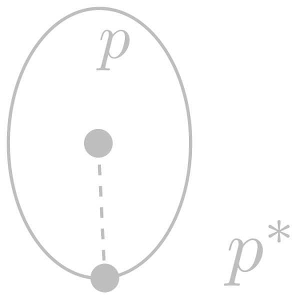
$$d((x_1, y_1), (x_2, y_2)) = |x_1 - x_2| + |y_1 - y_2|.$$

Then

$$W_2(p, p^{\perp}) \leq W_2(p^*, p^{\perp}) \text{ for any } p \in \Pi(\mathrm{d}x, \mathrm{d}x).$$

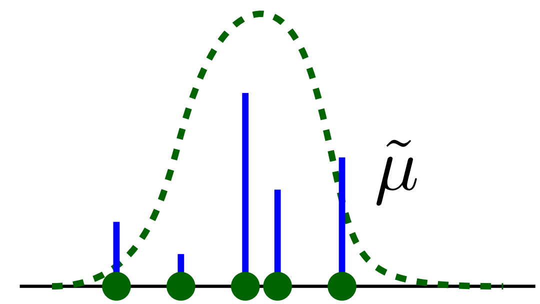
- Extension to metric spaces.
- Inherits properties of optimal transport.

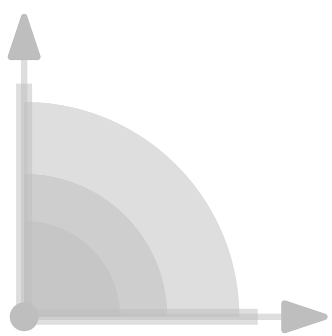
- Not tractable analytically.
- Sampling: curse of dimensionality.



1 - Measuring dependence with Wasserstein distance

2 - Why look at Lévy intensities: link with completely random measures



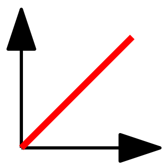
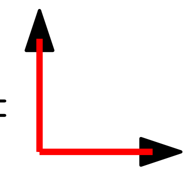


3 - Extended Wasserstein distance and index of dependence

Lévy intensities and Completely random measures

$\bar{\nu}$ measure on $(0, +\infty)$ with $\int s^2 \, \mathrm{d}\bar{\nu}(s) < +\infty$

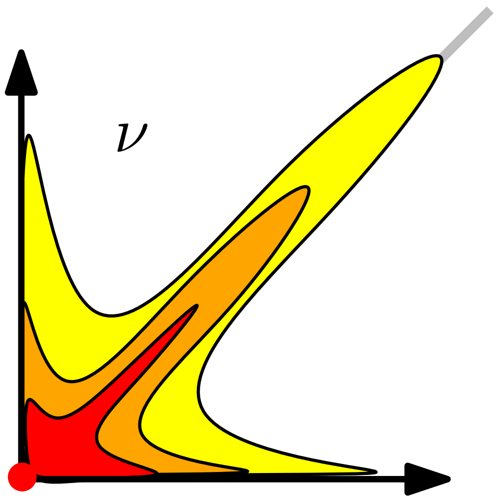
$\Pi(\bar{\nu}, \cdots, \bar{\nu}) = \{$Measure on $[0, +\infty)^d \setminus \{0\}$ with marginals $\bar{\nu}\}$

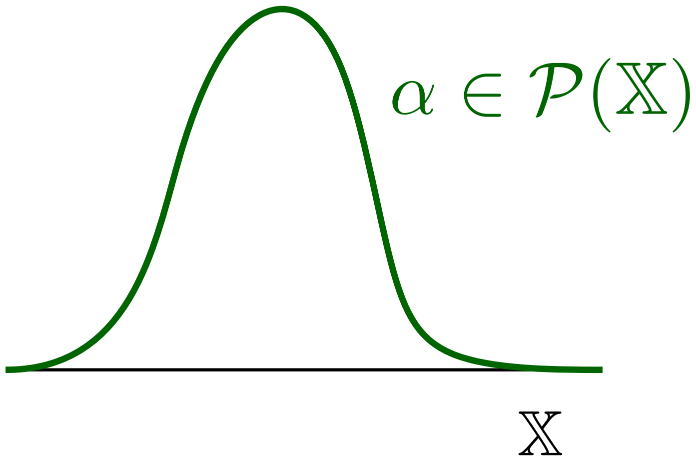
$$\operatorname*{argmax}_{\nu \in \Pi(\bar{\nu}, \dots, \bar{\nu})} W_* \Big(\nu,$$  $$\Big) =$$ 

Lévy intensities and Completely random measures

$\bar{\nu}$ measure on $(0, +\infty)$ with $\int s^2 \, \mathrm{d}\bar{\nu}(s) < +\infty$

$\Pi(\bar{\nu}, \cdots, \bar{\nu}) = \{$Measure on $[0, +\infty)^d \setminus \{0\}$ with marginals $\bar{\nu}\}$

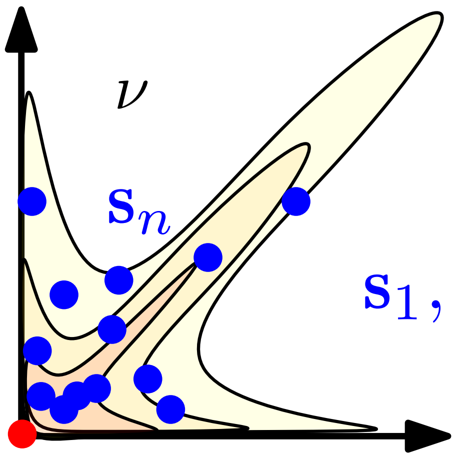




Lévy intensities and Completely random measures

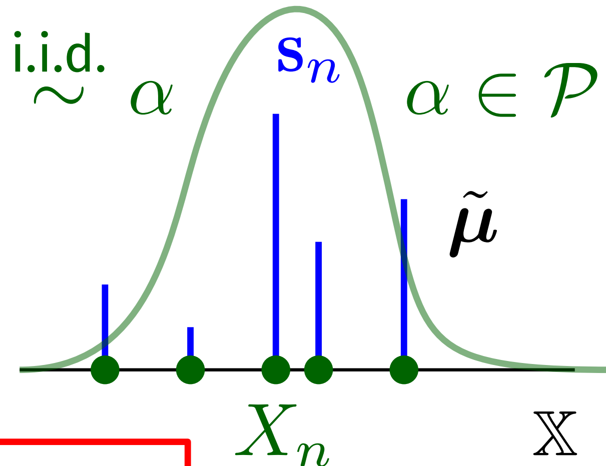
$\bar{\nu}$ measure on $(0,+\infty)$ with $\int s^2 \,\mathrm{d}\bar{\nu}(s) < +\infty$

$\Pi(\bar{\nu},\cdots,\bar{\nu}) = \{$Measure on $[0,+\infty)^d \setminus \{0\}$ with marginals $\bar{\nu}\}$



$X_1,\ldots,X_n,\ldots \overset{\text{i.i.d.}}{\sim} \alpha \qquad \alpha \in \mathcal{P}(\mathbb{X})$

$\mathbf{s}_1,\ldots,\mathbf{s}_n,\ldots \sim \mathrm{Poisson}(\nu)$



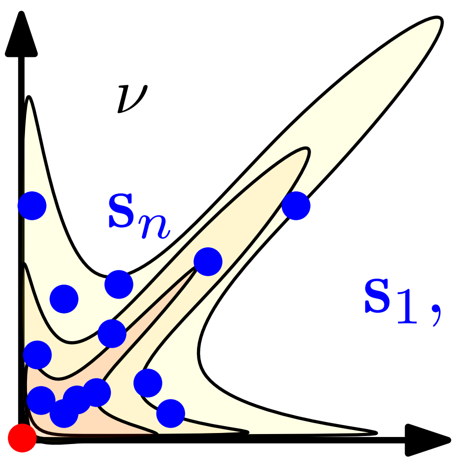
$$\tilde{\boldsymbol{\mu}} = \sum_{n\geq 1} \mathbf{s}_n \delta_{X_n} = (\tilde{\mu}_1,\ldots,\tilde{\mu}_d)$$

Collection of d random measures on $\mathcal{X}$

Lévy intensities and Completely random measures

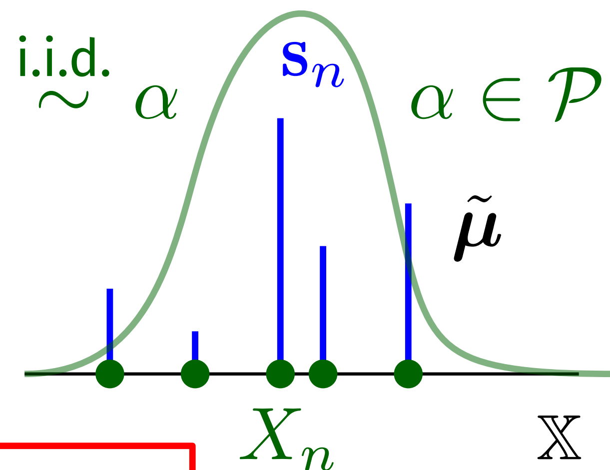
$\bar{\nu}$ measure on $(0,+\infty)$ with $\int s^2 \,\mathrm{d}\bar{\nu}(s) < +\infty$

$\Pi(\bar{\nu},\cdots,\bar{\nu}) = \{$Measure on $[0,+\infty)^d \setminus \{0\}$ with marginals $\bar{\nu}\}$



$X_1,\ldots,X_n,\ldots \overset{\text{i.i.d.}}{\sim} \alpha$ $\quad \alpha \in \mathcal{P}(\mathbb{X})$

$\mathbf{s}_1,\ldots,\mathbf{s}_n,\ldots \sim \mathrm{Poisson}(\nu)$



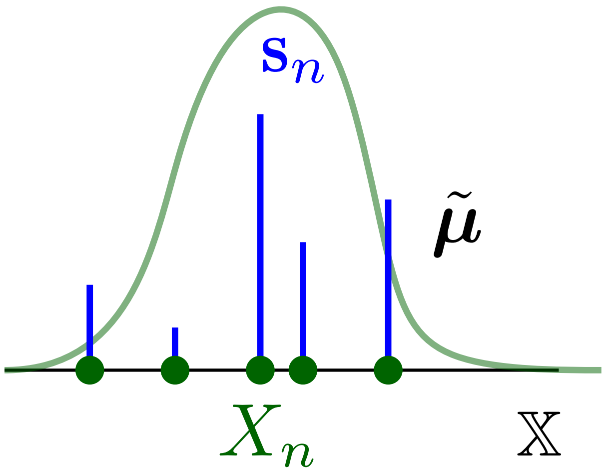
$$\tilde{\boldsymbol{\mu}} = \sum_{n\geq 1} \mathbf{s}_n \delta_{X_n} = (\tilde{\mu}_1,\ldots,\tilde{\mu}_d)$$

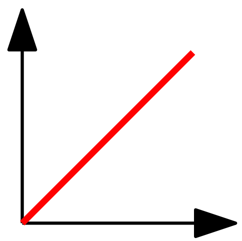
Completely Random Vector. For all $A_1,\ldots,A_n \subseteq \mathbb{X}$ disjoints, the vectors $\tilde{\boldsymbol{\mu}}(A_1),\ldots,\tilde{\boldsymbol{\mu}}(A_n)$ are independent random vectors in $\mathbb{R}_+^d$.

For $A \subseteq \mathbb{X}$, the random variables $\tilde{\mu}_1(A),\ldots,\tilde{\mu}_d(A)$ may be dependent.

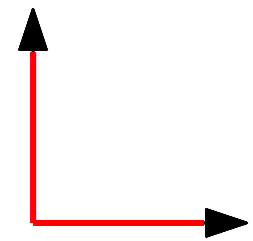
And then normalization

$\nu \in \Pi(\bar{\nu}, \dots, \bar{\nu})$ and $\alpha \in \mathcal{P}(\mathbb{X})$ gives law of $\tilde{\boldsymbol{\mu}}$ in $\mathcal{P}(\mathcal{M}_+(\mathbb{X})^d)$.





$\tilde{\mu}_1 = \dots = \tilde{\mu}_d$ a.s.

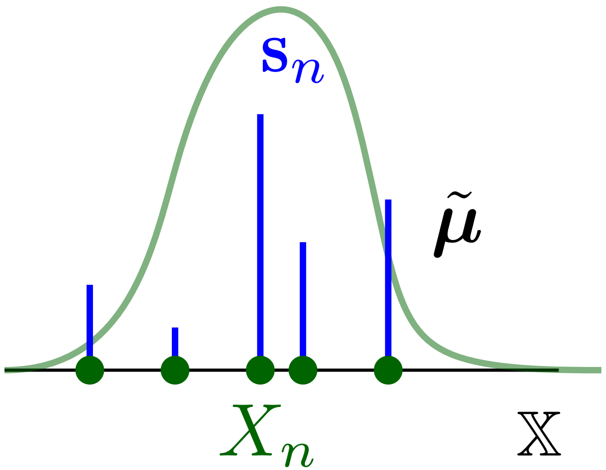


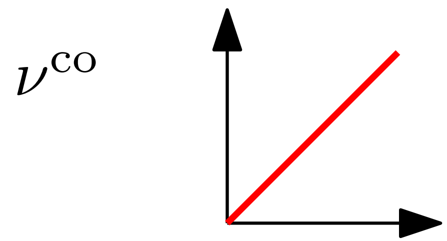
$\tilde{\mu}_1, \dots, \tilde{\mu}_d$ independent.

Lijoi, Nipoti and Prünster (2014). Bayesian inference with dependent normalized completely random measures.

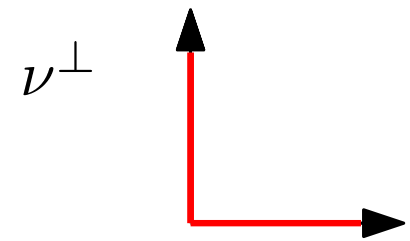
And then normalization

$\nu \in \Pi(\bar{\nu}, \ldots, \bar{\nu})$ and $\alpha \in \mathcal{P}(\mathbb{X})$ gives law of $\tilde{\boldsymbol{\mu}}$ in $\mathcal{P}(\mathcal{M}_+(\mathbb{X})^d)$.





$\tilde{\mu}_1 = \ldots = \tilde{\mu}_d$ a.s.



$\tilde{\mu}_1, \ldots, \tilde{\mu}_d$ independent.

Normalized version: $$\left(\frac{\tilde{\mu}_1}{\tilde{\mu}_1(\mathbb{X})}, \ldots, \frac{\tilde{\mu}_d}{\tilde{\mu}_d(\mathbb{X})} \right)$$

gives d random (dependent) probabilities, law in $\mathcal{P}(\mathcal{P}(\mathbb{X})^d)$.

Lijoi, Nipoti and Prünster (2014). Bayesian inference with dependent normalized completely random measures.

Why random probabilities? Prior in Bayesian Nonparametrics

Bayesian statistics

prior

$$\theta \sim \pi$$

data

$$X_1, \dots, X_n | \theta \overset{\text{i.i.d.}}{\sim} p_\theta$$

Lijoi and Prünster (2010). Models beyond the Dirichlet process.

Why random probabilities? Prior in Bayesian Nonparametrics

Bayesian statistics

prior

$$\theta \sim \pi$$

data

$$X_1, \ldots, X_n | \theta \overset{\text{i.i.d.}}{\sim} p_\theta$$

Remark: p_θ with $\theta \sim \pi$ is a random probability.

Bayesian NonParametrics: define directly $\tilde{p}$ a random probability instead of p_θ and π.

Lijoi and Prünster (2010). Models beyond the Dirichlet process.

Why random probabilities? Prior in Bayesian Nonparametrics

Bayesian statistics

prior

$$\theta \sim \pi$$

data

$$X_1, \ldots, X_n | \theta \overset{\text{i.i.d.}}{\sim} p_\theta$$

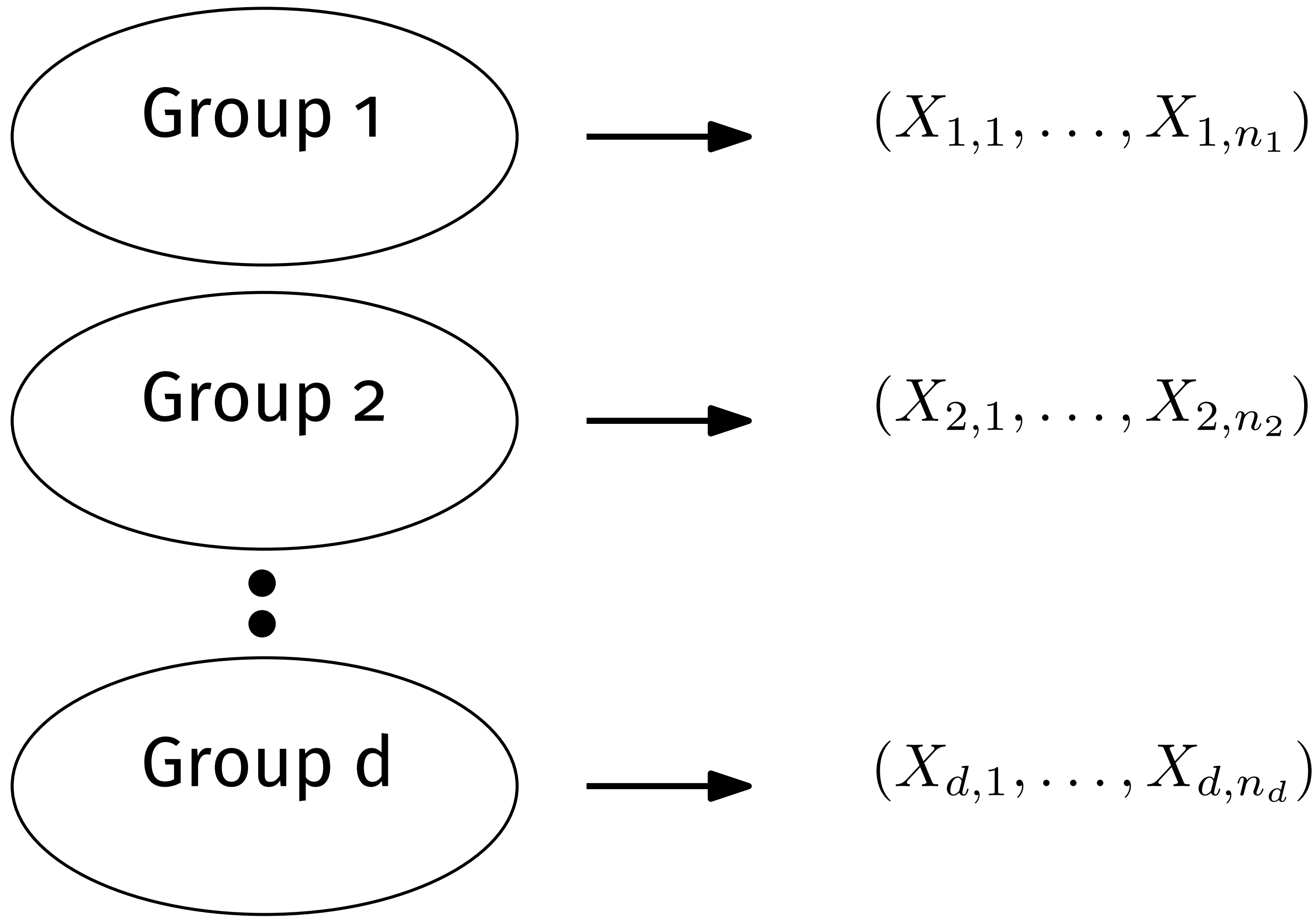
Remark: p_θ with $\theta \sim \pi$ is a random probability.

Bayesian NonParametrics: define directly $\tilde{p}$ a random probability instead of p_θ and π.

(Normalized) completely random measures: analytical tractability of the posterior distribution.

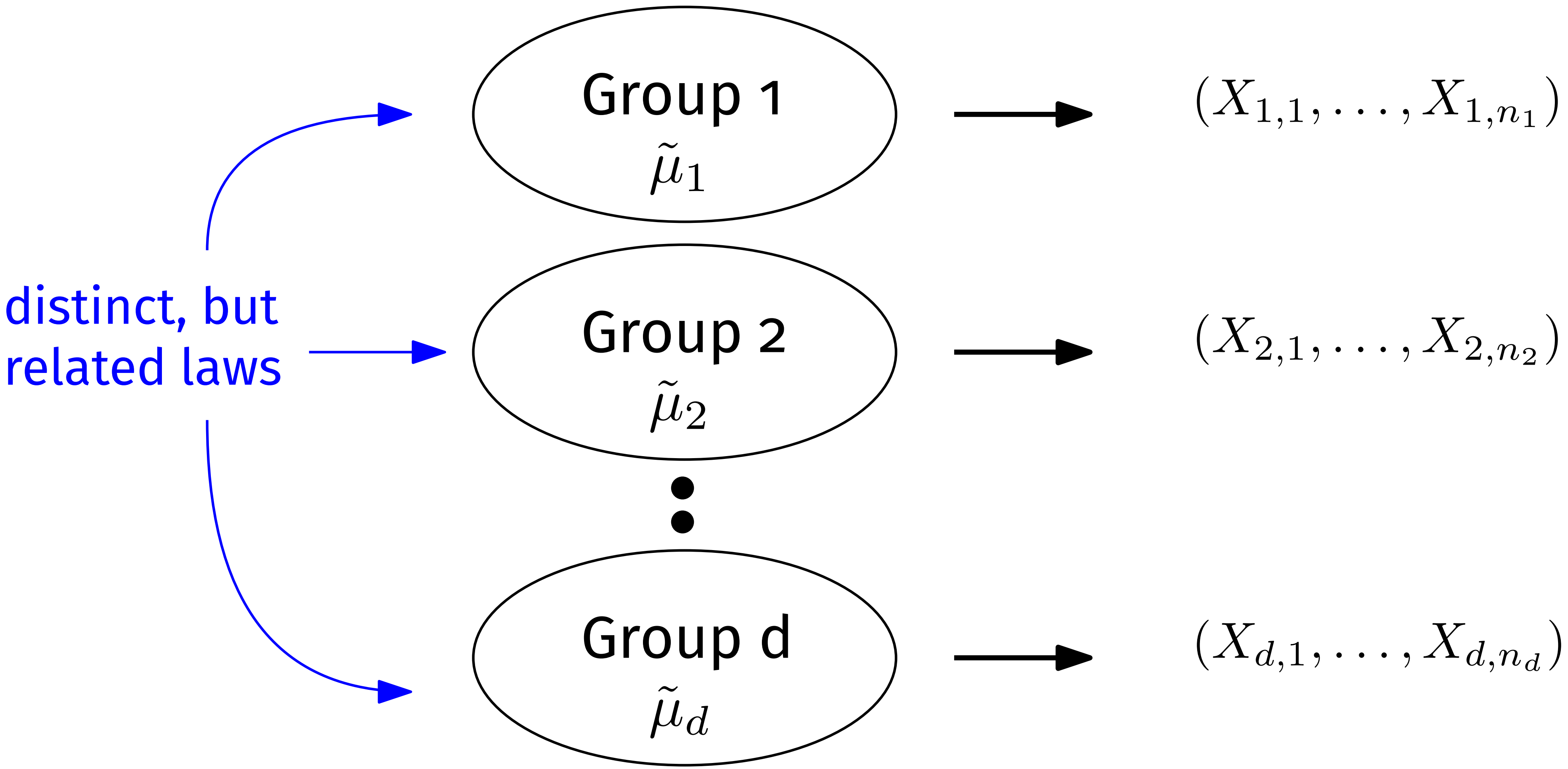
Lijoi and Prünster (2010). Models beyond the Dirichlet process.

Why quantifying dependence?



Catalano, Lijoi and Prünster (2021). Measuring dependence in the Wasserstein distance for Bayesian nonparametric models.

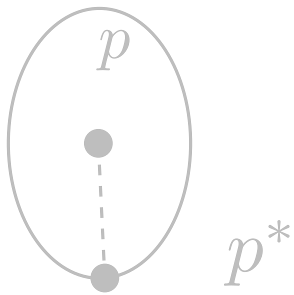
Why quantifying dependence?



Bayesian inference allows for borrowing of information

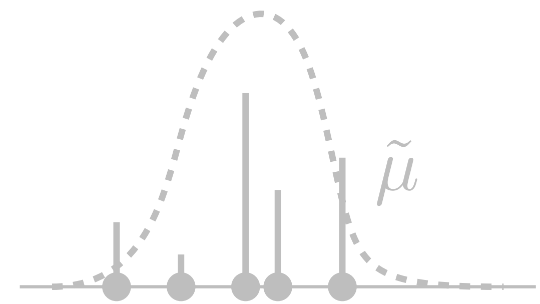
Goal: quantifying the amount of **dependence** between groups already present in the **prior**

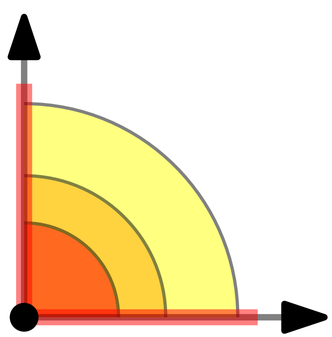
Catalano, Lijoi and Prünster (2021). Measuring dependence in the Wasserstein distance for Bayesian nonparametric models.



1 - Measuring dependence with Wasserstein distance

2 - Why look at Lévy intensities: link with completely random measures





3 - Extended Wasserstein distance and index of dependence

(Classical) optimal transport

Definition. If ν^1, ν^2 probability distributions, the Wasserstein distance is

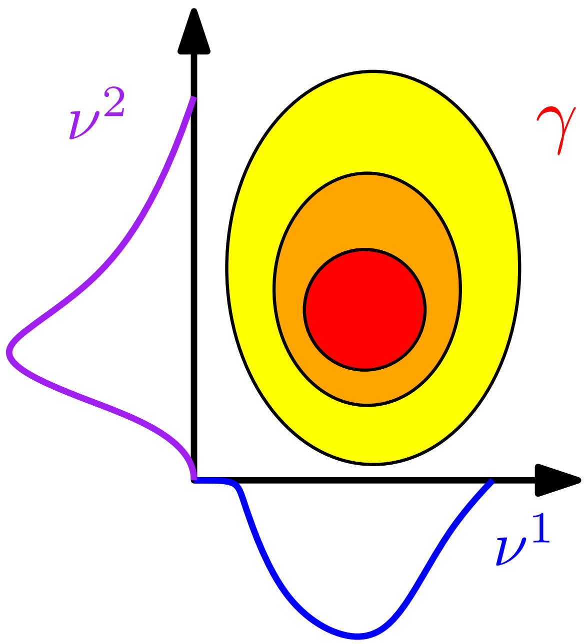
$$W_2(\nu^1, \nu^2)^2 = \min_{(X,Y)} \left\{ \mathbb{E}\left[\|X - Y\|^2 \right] \ : \ X \sim \nu^1 \text{ and } Y \sim \nu^2 \right\}$$

(Classical) optimal transport

Definition. If ν^1, ν^2 probability distributions, the Wasserstein distance is

$$W_2(\nu^1, \nu^2)^2 = \min_{(X,Y)} \left\{ \mathbb{E}\left[\|X - Y\|^2 \right] \ : \ X \sim \nu^1 \text{ and } Y \sim \nu^2 \right\}$$

$$= \min_{\gamma} \left\{ \iint \|x-y\|^2 \mathrm{d}\gamma(x,y) \ : \ \pi_1 \# \gamma = \nu^1 \text{ and } \pi_2 \# \gamma = \nu^2 \right\}$$



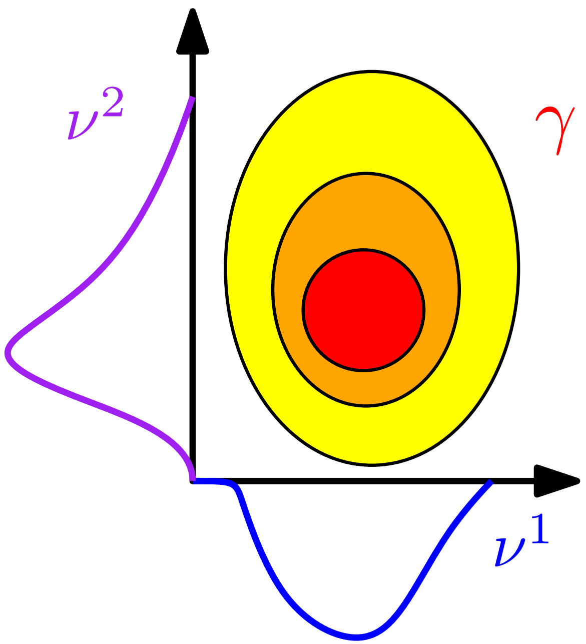
(Classical) optimal transport

Definition. If ν^1, ν^2 probability distributions, the Wasserstein distance is

$$W_2(\nu^1,\nu^2)^2 = \min_{(X,Y)} \left\{ \mathbb{E}\left[\|X-Y\|^2\right] \ : \ X \sim \nu^1 \text{ and } Y \sim \nu^2 \right\}$$

$$= \min_{\gamma} \left\{ \iint \|x-y\|^2 \mathrm{d}\gamma(x,y) \ : \ \pi_1 \# \gamma = \nu^1 \text{ and } \pi_2 \# \gamma = \nu^2 \right\}$$

$$\leq \int \|x\|^2 \mathrm{d}\nu^1(x) + \int \|y\|^2 \mathrm{d}\nu^2(y)$$



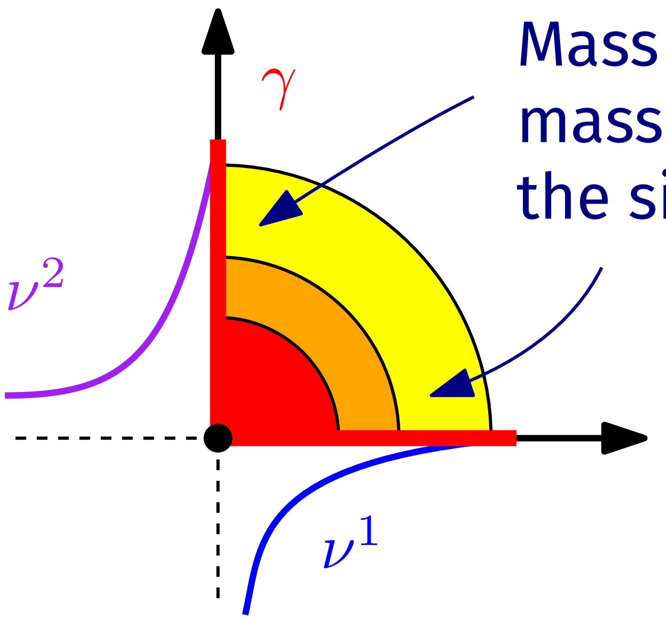
Observation. Naively, makes sense if ν^1, ν^2 have infinite mass but **finite** second moment.

Extended Wasserstein distance

Definition. If ν^1, ν^2 positive measures on $\mathbb{R}^d_+ \setminus \{0\}$ with **finite second moments**, the Wasserstein distance is

$$W_*(\nu^1,\nu^2)^2 = \min_\gamma \left\{ \iint \|x-y\|^2 \mathrm{d}\gamma(x,y) \;:\; \pi_1\#\gamma|_{\mathbb{R}^d_+\setminus\{0\}} = \nu^1 \text{ and } \pi_2\#\gamma|_{\mathbb{R}^d_+\setminus\{0\}} = \nu^2 \right\}$$

with γ measure on $\mathbb{R}^{2d}_+ \setminus \{(0,0)\}$.



Mass on $\mathbb{R}^d_{+,*} \times \{0\}$ and $\{0\} \times \mathbb{R}^d_{+,*}$: mass “destroyed” or “created” from the sink/reservoir $(0,0)$.

Figalli and Gigli (2010). A new transportation distance between non-negative measures, with applications to gradients flows with Dirichlet boundary conditions.

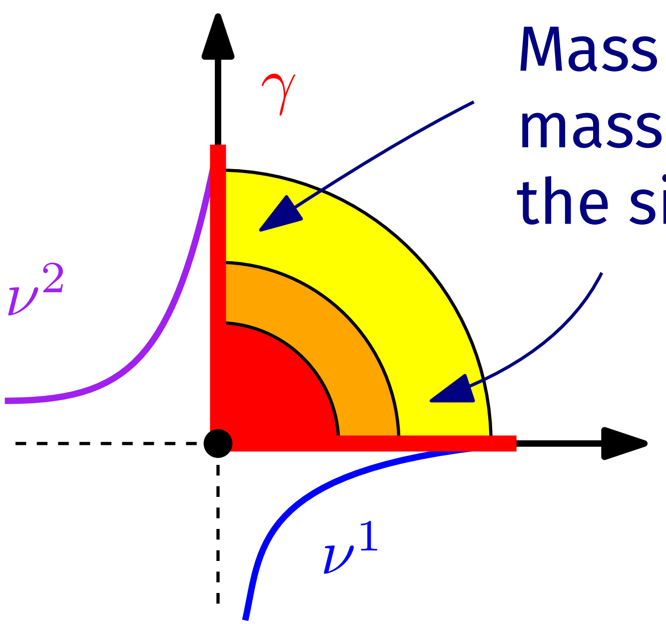
Guillen, Mou, Święch (2019). Coupling Lévy measures and comparison principles for viscosity solutions.

Extended Wasserstein distance

Definition. If ν^1, ν^2 positive measures on $\mathbb{R}^d_+ \setminus \{0\}$ with **finite second moments**, the Wasserstein distance is

$$W_*(\nu^1,\nu^2)^2 = \min_\gamma \left\{ \iint \|x-y\|^2 \mathrm{d}\gamma(x,y) \ : \ \pi_1\#\gamma|_{\mathbb{R}^d_+\setminus\{0\}} = \nu^1 \text{ and } \pi_2\#\gamma|_{\mathbb{R}^d_+\setminus\{0\}} = \nu^2 \right\}$$

with γ measure on $\mathbb{R}^{2d}_+ \setminus \{(0,0)\}$.



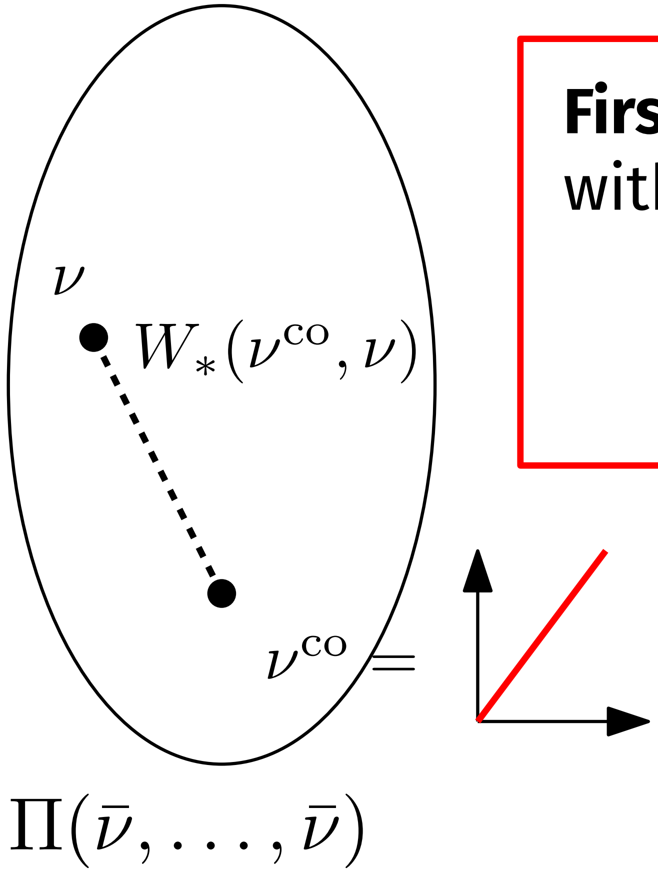
Mass on $\mathbb{R}^d_{+,*} \times \{0\}$ and $\{0\} \times \mathbb{R}^d_{+,*}$: mass "destroyed" or "created" from the sink/reservoir $(0,0)$.

Metrizes weak convergence and convergence of second moment with respect to 0 (work with I. Pinheiro).

Figalli and Gigli (2010). A new transportation distance between non-negative measures, with applications to gradients flows with Dirichlet boundary conditions.

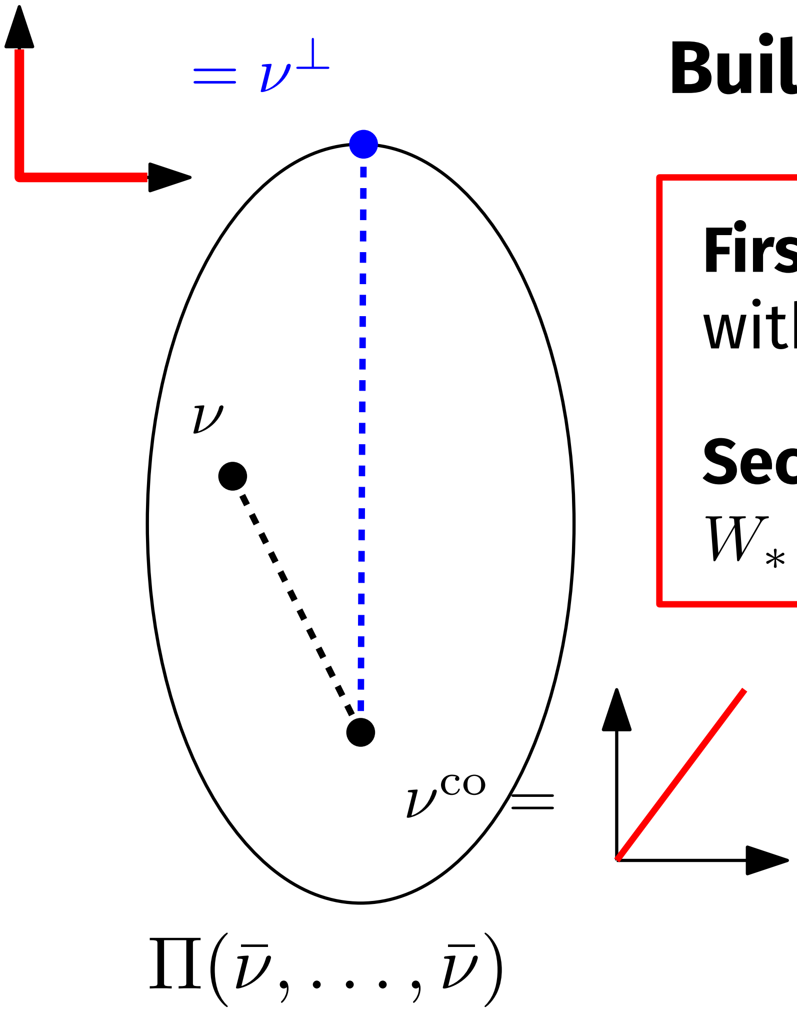
Guillen, Mou, Święch (2019). Coupling Lévy measures and comparison principles for viscosity solutions.

Building the index



First result. $W_*(\nu^{\text{co}}, \nu)$ can be computed with 1d integrals of tail functions.

Catalano, Lavenant, Lijoi, Prünster (2023+). A Wasserstein index of dependence for random measures.

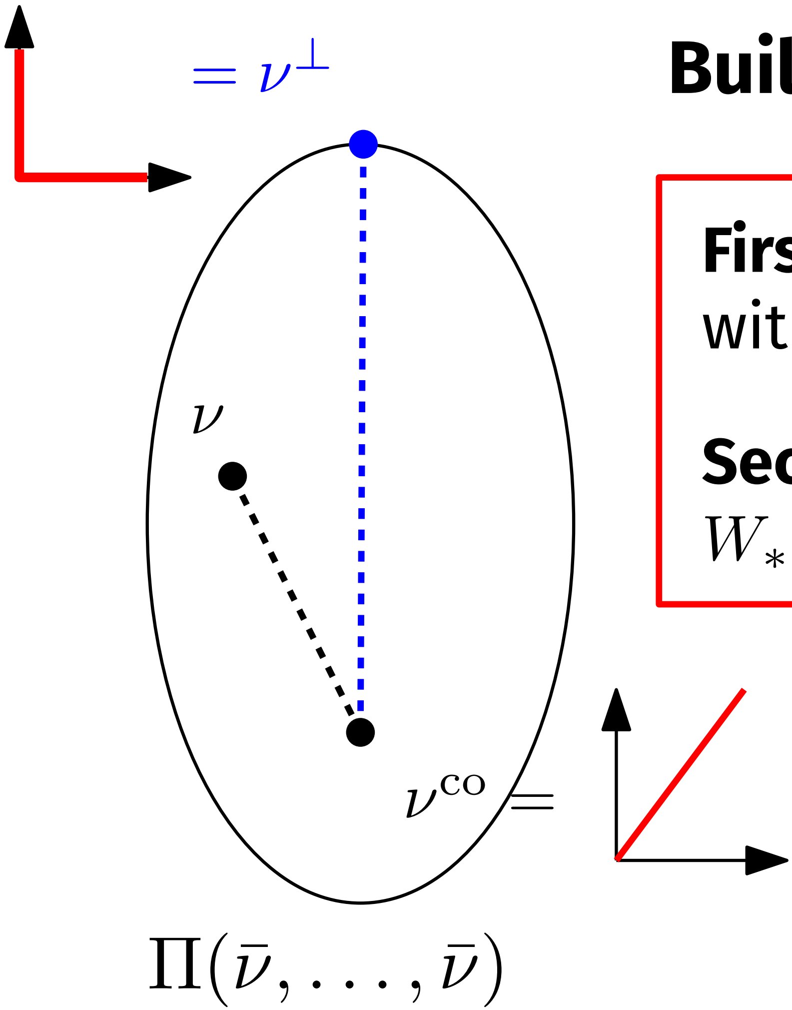


Building the index

First result. $W_*(\nu^{\mathrm{co}}, \nu)$ can be computed with 1d integrals of tail functions.

Second result. If ν^{co} has infinite mass, $W_*(\nu^{\mathrm{co}}, \nu)$ is maximized exactly for $\nu = \nu^{\perp}$.

Catalano, Lavenant, Lijoi, Prünster (2023+). A Wasserstein index of dependence for random measures.



Building the index

First result. $W_*(\nu^{\mathrm{co}}, \nu)$ can be computed with 1d integrals of tail functions.

Second result. If ν^{co} has infinite mass, $W_*(\nu^{\mathrm{co}}, \nu)$ is maximized exactly for $\nu = \nu^{\perp}$.

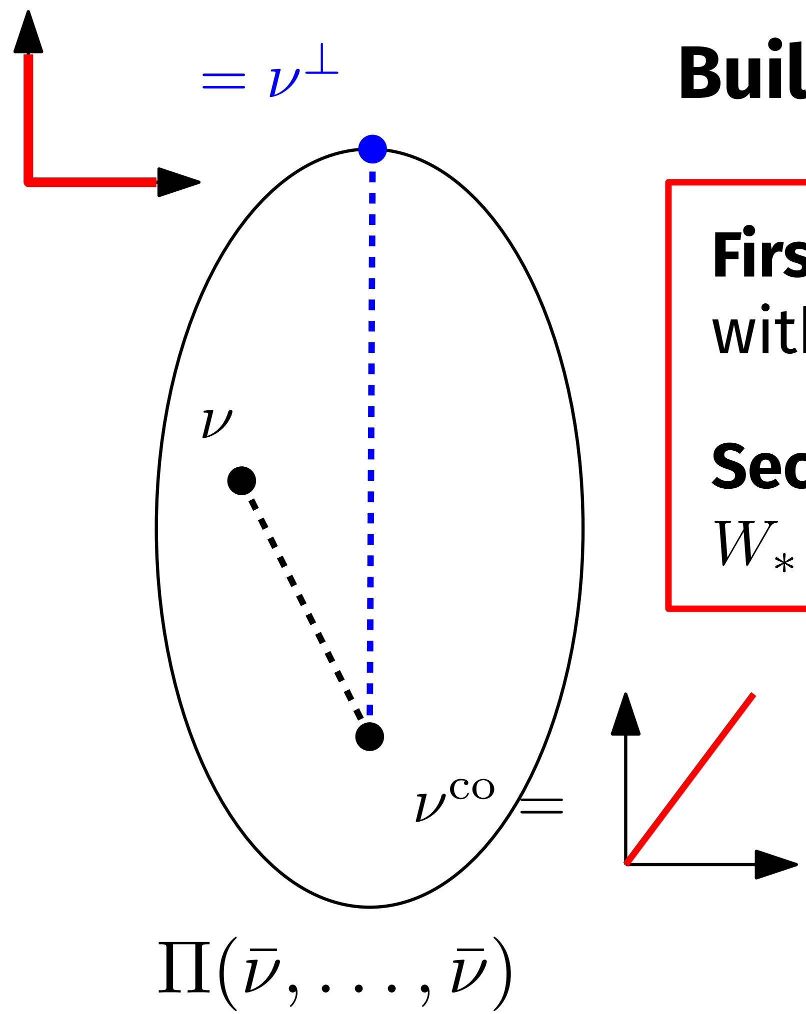
Define:

$$\mathcal{I}(\nu) = 1 - \frac{W_*(\nu^{\mathrm{co}}, \nu)^2}{W_*(\nu^{\mathrm{co}}, \nu^{\perp})^2}.$$

It belongs to $[0, 1]$ and satisfies:

$$\mathcal{I}(\nu^{\perp}) = 0, \qquad \mathcal{I}(\nu^{\mathrm{co}}) = 1.$$

Building the index



First result. $W_*(\nu^{\mathrm{co}}, \nu)$ can be computed with 1d integrals of tail functions.

Second result. If ν^{co} has infinite mass, $W_*(\nu^{\mathrm{co}}, \nu)$ is maximized exactly for $\nu = \nu^{\perp}$.

Define:

$$\mathcal{I}(\nu) = 1 - \frac{W_*(\nu^{\mathrm{co}}, \nu)^2}{W_*(\nu^{\mathrm{co}}, \nu^{\perp})^2}.$$

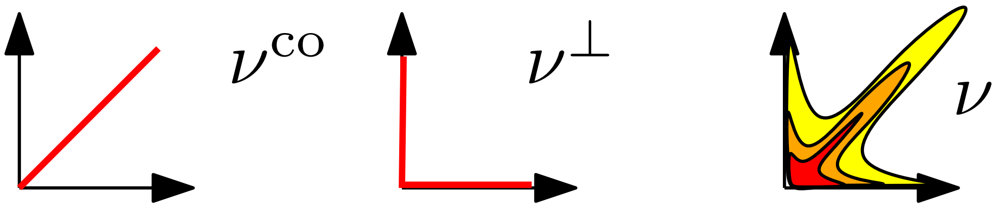
It belongs to $[0,1]$ and satisfies:

$$\mathcal{I}(\nu^{\perp}) = 0, \qquad \mathcal{I}(\nu^{\mathrm{co}}) = 1.$$

Consequence. We have an index of dependence for homogeneous infinitely active completely random vectors without fixed atoms, with equal marginals and finite second moments.

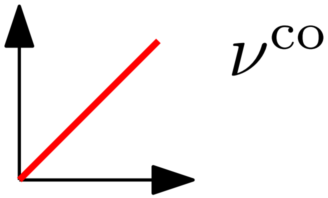
Catalano, Lavenant, Lijoi, Prünster (2023+). A Wasserstein index of dependence for random measures.

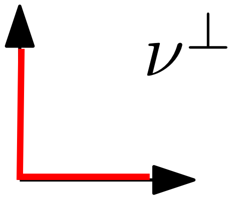
The proof of the main result (d=2)

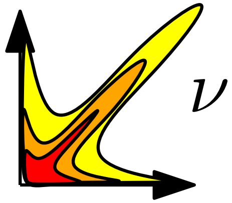


$$W_*^2(\nu, \nu^{\text{co}}) \leq W_*^2(\nu^{\perp}, \nu^{\text{co}})$$

The proof of the main result (d=2)



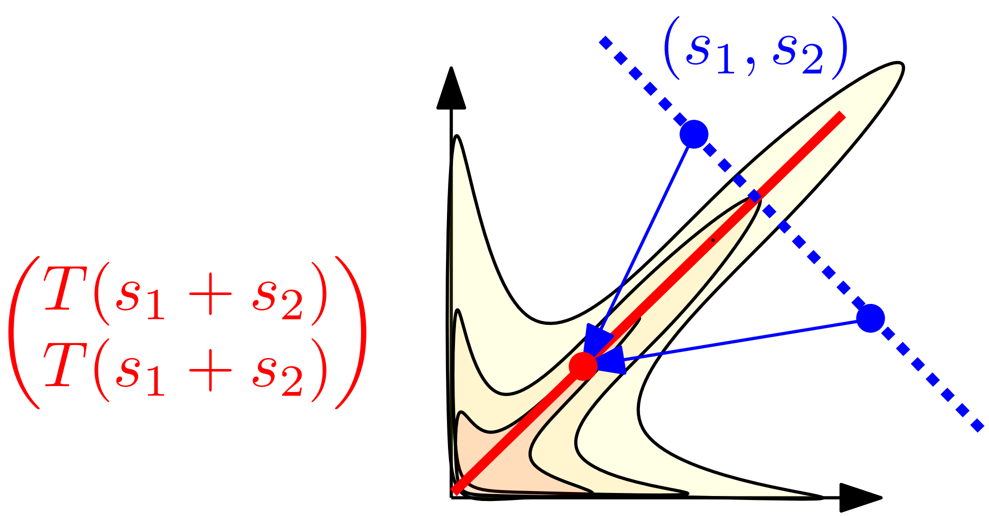




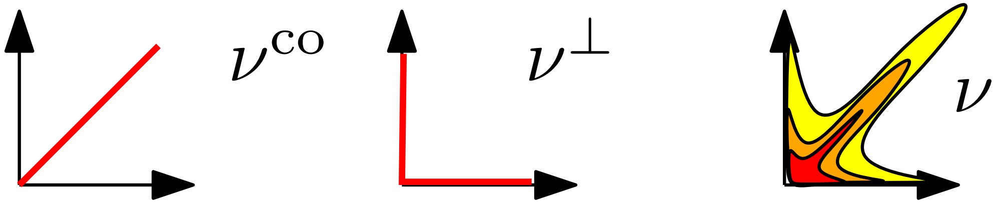
$$W_*^2(\nu, \nu^{\mathrm{co}}) \leq W_*^2(\nu^{\perp}, \nu^{\mathrm{co}})$$

Lemma.
The transport map from ν to ν^{co} is $(s_1, s_2) \mapsto \begin{pmatrix} T(s_1+s_2) \\ T(s_1+s_2) \end{pmatrix}$ with $T = u'$, for $u : \mathbb{R} \to \mathbb{R}$ convex and $u(0) = 0$. An optimal Kantorovich potential is

$$\varphi(s_1, s_2) = \frac{s_1^2 + s_2^2}{2} - u(s_1 + s_2).$$



The proof of the main result (d=2)



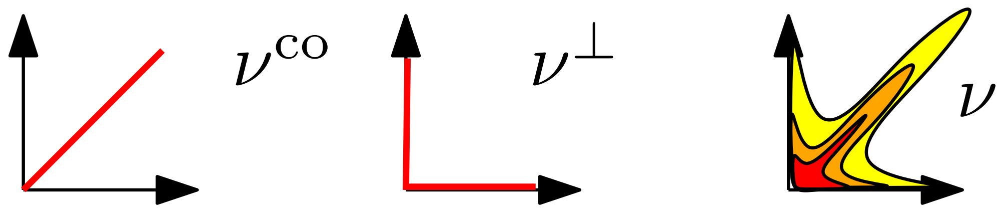
$$W_*^2(\nu, \nu^{\mathrm{co}}) \leq W_*^2(\nu^{\perp}, \nu^{\mathrm{co}})$$

Lemma.
The transport map from ν to ν^{co} is $(s_1, s_2) \mapsto \begin{pmatrix} T(s_1+s_2) \\ T(s_1+s_2) \end{pmatrix}$ with $T = u'$, for $u : \mathbb{R} \to \mathbb{R}$ convex and $u(0) = 0$. An optimal Kantorovich potential is

$$\varphi(s_1, s_2) = \frac{s_1^2 + s_2^2}{2} - u(s_1 + s_2).$$

$$\frac{W_*^2(\nu^{\perp}, \nu^{\mathrm{co}})}{2} \geq \int \varphi \, \mathrm{d}\nu^{\perp} + \int \psi \, \mathrm{d}\nu^{\mathrm{co}}$$
$$= \frac{W_*^2(\nu, \nu^{\mathrm{co}})}{2} + \int \varphi \, \mathrm{d}(\nu^{\perp} - \nu)$$

The proof of the main result (d=2)



$$\boxed{W_*^2(\nu, \nu^{\mathrm{co}}) \leq W_*^2(\nu^{\perp}, \nu^{\mathrm{co}})}$$

Lemma.
The transport map from ν to ν^{co} is $(s_1, s_2) \mapsto \begin{pmatrix} T(s_1+s_2) \\ T(s_1+s_2) \end{pmatrix}$ with $T = u'$, for $u : \mathbb{R} \to \mathbb{R}$ convex and $u(0) = 0$. An optimal Kantorovich potential is

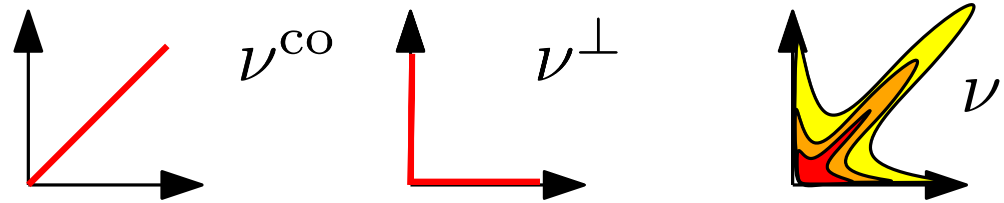
$$\varphi(s_1, s_2) = \frac{s_1^2 + s_2^2}{2} - u(s_1 + s_2).$$

$$\frac{W_*^2(\nu^{\perp}, \nu^{\mathrm{co}})}{2} \geq \int \varphi \,\mathrm{d}\nu^{\perp} + \int \psi \,\mathrm{d}\nu^{\mathrm{co}}$$

$$= \frac{W_*^2(\nu, \nu^{\mathrm{co}})}{2} + \int \varphi \,\mathrm{d}(\nu^{\perp} - \nu)$$

$$\int \varphi \,\mathrm{d}(\nu^{\perp} - \nu) = \int u(s_1+s_2)\,\mathrm{d}\nu(s_1, s_2) - \left(\int u(s_1)\mathrm{d}\bar{\nu}(s_1) + \int u(s_2)\mathrm{d}\bar{\nu}(s_2) \right)$$

The proof of the main result (d=2)



$$W_*^2(\nu,\nu^{\mathrm{co}}) \le W_*^2(\nu^{\perp},\nu^{\mathrm{co}})$$

Lemma.
The transport map from ν to ν^{co} is $(s_1,s_2) \mapsto \begin{pmatrix} T(s_1+s_2) \\ T(s_1+s_2) \end{pmatrix}$ with $T = u'$, for $u : \mathbb{R} \to \mathbb{R}$ convex and $u(0) = 0$. An optimal Kantorovich potential is

$$\varphi(s_1,s_2) = \frac{s_1^2+s_2^2}{2} - u(s_1+s_2).$$

$$\frac{W_*^2(\nu^{\perp},\nu^{\mathrm{co}})}{2} \ge \int \varphi \,\mathrm{d}\nu^{\perp} + \int \psi \,\mathrm{d}\nu^{\mathrm{co}}$$

$$= \frac{W_*^2(\nu,\nu^{\mathrm{co}})}{2} + \int \varphi \,\mathrm{d}(\nu^{\perp}-\nu)$$

Has the right sign (superadditivity of convex functions)

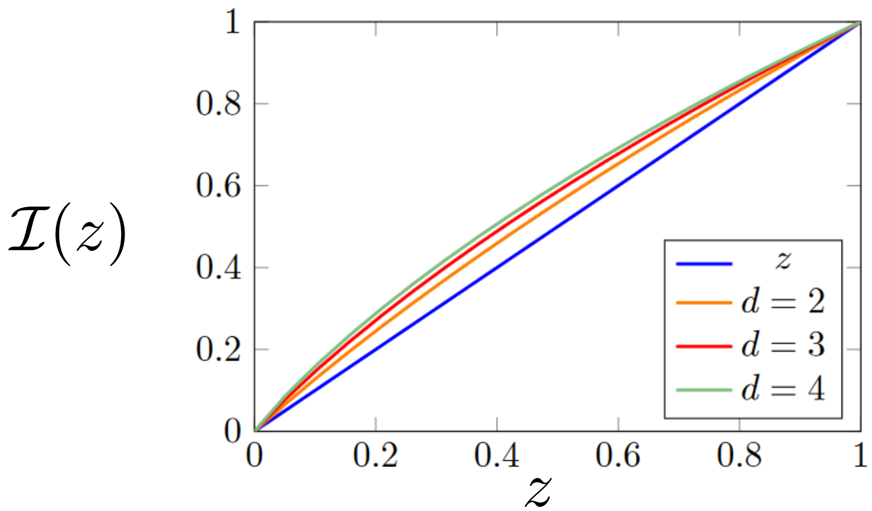
$$\int \varphi \,\mathrm{d}(\nu^{\perp}-\nu) = \int u(s_1+s_2)\,\mathrm{d}\nu(s_1,s_2) - \left(\int u(s_1)\mathrm{d}\bar{\nu}(s_1) + \int u(s_2)\mathrm{d}\bar{\nu}(s_2) \right)$$

Examples

Additive model

Parameter $z \in [0, 1]$,

$$\nu = (1 - z)\nu^{\perp} + z\nu^{\mathrm{co}}$$



$\mathcal{I}(z) \geq z$ [$=$ Covariance if $d = 2$]

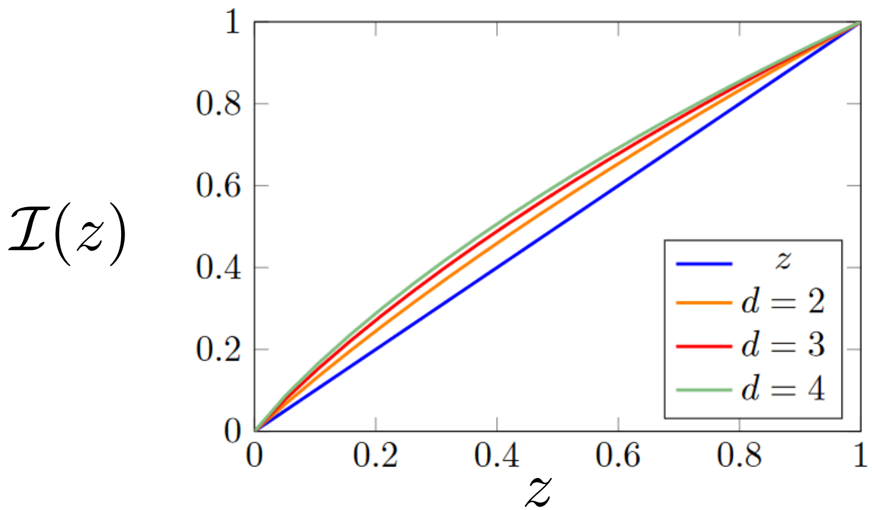
Lijoi, Nipoti and Prünster (2014). Bayesian inference with dependent normalized completely random measures.

Examples

Additive model

Parameter $z \in [0,1]$,

$$\nu = (1-z)\nu^{\perp} + z\nu^{\mathrm{co}}$$



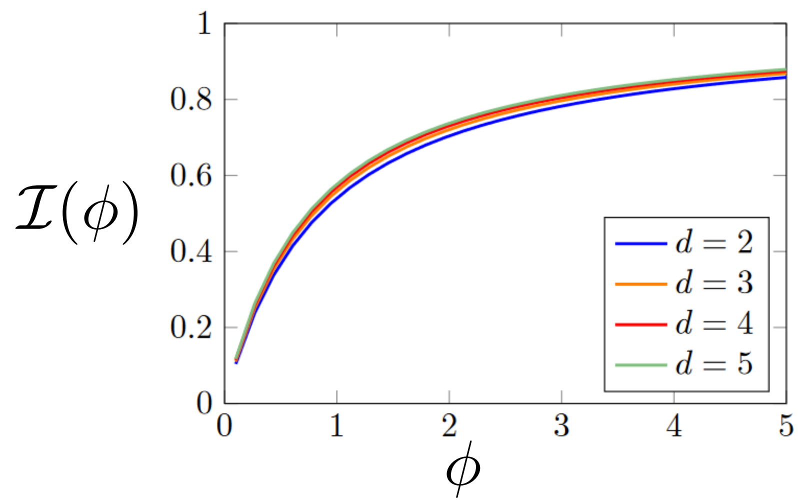
$\mathcal{I}(z) \geq z$ [$=$ Covariance if $d = 2$]

Compound random measures

Parameter ϕ measures dependence

$$\nu(s_1, \dots, s_d) = \int_0^{+\infty} h^{\phi}\left(\frac{s_1}{u}, \dots, \frac{s_d}{u}\right) \mathrm{d}\nu_*^{\phi}(u)$$

for well chosen h^{ϕ}, ν_*^{ϕ}.



Lijoi, Nipoti and Prünster (2014). Bayesian inference with dependent normalized completely random measures.
Griffin and Leisen (2017). Compound random measures and their use in bayesian non-parametrics.

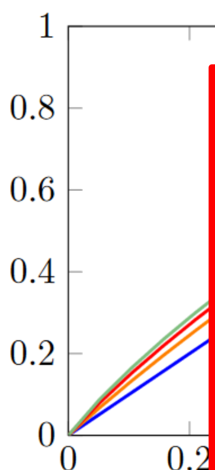
Examples

Additive model

Parameter $z \in [0, 1]$,

$$\nu = (1 - z)\nu^{\perp} + z\nu^{\text{co}}$$

$\mathcal{I}(z)$

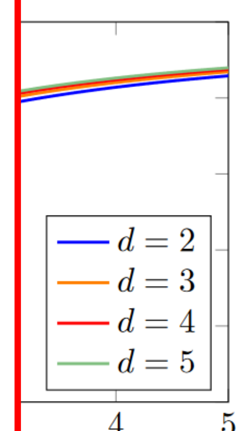


$\mathcal{I}(z) \geq z \, [=$

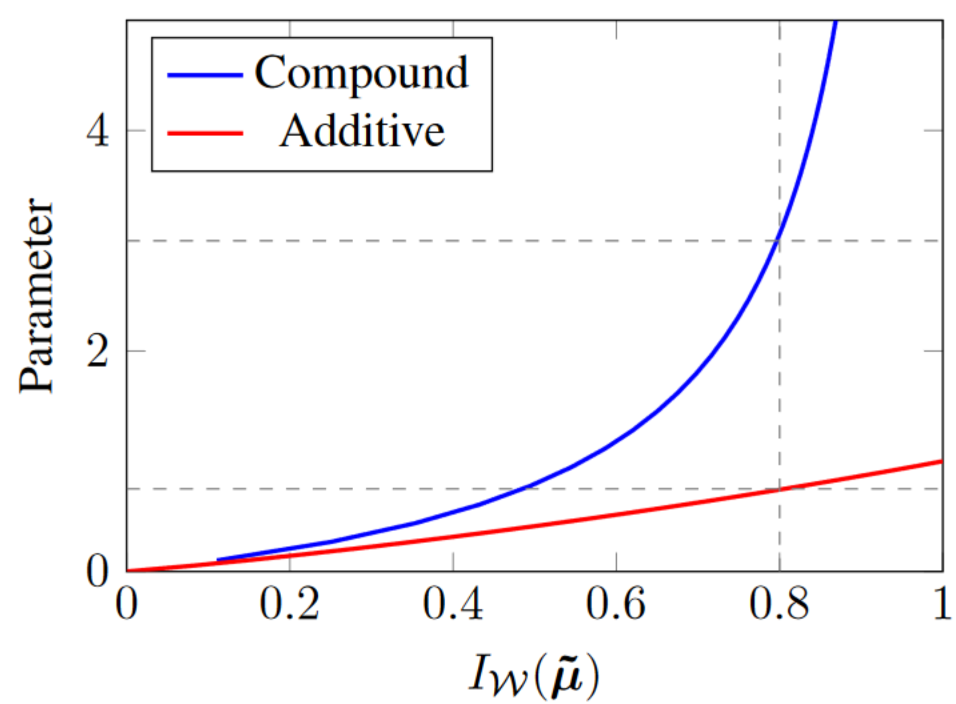
Compound random measures

Parameter ϕ measures dependence

$$\nu(s_1, \ldots, s_d) = \int_0^{+\infty} h^{\phi}\left(\frac{s_1}{u}, \ldots, \frac{s_d}{u}\right) \mathrm{d}\nu_*^{\phi}(u)$$



Model comparision



Lijoi, Nipoti and Prünste

Griffin and Leisen (2017)

Conclusion

What is done:

- Wasserstein distance between Lévy measures.
- Index of dependence between Completely Random Vectors.

What's next?:

- Study dependence in the posterior.
- Use this distance for other purposes: merging of opinions, hypothesis testing, etc.

For this, need to extend the distance to couple both atoms and jumps, to Cox processes.

Catalano, Lavenant (2023+). Merging Rate of Opinions via Optimal Transport on Random Measures.

Conclusion

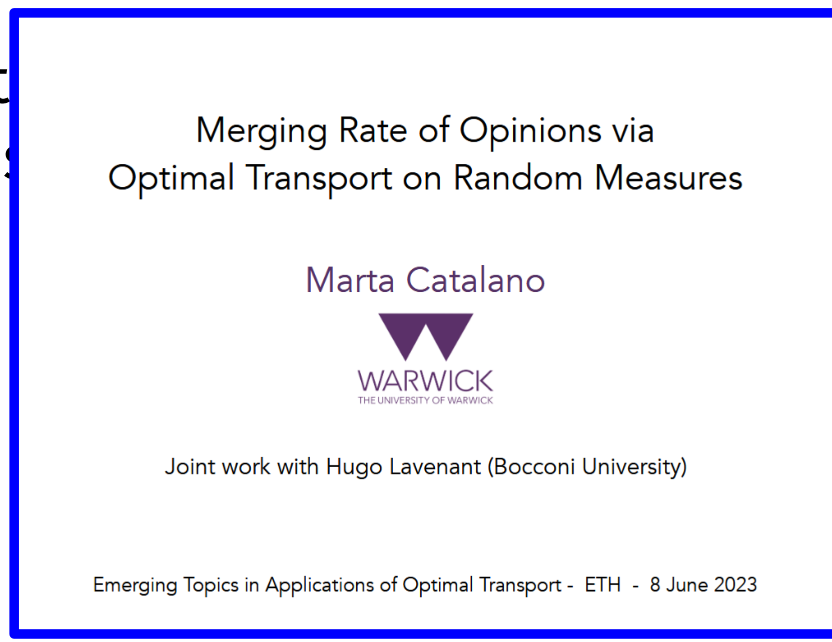
What is done:

- Wasserstein distance between Lévy measures.
- Index of dependence between Completely Random Vectors.

What's next?:

- Study dependence in the posterior.
- Use this distance for other purposes: merging of opinions, hypothesis testing, etc.

For this, need to ext oth atoms and jumps, to Cox proce



To be presented in a few minutes!

Catalano, Lavenant (2023+). Merging Rate of Opinions via Optimal Transport on Random Measures.

Conclusion

What is done:

- Wasserstein distance between Lévy measures.
- Index of dependence between Completely Random Vectors.

What's next?:

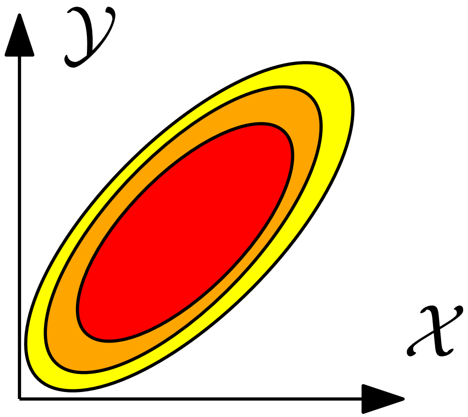
- Study dependence in the posterior.
- Use this distance for other purposes: merging of opinions, hypothesis testing, etc.

For this, need to extend the distance to couple both atoms and jumps, to Cox processes.

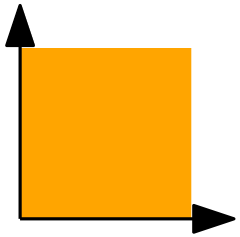
Thank you for your attention

With the Manhattan distance

$(\mathcal{X}, d_\mathcal{X})$ and $(\mathcal{Y}, d_\mathcal{Y})$ metric spaces with probabilies $p_\mathcal{X}, p_\mathcal{Y}$



$p \in \Pi(p_\mathcal{X}, p_\mathcal{Y})$

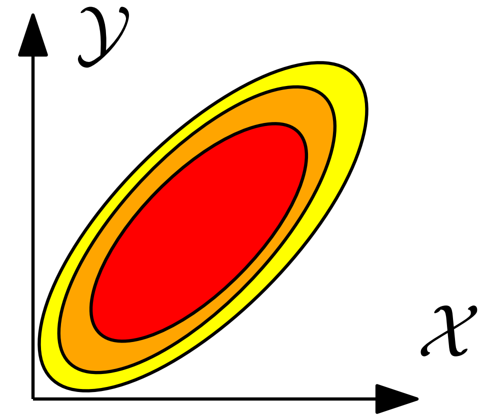


$$p^\perp = p_\mathcal{X} \otimes p_\mathcal{Y}$$

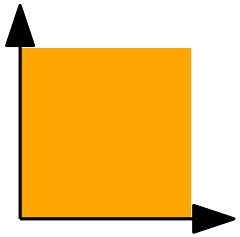
Nies, Staudt, and Munk (2021). Transport dependency: Optimal transport based dependency measures.

With the Manhattan distance

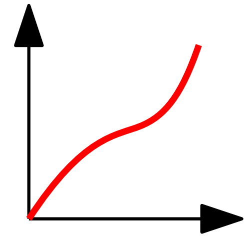
$(\mathcal{X}, d_{\mathcal{X}})$ and $(\mathcal{Y}, d_{\mathcal{Y}})$ metric spaces with probabilies $p_{\mathcal{X}}, p_{\mathcal{Y}}$



$p \in \Pi(p_{\mathcal{X}}, p_{\mathcal{Y}})$



$p^{\perp} = p_{\mathcal{X}} \otimes p_{\mathcal{Y}}$



p^*

Theorem.
Define $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$ with

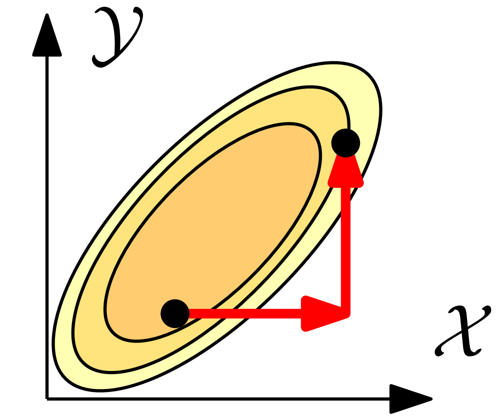
Then $$W_2(p, p^{\perp})^2 \leq \iint d_{\mathcal{Y}}(y_1, y_2)^2 \,\mathrm{d}p_{\mathcal{Y}}(y)\mathrm{d}p_{\mathcal{Y}}(y')$$

with equality iff p on the graph of 1 -Lipschitz function.

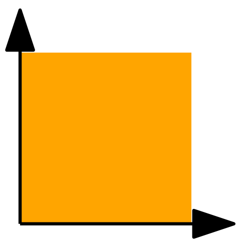
Nies, Staudt, and Munk (2021). Transport dependency: Optimal transport based dependency measures.

With the Manhattan distance

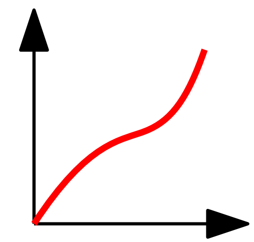
$(\mathcal{X}, d_{\mathcal{X}})$ and $(\mathcal{Y}, d_{\mathcal{Y}})$ metric spaces with probabilies $p_{\mathcal{X}}, p_{\mathcal{Y}}$



$p \in \Pi(p_{\mathcal{X}}, p_{\mathcal{Y}})$



$p^{\perp} = p_{\mathcal{X}} \otimes p_{\mathcal{Y}}$



p^*

Theorem.
Define $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$ with

$$d_{\mathcal{Z}}((x_1, y_1), (x_2, y_2)) = \quad d_{\mathcal{X}}(x_1, x_2) + d_{\mathcal{Y}}(y_1, y_2).$$

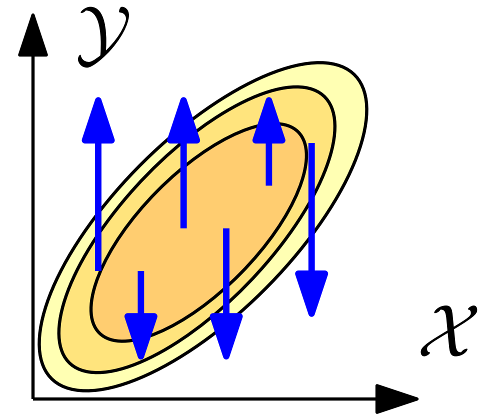
Then $$W_2(p, p^{\perp})^2 \leq \iint d_{\mathcal{Y}}(y_1, y_2)^2 \, \mathrm{d}p_{\mathcal{Y}}(y) \mathrm{d}p_{\mathcal{Y}}(y')$$

with equality iff p on the graph of 1 -Lipschitz function.

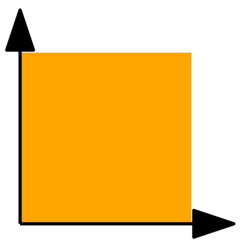
Nies, Staudt, and Munk (2021). Transport dependency: Optimal transport based dependency measures.

With the Manhattan distance

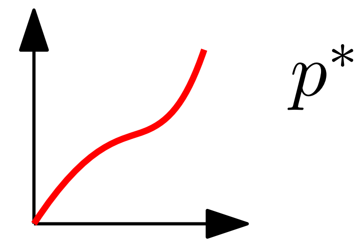
$(\mathcal{X}, d_\mathcal{X})$ and $(\mathcal{Y}, d_\mathcal{Y})$ metric spaces with probabilies $p_\mathcal{X}, p_\mathcal{Y}$



$p \in \Pi(p_\mathcal{X}, p_\mathcal{Y})$



$p^\perp = p_\mathcal{X} \otimes p_\mathcal{Y}$



Theorem.
Define $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$ with

$$d_\mathcal{Z}((x_1, y_1), (x_2, y_2)) = d_\mathcal{X}(x_1, x_2) + d_\mathcal{Y}(y_1, y_2).$$

Then $$W_2(p, p^\perp)^2 \leq \iint d_\mathcal{Y}(y_1, y_2)^2 \, \mathrm{d}p_\mathcal{Y}(y) \mathrm{d}p_\mathcal{Y}(y')$$

with equality iff p on the graph of 1-Lipschitz function.

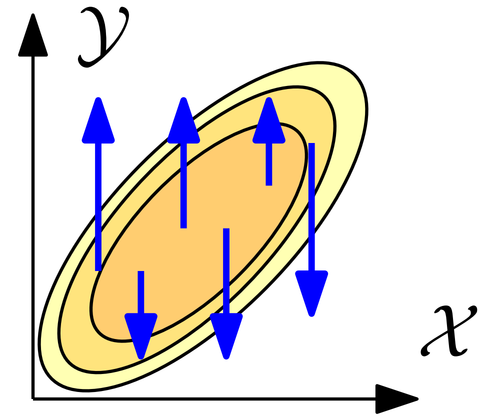
Proof idea. Send (x, y) onto (x, y') with y' independent of (x, y).

Annex

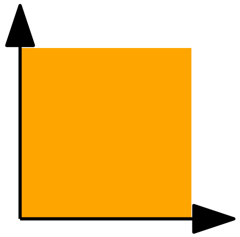
Nies, Staudt, and Munk (2021). Transport dependency: Optimal transport based dependency measures.

With the Manhattan distance

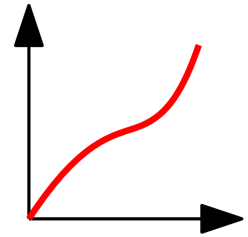
$(\mathcal{X}, d_{\mathcal{X}})$ and $(\mathcal{Y}, d_{\mathcal{Y}})$ metric spaces with probabilies $p_{\mathcal{X}}, p_{\mathcal{Y}}$



$p \in \Pi(p_{\mathcal{X}}, p_{\mathcal{Y}})$



$p^{\perp} = p_{\mathcal{X}} \otimes p_{\mathcal{Y}}$



p^*

Theorem.
Define $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$ with

$$d_{\mathcal{Z}}((x_1, y_1), (x_2, y_2)) = \alpha \, d_{\mathcal{X}}(x_1, x_2) + d_{\mathcal{Y}}(y_1, y_2).$$

Then $$W_2(p, p^{\perp})^2 \leq \iint d_{\mathcal{Y}}(y_1, y_2)^2 \, \mathrm{d}p_{\mathcal{Y}}(y) \mathrm{d}p_{\mathcal{Y}}(y')$$

with equality iff p on the graph of α -Lipschitz function.

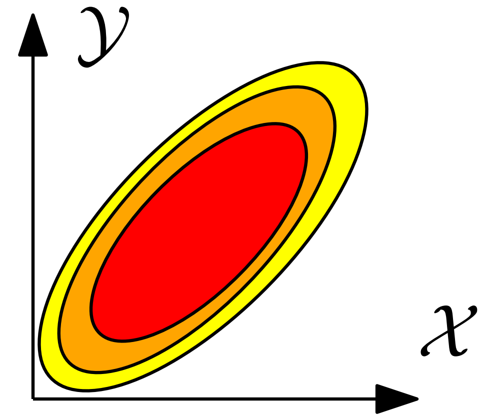
Proof idea. Send (x, y) onto (x, y') with y' independent of (x, y).

Annex

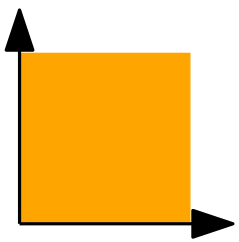
Nies, Staudt, and Munk (2021). Transport dependency: Optimal transport based dependency measures.

With the Manhattan distance

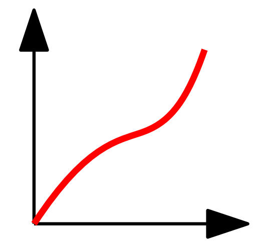
$(\mathcal{X}, d_{\mathcal{X}})$ and $(\mathcal{Y}, d_{\mathcal{Y}})$ metric spaces with probabilies $p_{\mathcal{X}}, p_{\mathcal{Y}}$



$p \in \Pi(p_{\mathcal{X}}, p_{\mathcal{Y}})$



$p^{\perp} = p_{\mathcal{X}} \otimes p_{\mathcal{Y}}$



p^*

Theorem.
Define $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$ with

$$d_{\mathcal{Z}}((x_1, y_1), (x_2, y_2)) = \alpha\, d_{\mathcal{X}}(x_1, x_2) + d_{\mathcal{Y}}(y_1, y_2).$$

Then $$W_2(p, p^{\perp})^2 \leq \iint d_{\mathcal{Y}}(y_1, y_2)^2 \, \mathrm{d}p_{\mathcal{Y}}(y) \mathrm{d}p_{\mathcal{Y}}(y')$$

with equality iff p on the graph of α -Lipschitz function.

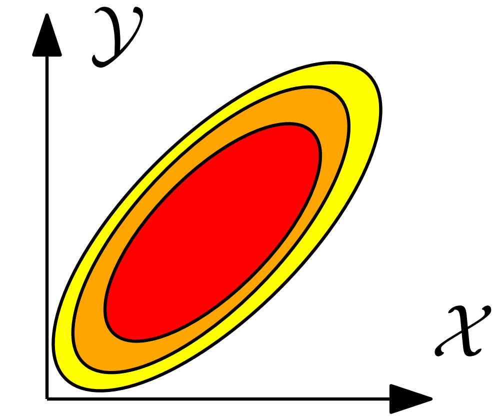
- Defined in arbitrary metric spaces.
- Inherits properties of optimal transport.

Nies, Staudt, and Munk (2021). Transport dependency: Optimal transport based dependency measures.

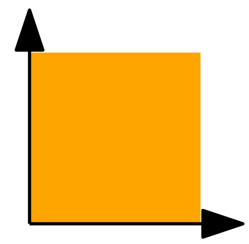
Annex

With the Manhattan distance

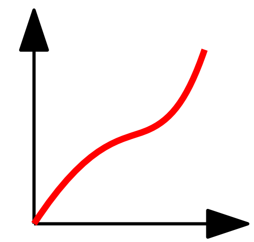
$(\mathcal{X}, d_\mathcal{X})$ and $(\mathcal{Y}, d_\mathcal{Y})$ metric spaces with probabilies $p_\mathcal{X}, p_\mathcal{Y}$



$p \in \Pi(p_\mathcal{X}, p_\mathcal{Y})$



$p^\perp = p_\mathcal{X} \otimes p_\mathcal{Y}$



p^*

Theorem.
Define $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$ with

$$d_\mathcal{Z}((x_1, y_1), (x_2, y_2)) = \alpha\, d_\mathcal{X}(x_1, x_2) + d_\mathcal{Y}(y_1, y_2).$$

Then $$W_2(p, p^\perp)^2 \leq \iint d_\mathcal{Y}(y_1, y_2)^2 \,\mathrm{d}p_\mathcal{Y}(y)\mathrm{d}p_\mathcal{Y}(y')$$

with equality iff p on the graph of α -Lipschitz function.

- Defined in arbitrary metric spaces.
- Inherits properties of optimal transport.

- Not tractable analytically.
- Sampling: curse of dimensionality.

Annex

Nies, Staudt, and Munk (2021). Transport dependency: Optimal transport based dependency measures.