Wasserstein distance between Lévy measures with applications to Bayesian nonparametrics

Hugo Lavenant

Bocconi University

2022 ISBA world meeting, Montreal (Canada), June 28 2022

Joint work with:

Marta Catalano

Antonio Lijoi

Igor Prünster

Joint work with:

WARWICK
THE UNIVERSITY OF WARWICK

Marta Catalano

Antonio Lijoi

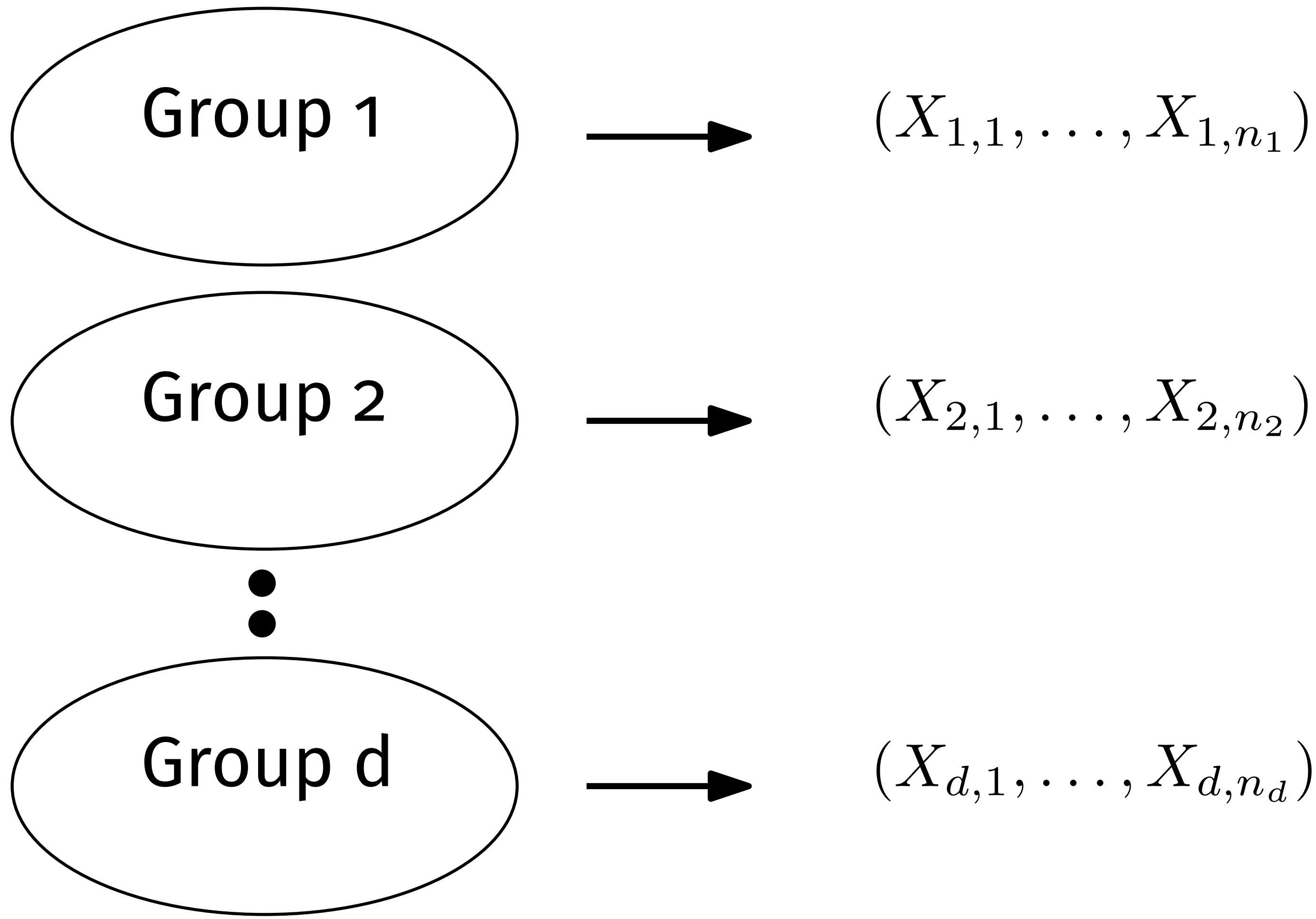
Igor Prünster

Disclaimer

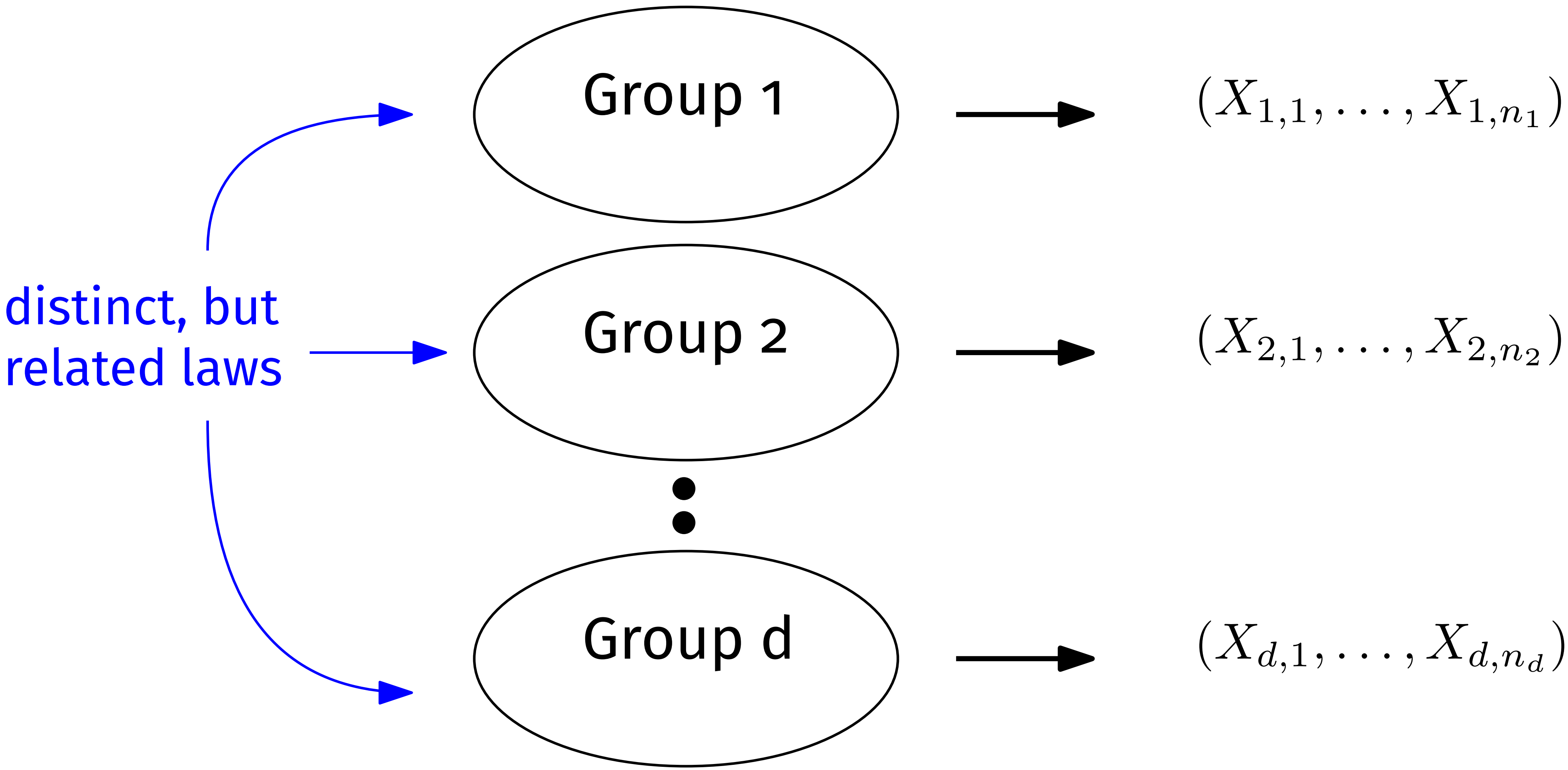
I am not a (Bayesian) statistican.

My background: mathematical analysis, optimal transport.

Quantifying dependence

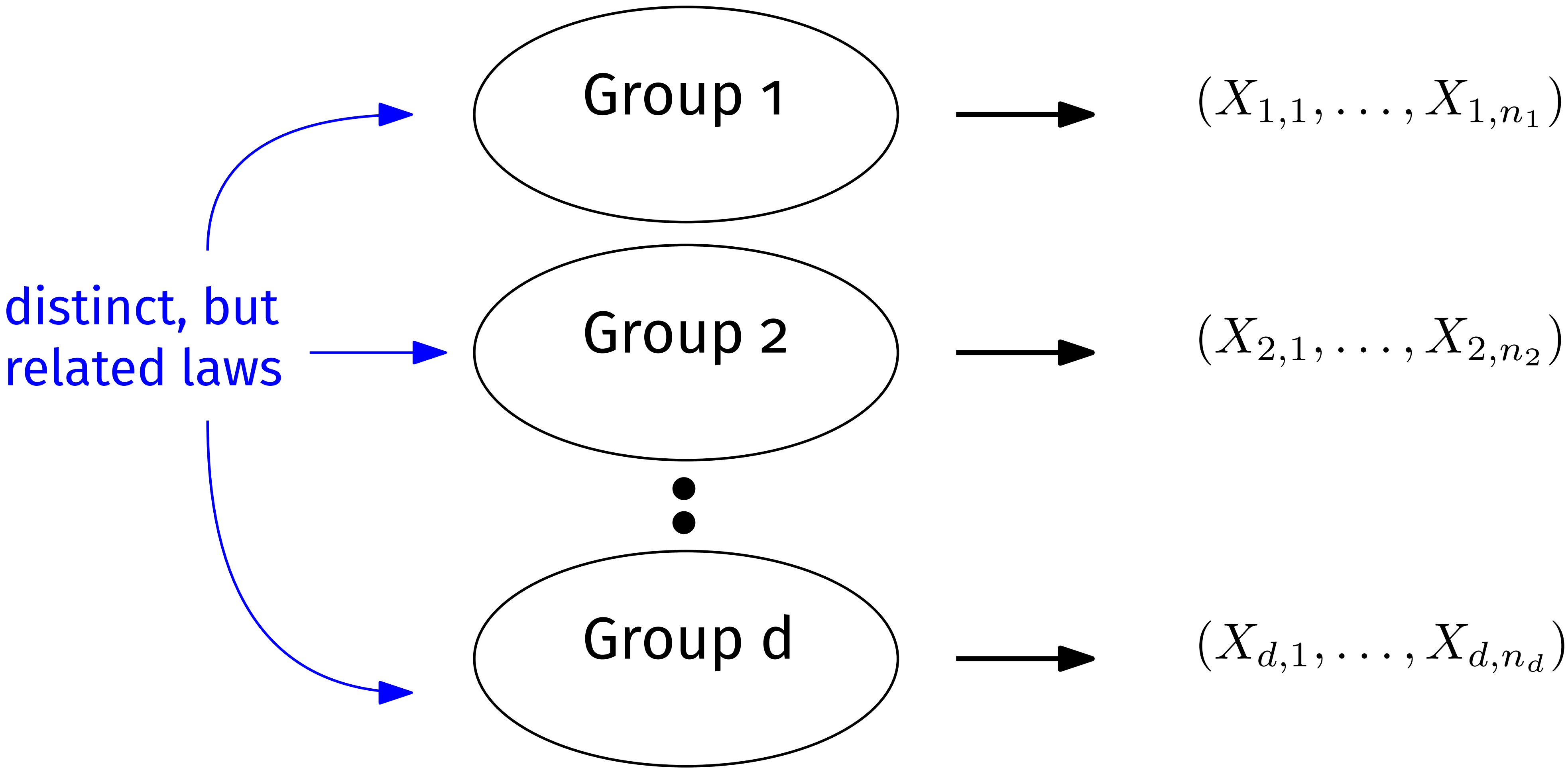


Quantifying dependence



Bayesian inference allows for borrowing of information

Quantifying dependence



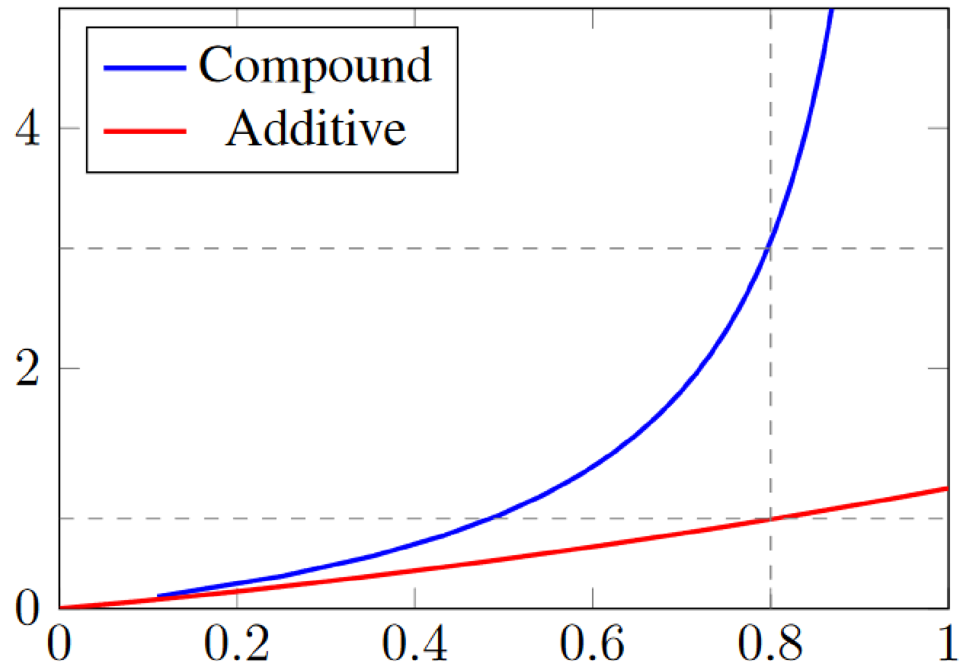
Bayesian inference allows for borrowing of information

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

Snapshot of the final result

Our contribution: an index of dependence quantifying dependence in the prior

Different parametrized models of prior

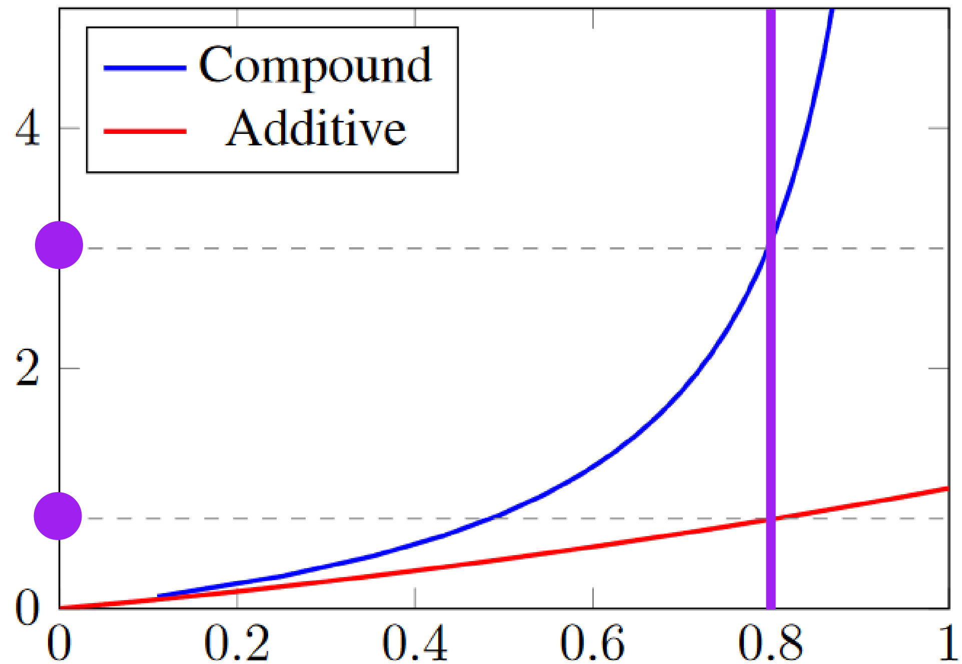


Value of the index

Snapshot of the final result

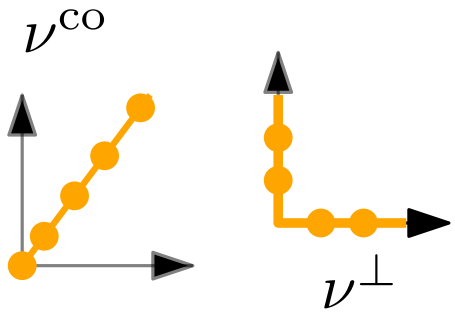
Our contribution: an index of dependence quantifying dependence in the prior

"Compound" with parameter 3 has same dependence as "Additive" with parameter 0.75.



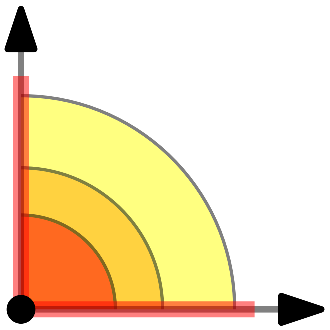
Value of the index

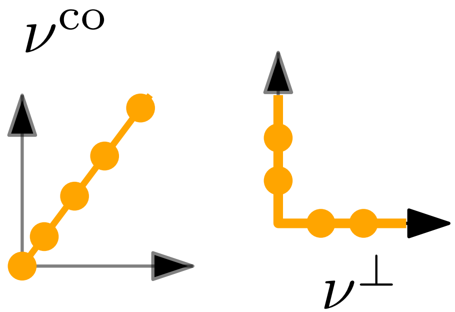
Allow for comparision between **different** priors



1 - Context, general strategy

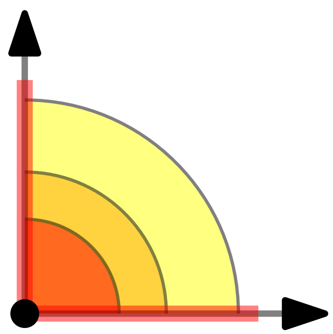
2 - Building the index with optimal transport





1 - Context, general strategy

2 - Building the index with optimal transport



Bayesian Non Parametrics

$\tilde{p}$ random probability measure on $\mathbb{X}$

$$X_1, X_2, \ldots, X_n | \tilde{p} \overset{\text{i.i.d.}}{\sim} \tilde{p}$$

(justified by exchangeability)

Bayesian Non Parametrics

$\tilde{p}$ random probability measure on $\mathbb{X}$

$$X_1, X_2, \ldots, X_n \,|\, \tilde{p} \overset{\text{i.i.d.}}{\sim} \tilde{p}$$

(justified by exchangeability)

Group 1 $\tilde{p}_1$ $\longrightarrow$ $X_{1,1}, \ldots, X_{1,n_1} \,|\, (\tilde{p}_1, \ldots, \tilde{p}_d) \overset{\text{i.i.d.}}{\sim} \tilde{p}_1$

Group 2 $\tilde{p}_2$ $\longrightarrow$ $X_{2,1}, \ldots, X_{2,n_2} \,|\, (\tilde{p}_1, \ldots, \tilde{p}_d) \overset{\text{i.i.d.}}{\sim} \tilde{p}_2$

$\vdots$

Group d $\tilde{p}_d$ $\longrightarrow$ $X_{d,1}, \ldots, X_{d,n_d} \,|\, (\tilde{p}_1, \ldots, \tilde{p}_d) \overset{\text{i.i.d.}}{\sim} \tilde{p}_d$

(justified by partial exchangeability)

Specific setting: Completely Random Vectors

$\tilde{\boldsymbol{\mu}} = (\tilde{\mu}_1, \tilde{\mu}_2, \ldots, \tilde{\mu}_d)$ Completely Random Vector

$$X_{1,1}, X_{1,2}, \ldots, X_{1,n_1} \,|\, \tilde{\boldsymbol{\mu}} \overset{\text{i.i.d.}}{\sim} \frac{\tilde{\mu}_1}{\tilde{\mu}_1(\mathbb{X})}$$

$$X_{2,1}, X_{2,2}, \ldots, X_{2,n_2} \,|\, \tilde{\boldsymbol{\mu}} \overset{\text{i.i.d.}}{\sim} \frac{\tilde{\mu}_2}{\tilde{\mu}_2(\mathbb{X})}$$

$$\vdots$$

$$X_{d,1}, X_{d,2}, \ldots, X_{d,n_d} \,|\, \tilde{\boldsymbol{\mu}} \overset{\text{i.i.d.}}{\sim} \frac{\tilde{\mu}_d}{\tilde{\mu}_d(\mathbb{X})}$$

Definition (CRV). 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 $\mathbb{R}_+^d$.

Specific setting: Completely Random Vectors

$\tilde{\boldsymbol{\mu}} = (\tilde{\mu}_1, \tilde{\mu}_2, \ldots, \tilde{\mu}_d)$ Completely Random Vector

Contains all dependence in the prior

$$X_{1,1}, X_{1,2}, \ldots, X_{1,n_1} \mid \tilde{\boldsymbol{\mu}} \overset{\text{i.i.d.}}{\sim} \frac{\tilde{\mu}_1}{\tilde{\mu}_1(\mathbb{X})}$$

$$X_{2,1}, X_{2,2}, \ldots, X_{2,n_2} \mid \tilde{\boldsymbol{\mu}} \overset{\text{i.i.d.}}{\sim} \frac{\tilde{\mu}_2}{\tilde{\mu}_2(\mathbb{X})}$$

$$\vdots$$

$$X_{d,1}, X_{d,2}, \ldots, X_{d,n_d} \mid \tilde{\boldsymbol{\mu}} \overset{\text{i.i.d.}}{\sim} \frac{\tilde{\mu}_d}{\tilde{\mu}_d(\mathbb{X})}$$

Definition (CRV). 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 $\mathbb{R}_+^d$.

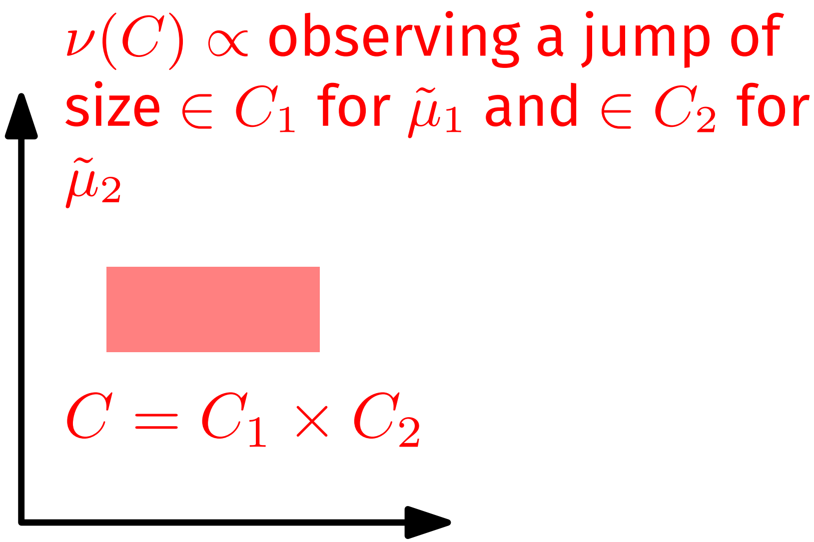
For $A \subseteq \mathbb{X}$, the random variables $\tilde{\mu}_1(A), \ldots, \tilde{\mu}_d(A)$ may be dependent.

Lévy measure of a Completely Random Vector

Assumptions of **homogeneity** and no fixed atoms:

$$\tilde{\boldsymbol{\mu}} = \sum_{i=1}^{\infty} \tilde{\mathbf{J}}_i \delta_{Y_i}$$

where $(Y_i)_i \in \mathbb{X}$ (**atoms**) follow base measure P_0; and $(\tilde{\mathbf{J}}_i)_i$ (**jumps**) independent from $(Y_i)_i$ follow Poisson point cloud on $\mathbb{R}_+^d$ with intensity measure ν (**Lévy measure**).

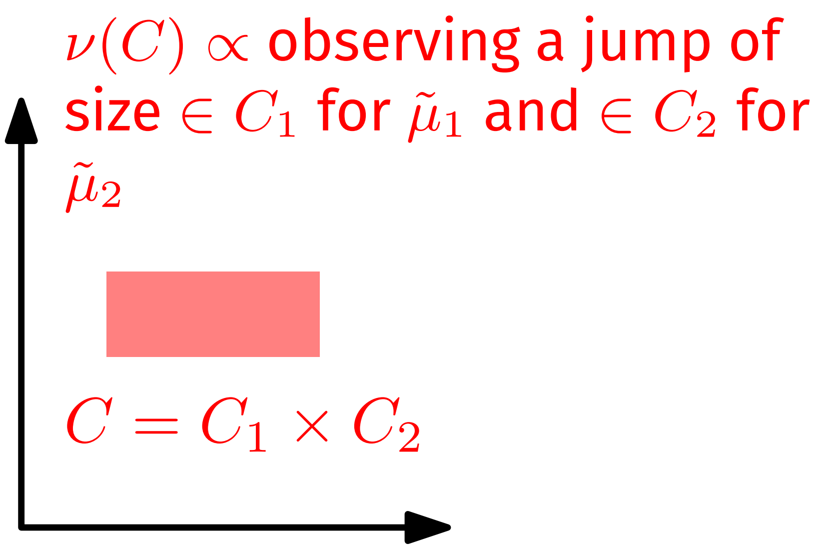


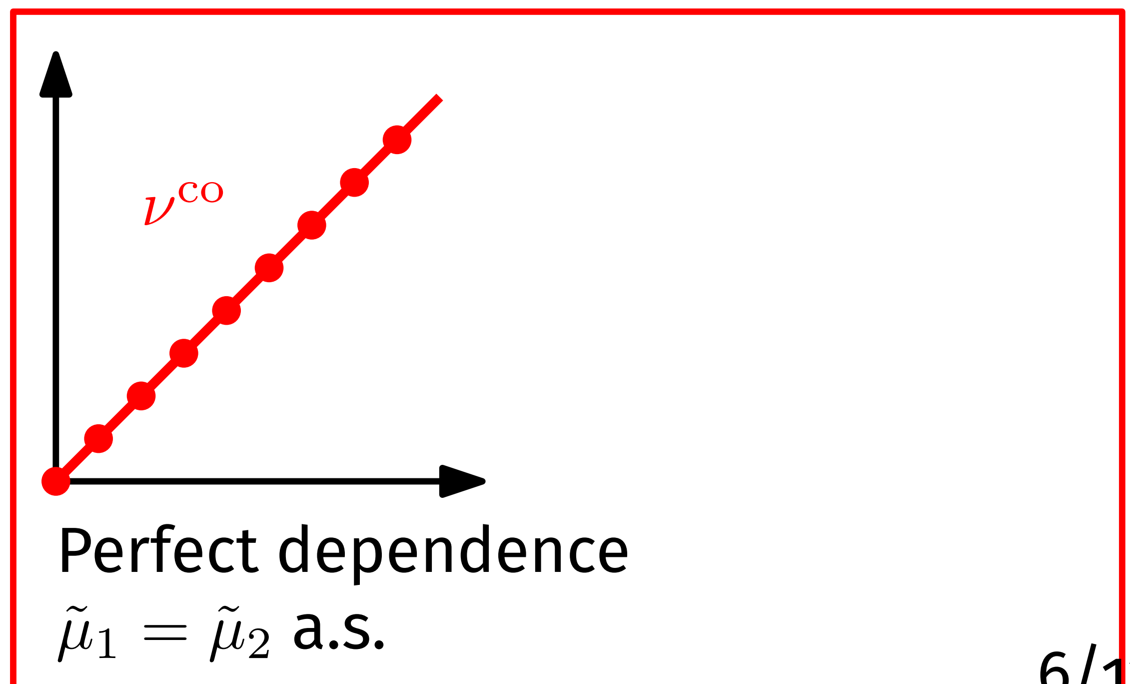
Lévy measure of a Completely Random Vector

Assumptions of **homogeneity** and no fixed atoms:

$$\tilde{\boldsymbol{\mu}} = \sum_{i=1}^{\infty} \tilde{\mathbf{J}}_i \delta_{Y_i}$$

where $(Y_i)_i \in \mathbb{X}$ (**atoms**) follow base measure P_0; and $(\tilde{\mathbf{J}}_i)_i$ (**jumps**) independent from $(Y_i)_i$ follow Poisson point cloud on $\mathbb{R}_+^d$ with intensity measure ν (**Lévy measure**).



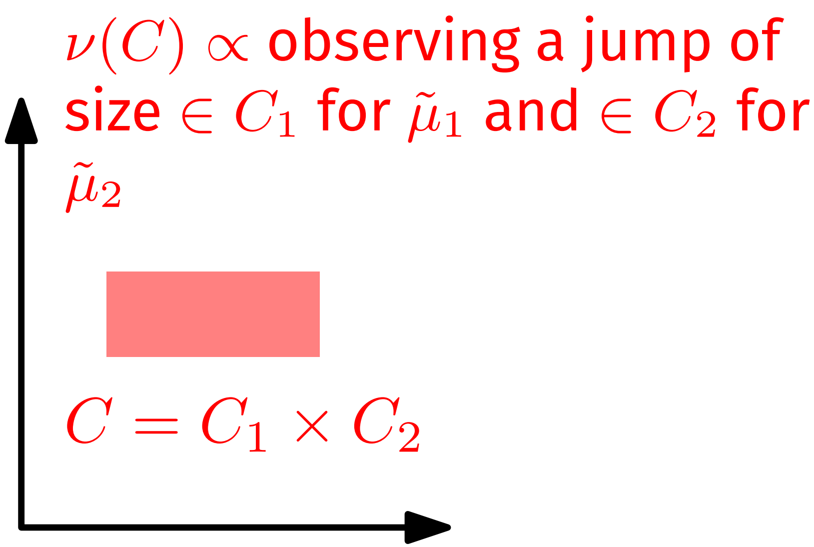


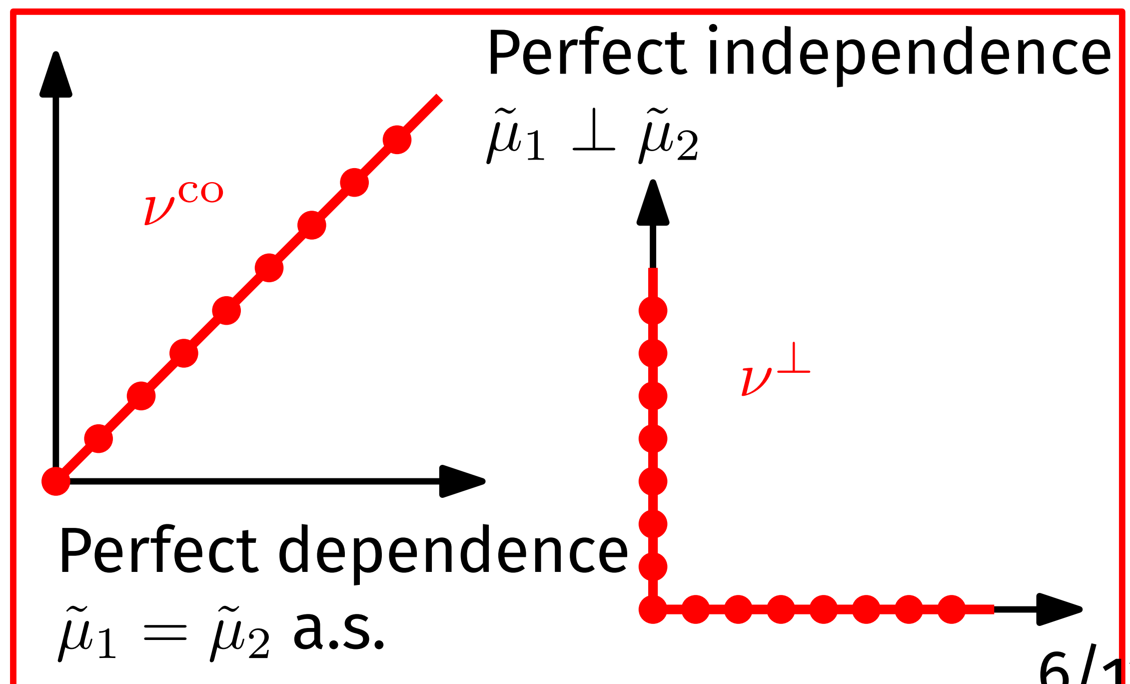
Lévy measure of a Completely Random Vector

Assumptions of **homogeneity** and no fixed atoms:

$$\tilde{\boldsymbol{\mu}} = \sum_{i=1}^{\infty} \tilde{\mathbf{J}}_i \delta_{Y_i}$$

where $(Y_i)_i \in \mathbb{X}$ (**atoms**) follow base measure P_0; and $(\tilde{\mathbf{J}}_i)_i$ (**jumps**) independent from $(Y_i)_i$ follow Poisson point cloud on $\mathbb{R}_+^d$ with intensity measure ν (**Lévy measure**).



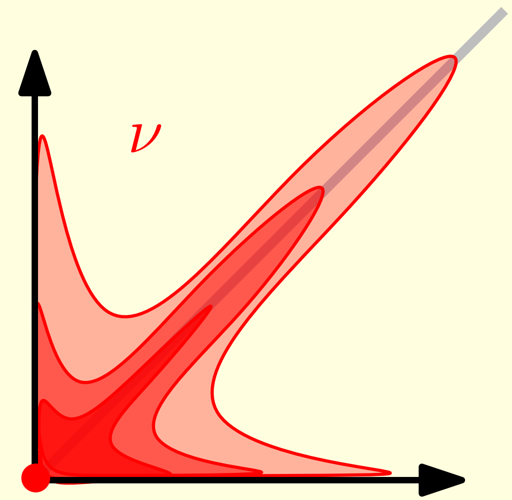


Lévy measure of a Completely Random Vector

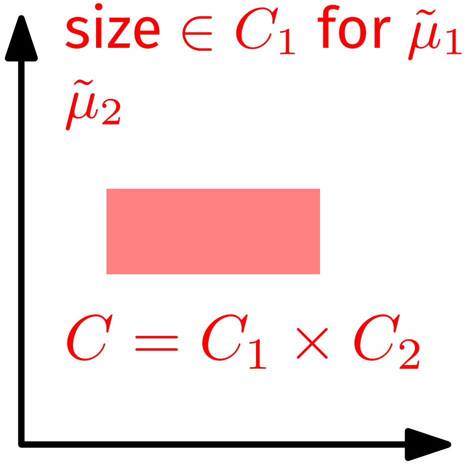
Assumpti

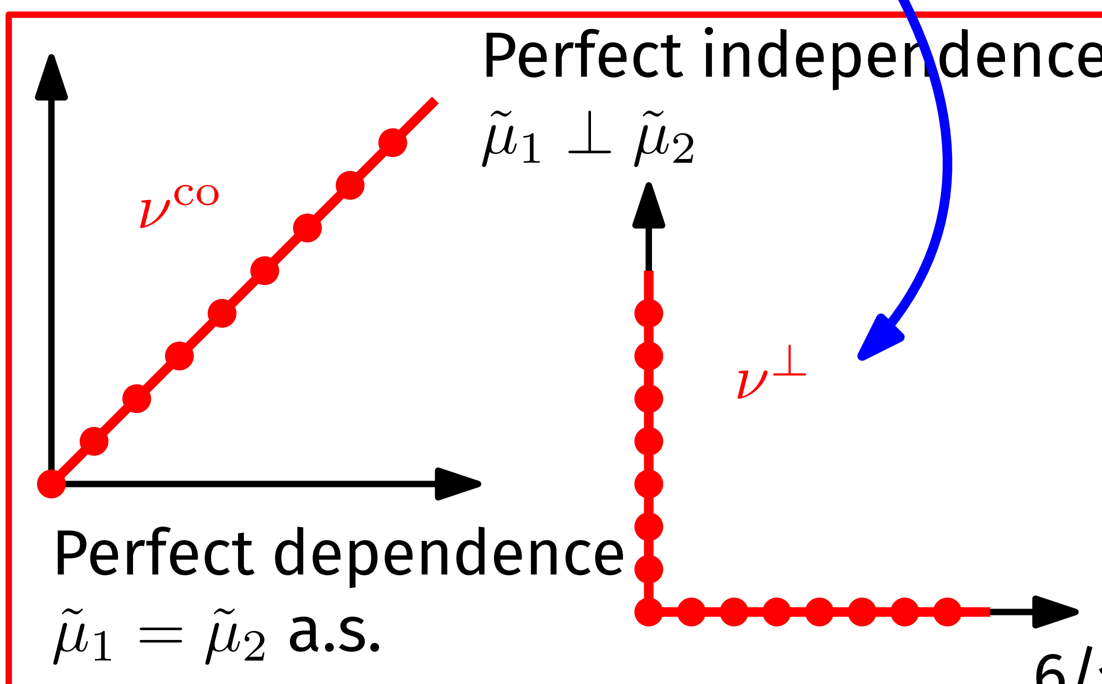
where $(Y$ $(\tilde{\mathbf{J}}_i)_i$
(jumps) i cloud
on $\mathbb{R}_+^d$ wi

Goal: distinguish between these two cases.



$\nu(C) \propto$ observing a jump of size $\in C_1$ for $\tilde{\mu}_1$ and $\in C_2$ for $\tilde{\mu}_2$

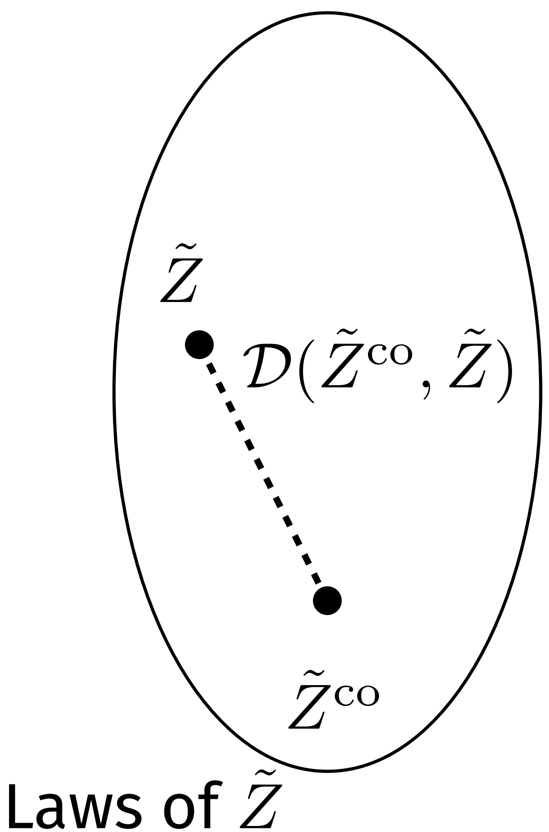




A general method to construct an index

Ingredients:

- $\tilde{Z}$ random object, $\tilde{Z}^{\mathrm{co}}$ "most dependent".
- $\mathcal{D}$ "discrepancy" between laws of random objects.

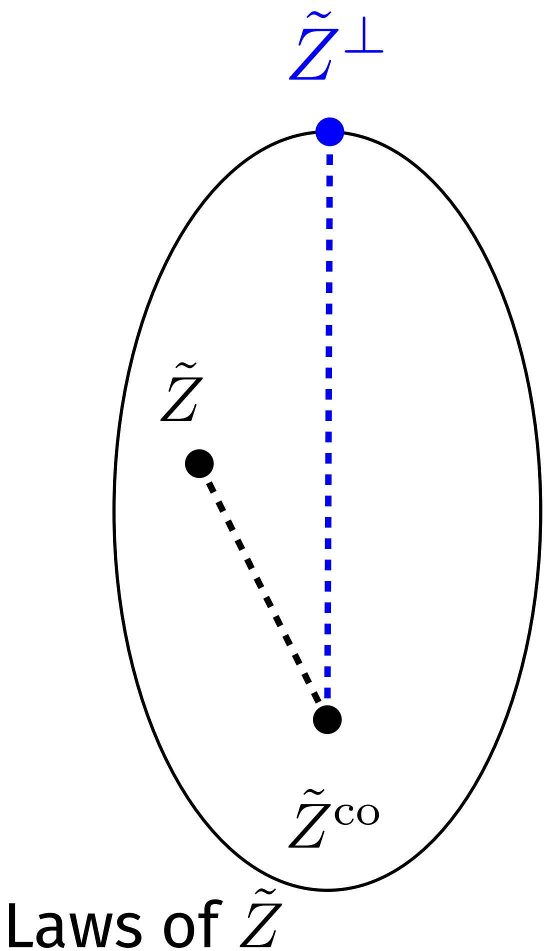


Móri, and Székely (2020). The Earth Mover's correlation.

A general method to construct an index

Ingredients:

- $\tilde{Z}$ random object, $\tilde{Z}^{\mathrm{co}}$ “most dependent”.
- $\mathcal{D}$ “discrepancy” between laws of random objects.



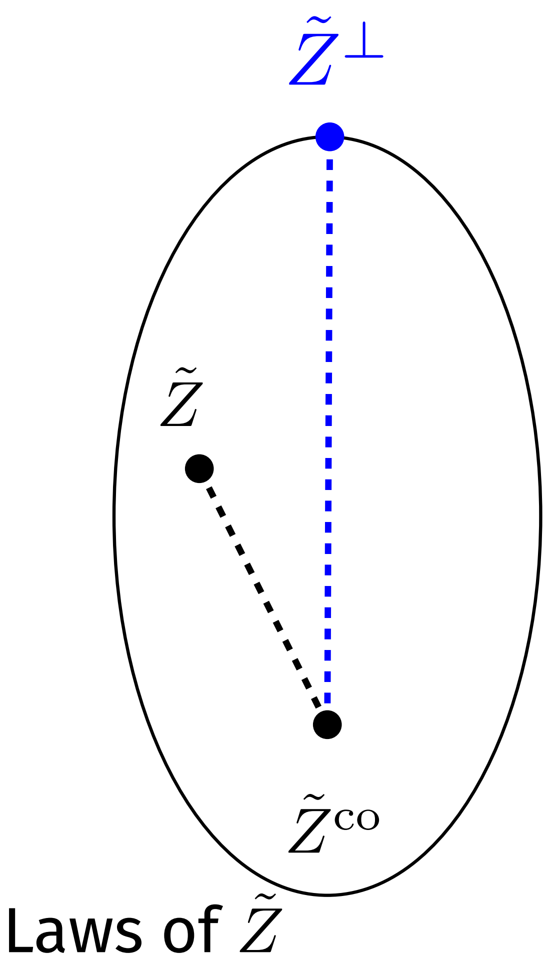
To check: $\mathcal{D}(\tilde{Z}^{\mathrm{co}}, \tilde{Z})$ is maximized when $\tilde{Z} = \tilde{Z}^{\perp}$ the most independent structure.

Móri, and Székely (2020). The Earth Mover’s correlation.

A general method to construct an index

Ingredients:

- $\tilde{Z}$ random object, $\tilde{Z}^{\mathrm{co}}$ "most dependent".
- $\mathcal{D}$ "discrepancy" between laws of random objects.



To check: $\mathcal{D}(\tilde{Z}^{\mathrm{co}}, \tilde{Z})$ is maximized when $\tilde{Z} = \tilde{Z}^{\perp}$ the most independent structure.

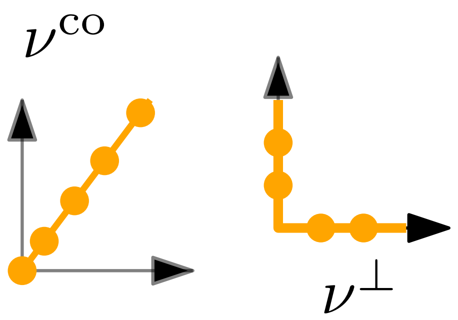
Then **define**:

$$\mathcal{I}(\tilde{Z}) = 1 - \frac{\mathcal{D}(\tilde{Z}^{\mathrm{co}}, \tilde{Z})}{\mathcal{D}(\tilde{Z}^{\mathrm{co}}, \tilde{Z}^{\perp})}.$$

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

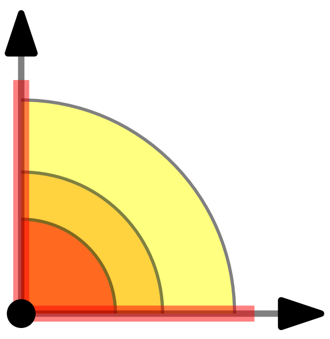
$$\mathcal{I}(\tilde{Z}^{\perp}) = 0, \qquad \mathcal{I}(\tilde{Z}^{\mathrm{co}}) = 1.$$

Móri, and Székely (2020). The Earth Mover's correlation.



1 - Context, general strategy

2 - Building the index with optimal transport



How to measure discrepancy between Completly Random Vectors?

Law of a Completly Random Vector $\tilde{\mu}$

How to measure discrepancy between Completly Random Vectors?

Law of a Completly Random Vector $\tilde{\boldsymbol{\mu}}$

Setwise evaluation $\tilde{\boldsymbol{\mu}}(A)$.

How to measure discrepancy between Completly Random Vectors?

Law of a Completly Random Vector $\tilde{\boldsymbol{\mu}}$

Setwise evaluation $\tilde{\boldsymbol{\mu}}(A)$.

Problem: not tractable

How to measure discrepancy between Completly Random Vectors?

Law of a Completly Random Vector $\tilde{\boldsymbol{\mu}}$

Setwise evaluation $\tilde{\boldsymbol{\mu}}(A)$.

Problem: not tractable

Under assumption same base measure, characterized by Lévy measure ν.

How to measure discrepancy between Completly Random Vectors?

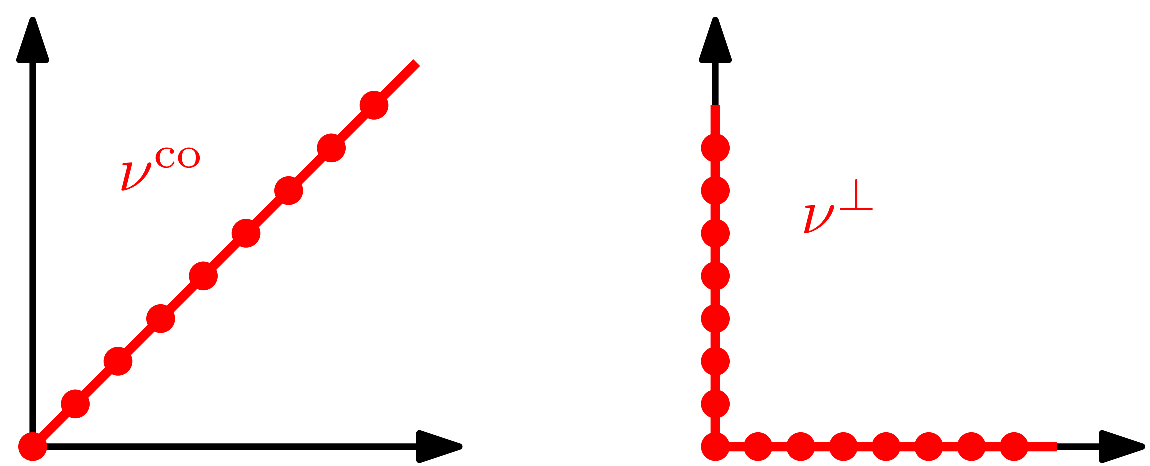
Law of a Completly Random Vector $\tilde{\mu}$

Setwise evaluation $\tilde{\boldsymbol{\mu}}(A)$.

Problem: not tractable

Under assumption same base measure, characterized by Lévy measure ν.

Problem: comparing measures with different supports



How to measure discrepancy between Completly Random Vectors?

Law of a Completly Random Vector $\tilde{\boldsymbol{\mu}}$

Setwise evaluation $\tilde{\boldsymbol{\mu}}(A)$.

Problem: not tractable

×

Under assumption same base measure, characterized by Lévy measure ν.

Problem: comparing measures with different supports

Use Optimal transport

How to measure discrepancy between Completly Random Vectors?

Law of a Completly Random Vector $\tilde{\mu}$

Setwise evaluation $\tilde{\mu}(A)$.

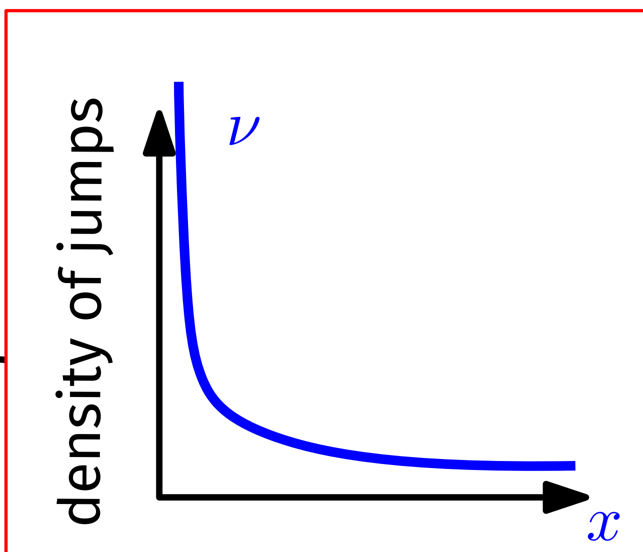
Problem: not tractable

Under assumption same base measure, characterized by Lévy measure ν.

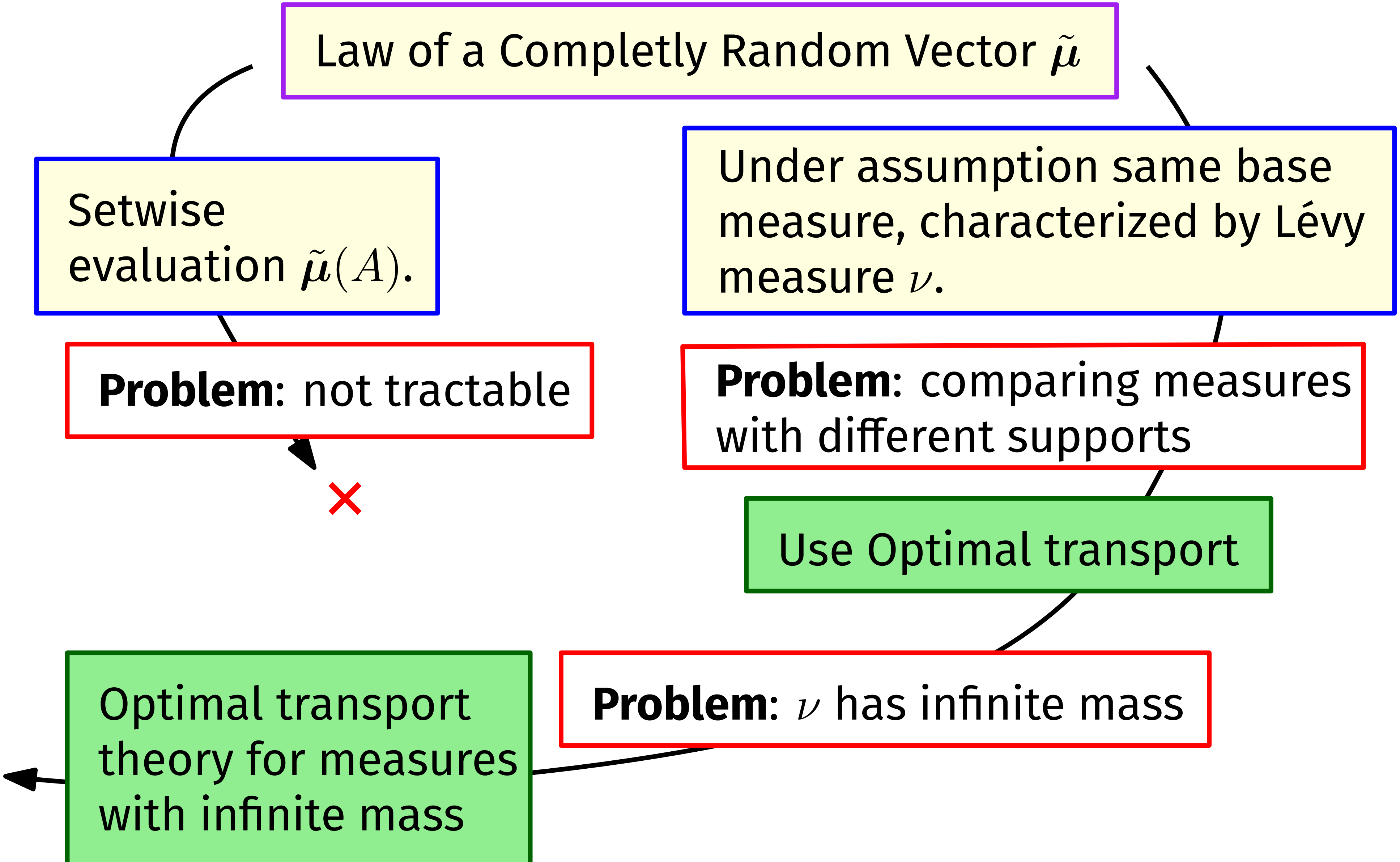
Problem: comparing measures with different supports

Use Optimal transport

Problem: ν has infinite mass



How to measure discrepancy between Completly Random Vectors?



How to measure discrepancy between Completly Random Vectors?

Law of a Completly Random Vector $\tilde{\mu}$

Setwise evaluation $\tilde{\mu}(A)$.

Problem: not tractable

Under assumption same base measure, characterized by Lévy measure ν.

Problem: comparing measures with different supports

Use Optimal transport

Problem: ν has infinite mass

Optimal transport theory for measures with infinite mass

Upper bound

(Classical) optimal transport

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

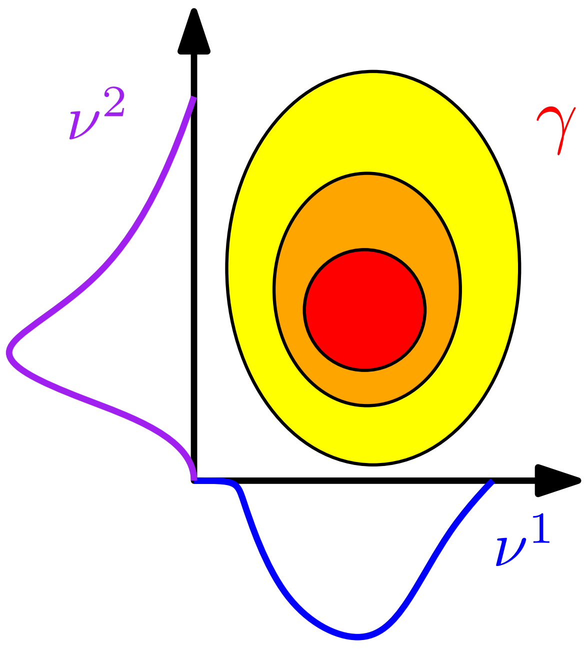
$$\mathcal{W}(\nu^1, \nu^2)^2 = \min_{(X,Y)} \left\{ \mathbb{E}\left[\|X - Y\|^2 \right] \ : \ X \sim \nu^1 \text{ and } Y \sim \nu^2 \right\}$$

(Classical) optimal transport

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

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

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



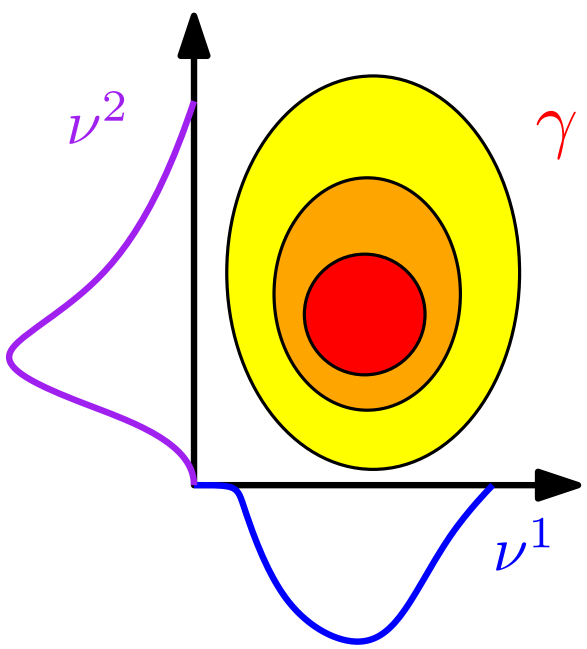
(Classical) optimal transport

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

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

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

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



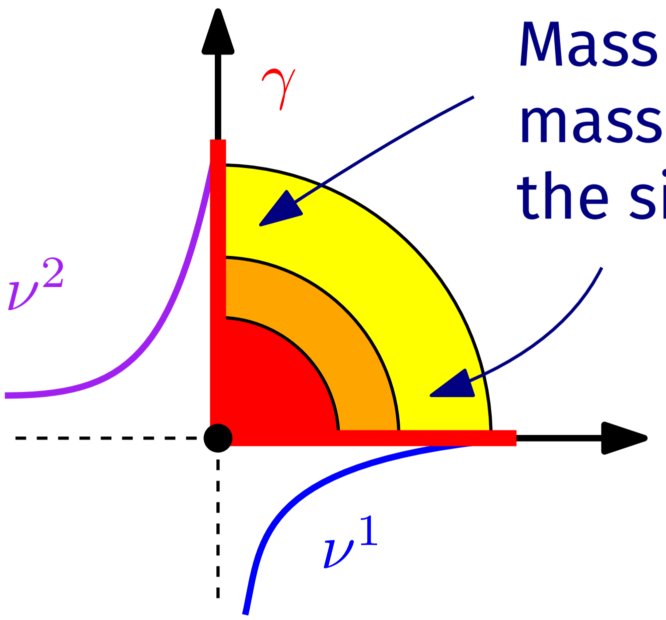
Observation. Naively, makes sense if ν^1, ν^2 have infinite mass but **finite** second moment.

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

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

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.

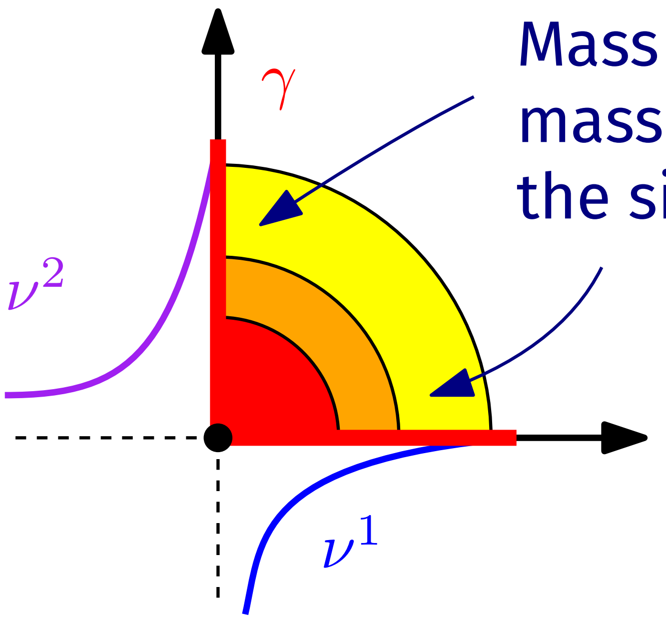
Guillen, Mou, Święch (2019). Coupling Lévy measures and comparison principles for viscosity solutions.

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

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

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



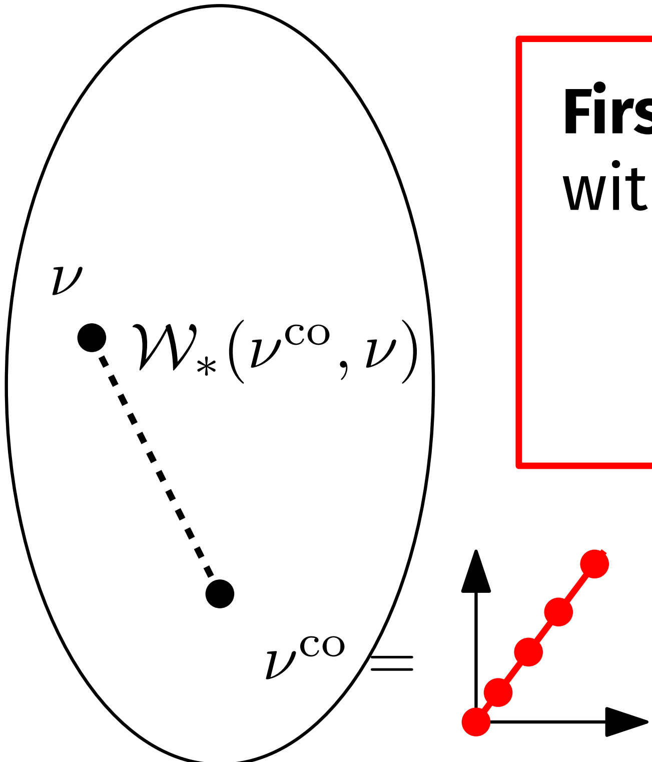
Mass on $\mathbb{R}^d_{+,*} \times \{0\}$ and $\{0\} \times \mathbb{R}^d_{+,*}$: mass "destroyed" or "created" from the sink/reservoir $(0,0)$.

In progress: extending this idea to couple also the law of atoms for inhomogeneous CRV

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.

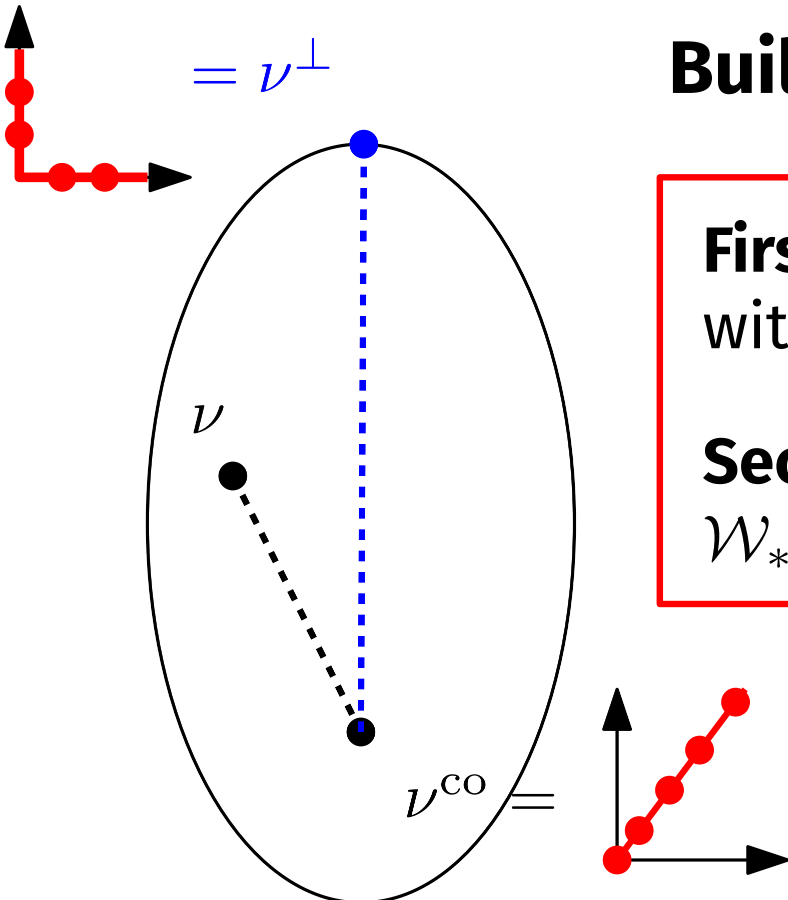
Building the index



Space of Lévy measure over $\mathbb{R}_+^d$ having same marginals

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

Catalano, Lavenant, Lijoi, Prünster (2022+). A Wasserstein index of dependence.

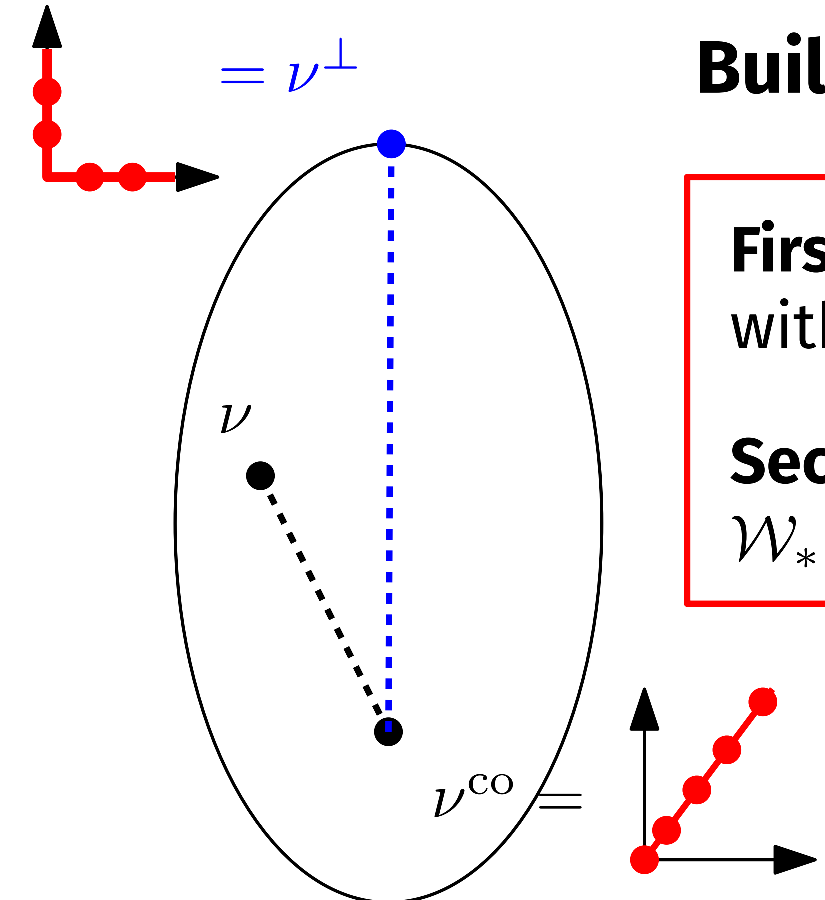


Space of Lévy measure over $\mathbb{R}_+^d$ having same marginals

Building the index

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

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



Space of Lévy measure over $\mathbb{R}_+^d$ having same marginals

Building the index

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

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

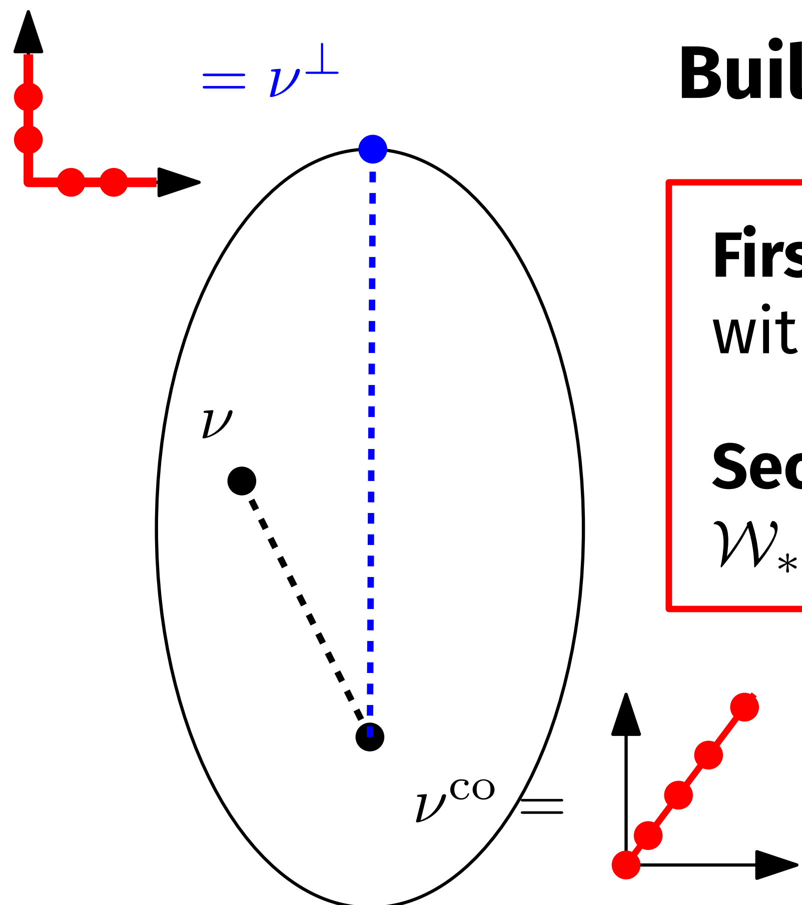
Define:

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

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

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

Catalano, Lavenant, Lijoi, Prünster (2022+). A Wasserstein index of dependence.



Space of Lévy measure over $\mathbb{R}_+^d$ having same marginals

Building the index

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

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

Define:

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

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 (2022+). A Wasserstein index of dependence.

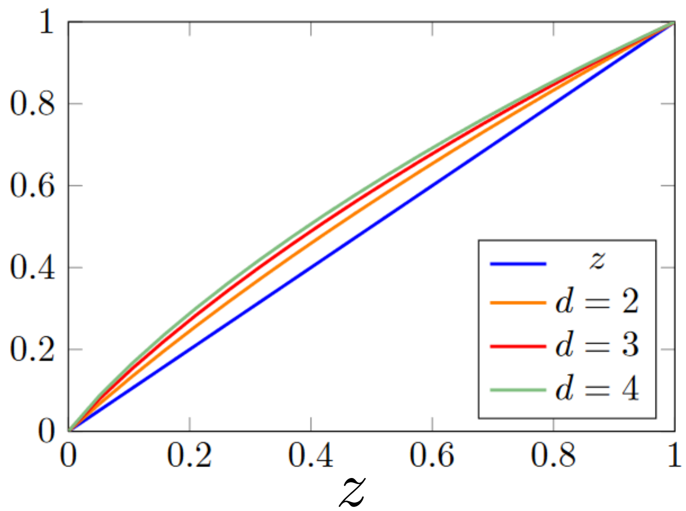
Examples

Additive model

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

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

$\mathcal{I}(z)$



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

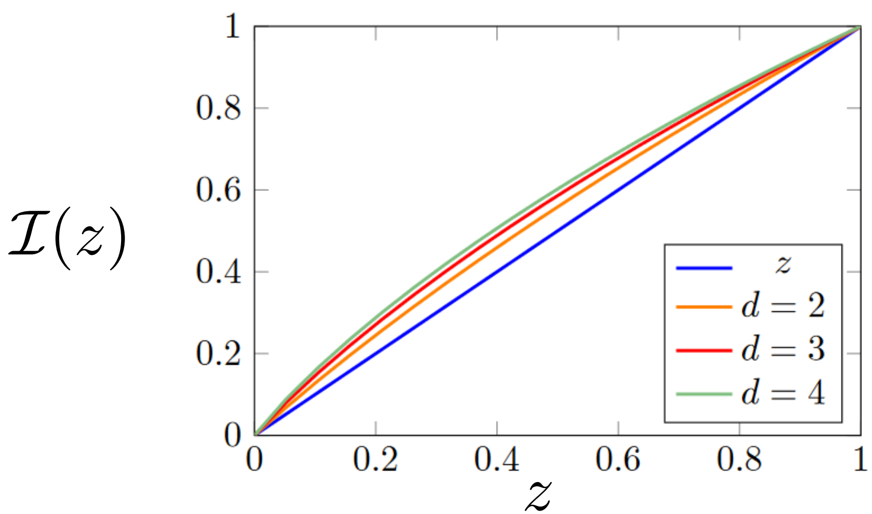
Lijoi, Nipoti and Prünster (2014). Bayesian inference with dependent normalized completely random measures.

Examples

Additive model

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

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



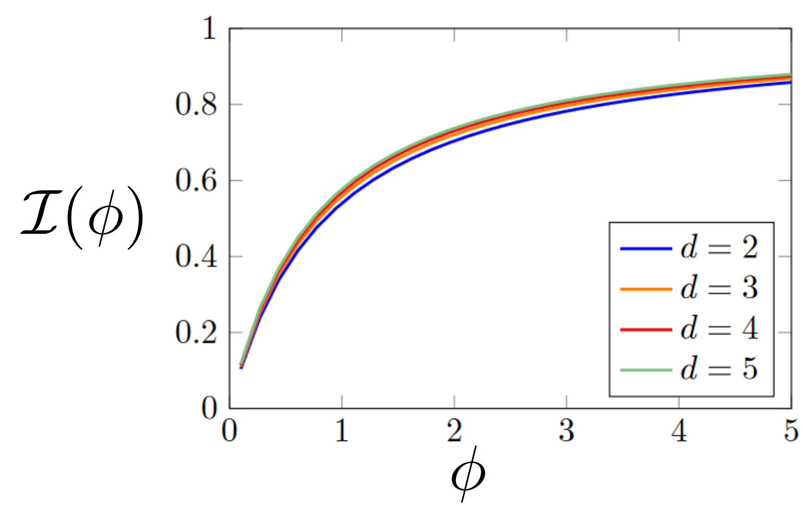
$\mathcal{I}(z) \geq z$ [$=$ Covariance if $d=2$]

Compound random measures

Parameter ϕ measures dependence

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

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.

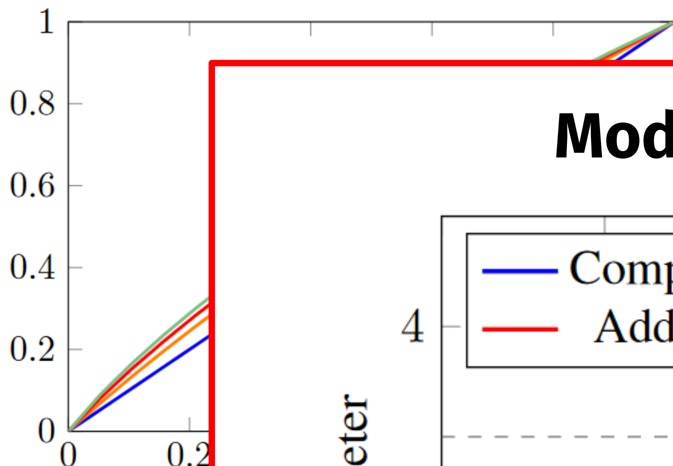
Examples

Additive model

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

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

$\mathcal{I}(z)$

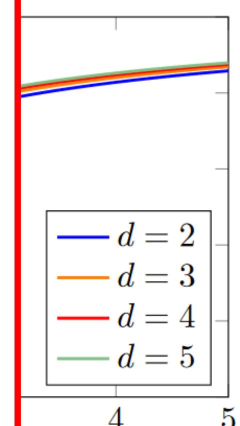


$\mathcal{I}(z) \geq z$ [=

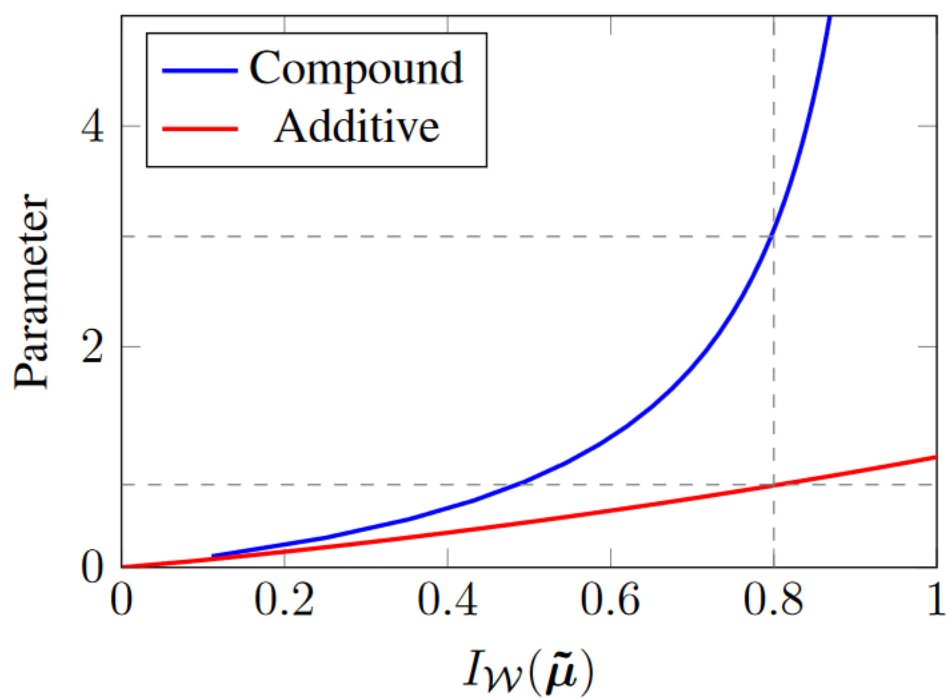
Compound random measures

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: convergence of posterior, distance of a prior/posterior to a reference one, etc.

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

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: convergence of posterior, distance of a prior/posterior to a reference one, etc.

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

Thank you for your attention