What distance to use between probabilities over probabilities?

Hugo Lavenant

Bocconi University

Optimal Transport Cargese Workshop

Cargèse (France), April 9, 2024

Joint work with

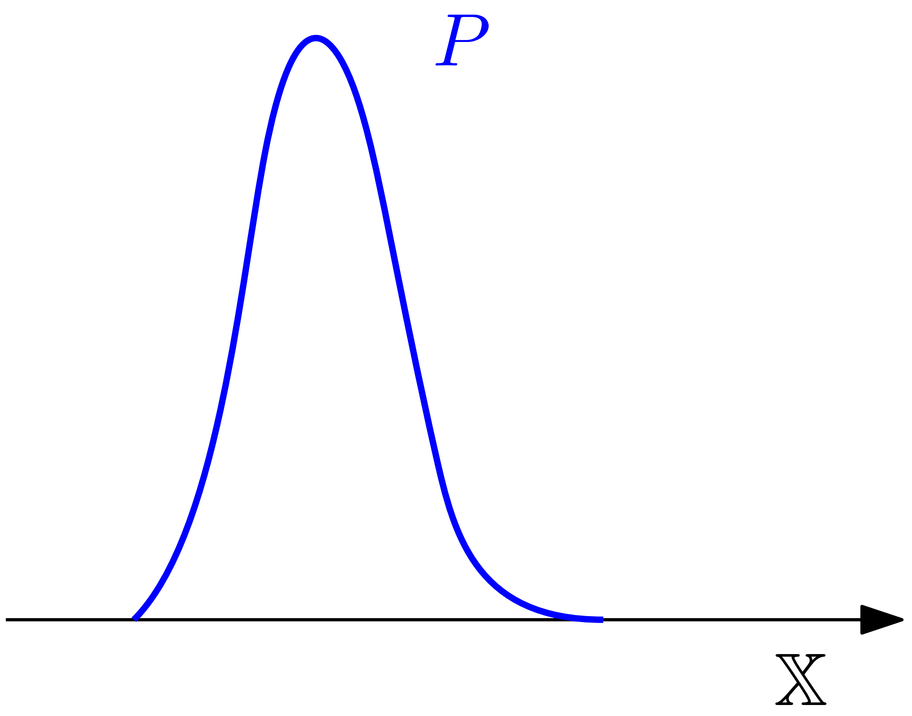
Marta Catalano (Luiss University)

Our article *Hierarchical Integral Probability Metrics: A distance on random probability measures with low sample complexity* is on arxiv!

Probabilities over Probabilities

$\mathbb{X}$ set (think subset of $\mathbb{R}^d$)

Probabilities over Probabilities

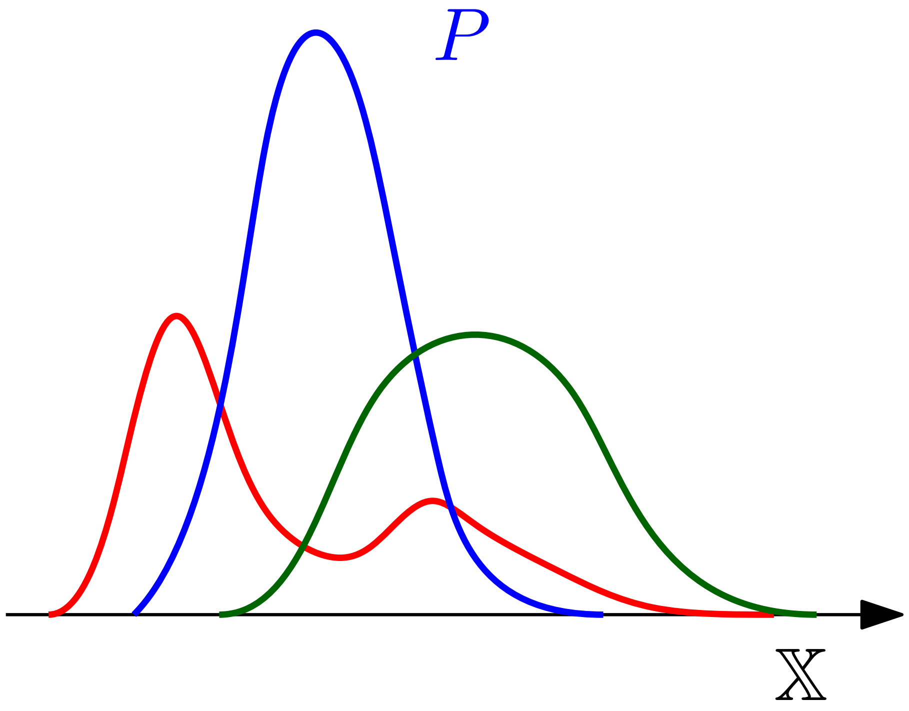


$\mathbb{X}$ set (think subset of $\mathbb{R}^d$)

$\mathcal{P}(\mathbb{X})$ probability distributions over $\mathbb{X}$

Typical element P

Probabilities over Probabilities



$\mathbb{X}$ set (think subset of $\mathbb{R}^d$)

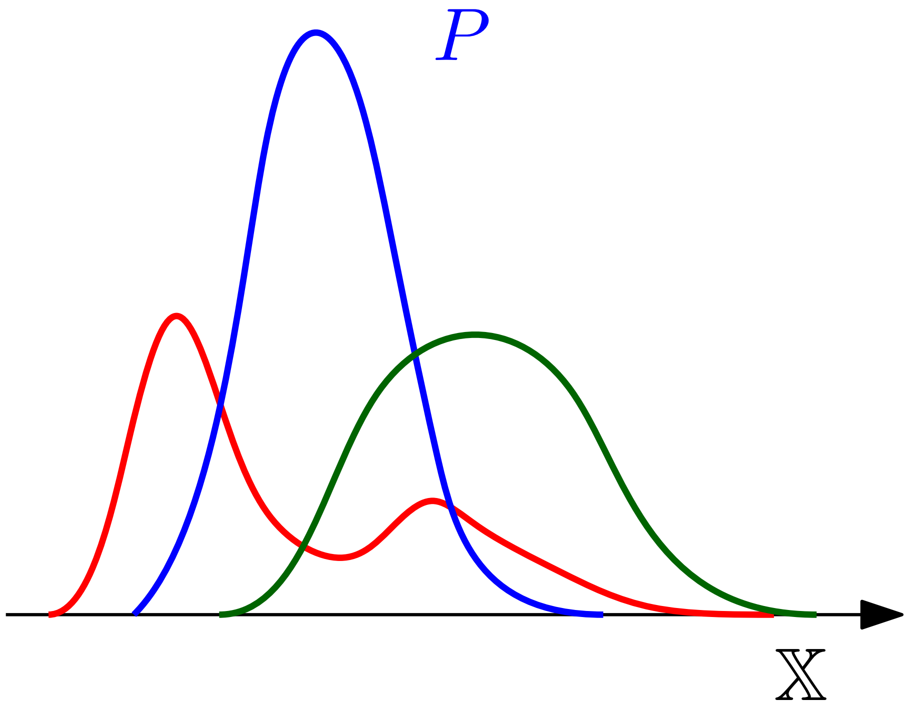
$\mathcal{P}(\mathbb{X})$ probability distributions over $\mathbb{X}$

Typical element P

$\mathcal{P}(\mathcal{P}(\mathbb{X}))$ probability distributions over $\mathcal{P}(\mathbb{X})$

Typical element $\mathbb{Q}$, or $\tilde{P} \sim \mathbb{Q}$ random probability

Probabilities over Probabilities



$\mathbb{X}$ set (think subset of $\mathbb{R}^d$)

$\mathcal{P}(\mathbb{X})$ probability distributions over $\mathbb{X}$

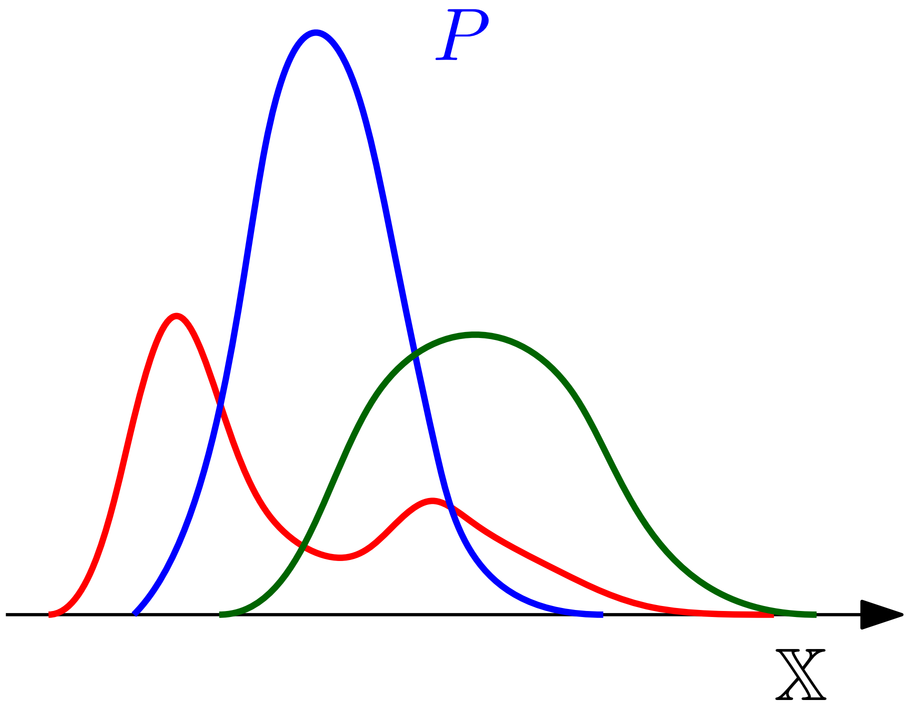
Typical element P

$\mathcal{P}(\mathcal{P}(\mathbb{X}))$ probability distributions over $\mathcal{P}(\mathbb{X})$

Typical element $\mathbb{Q}$, or $\tilde{P} \sim \mathbb{Q}$ random probability

What distance to put on the space $\mathcal{P}(\mathcal{P}(\mathbb{X}))$?

Probabilities over Probabilities



$\mathbb{X}$ set (think subset of $\mathbb{R}^d$)

$\mathcal{P}(\mathbb{X})$ probability distributions over $\mathbb{X}$

Typical element P

$\mathcal{P}(\mathcal{P}(\mathbb{X}))$ probability distributions over $\mathcal{P}(\mathbb{X})$

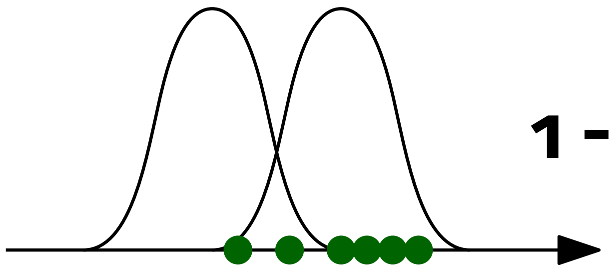
Typical element $\mathbb{Q}$, or $\tilde{P} \sim \mathbb{Q}$ random probability

$\boxed{\text{What distance to put on the space } \mathcal{P}(\mathcal{P}(\mathbb{X}))\text{?}}$

Desiderata:

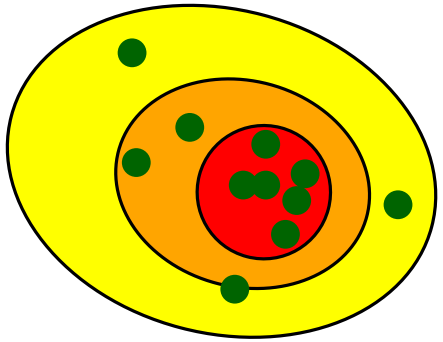
This talk

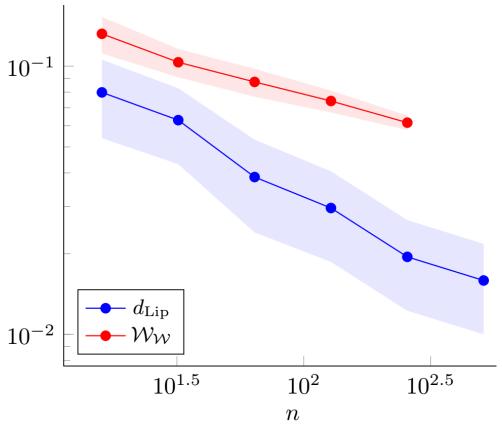
- Metrizing weak topology.
- Computation from samples: satistical and numerical complexity.
- Explicit formula, upper and lower bounds.



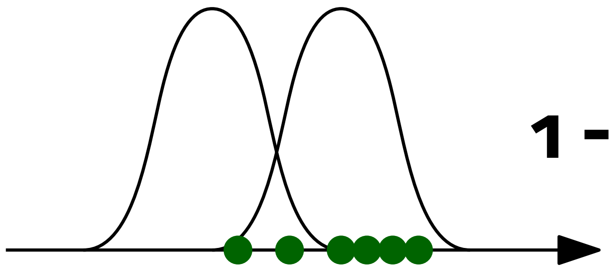
1 - Why? Bayesian Nonparametric Statistics

2 - Wasserstein over Wasserstein and its sample complexity



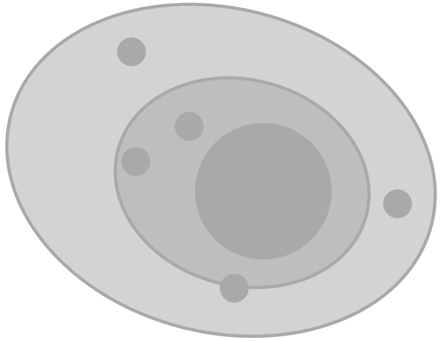


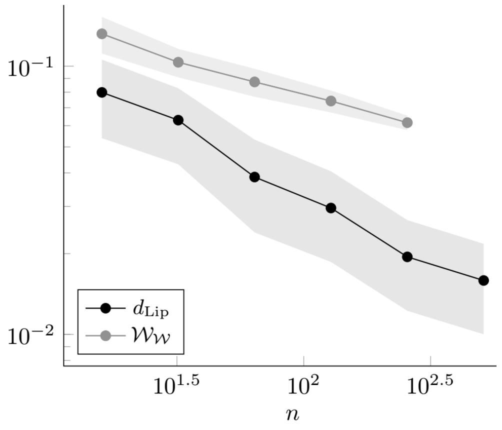
3 - A new distance with a better sample complexity



1 - Why? Bayesian Nonparametric Statistics

2 - Wasserstein over Wasserstein and its sample complexity



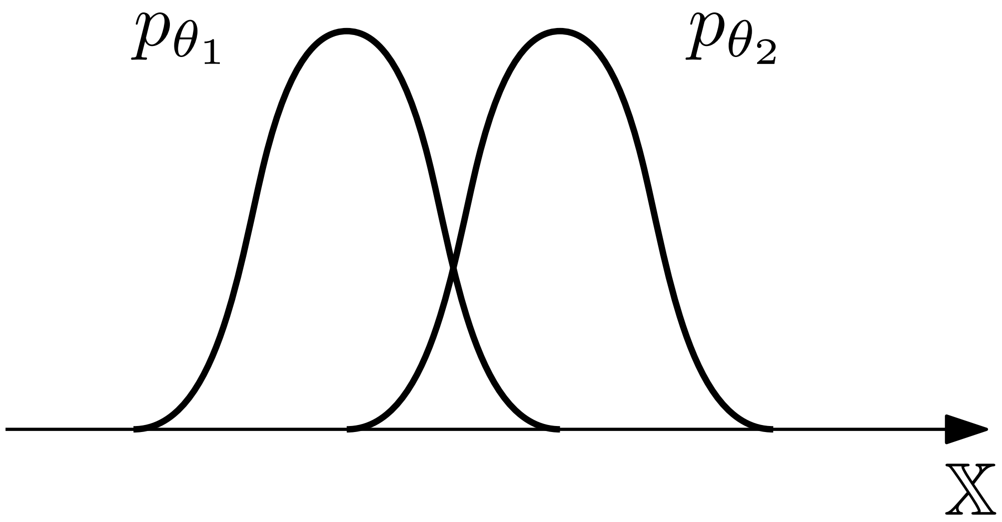


3 - A new distance with a better sample complexity

Bayesian (Nonparametric) Statistics

p_θ distributions over $\mathbb{X}$ indexed by $\theta \in \Theta$.

Goal: infer θ from data.



Bayesian (Nonparametric) Statistics

p_θ distributions over $\mathbb{X}$ indexed by $\theta \in \Theta$.

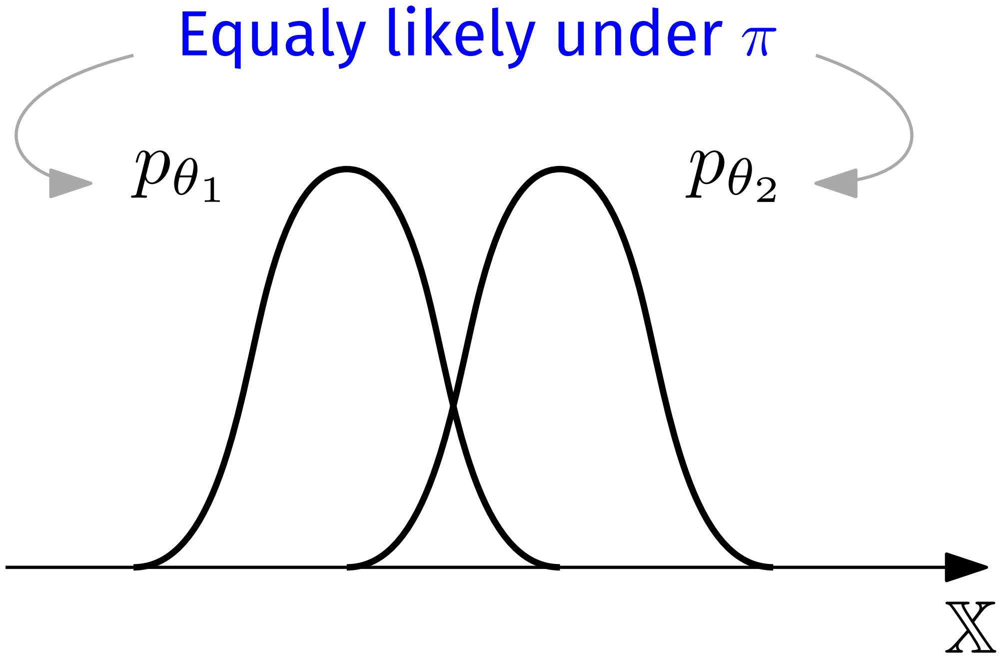
Goal: infer θ from data.

prior in $\mathcal{P}(\Theta)$

$$\theta \sim \pi$$

data in $\mathbb{X}$

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



Bayesian (Nonparametric) Statistics

p_θ distributions over $\mathbb{X}$ indexed by $\theta \in \Theta$.

Goal: infer θ from data.

prior in $\mathcal{P}(\Theta)$

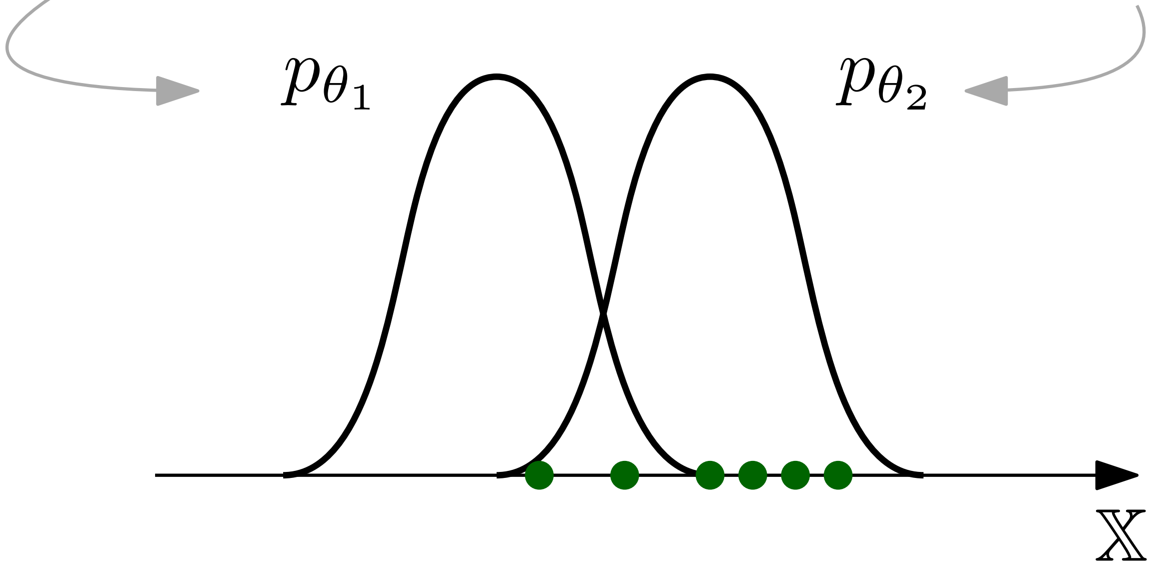
$$\theta \sim \pi$$

data in $\mathbb{X}$

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

Inference gives posterior $\theta | X_1, \ldots, X_n$.

$$\pi(\theta_1 | X_1, \ldots, X_n) \ll \pi(\theta_2 | X_1, \ldots, X_n)$$



Bayesian (Nonparametric) Statistics

p_θ distributions over $\mathbb{X}$ indexed by $\theta \in \Theta$.

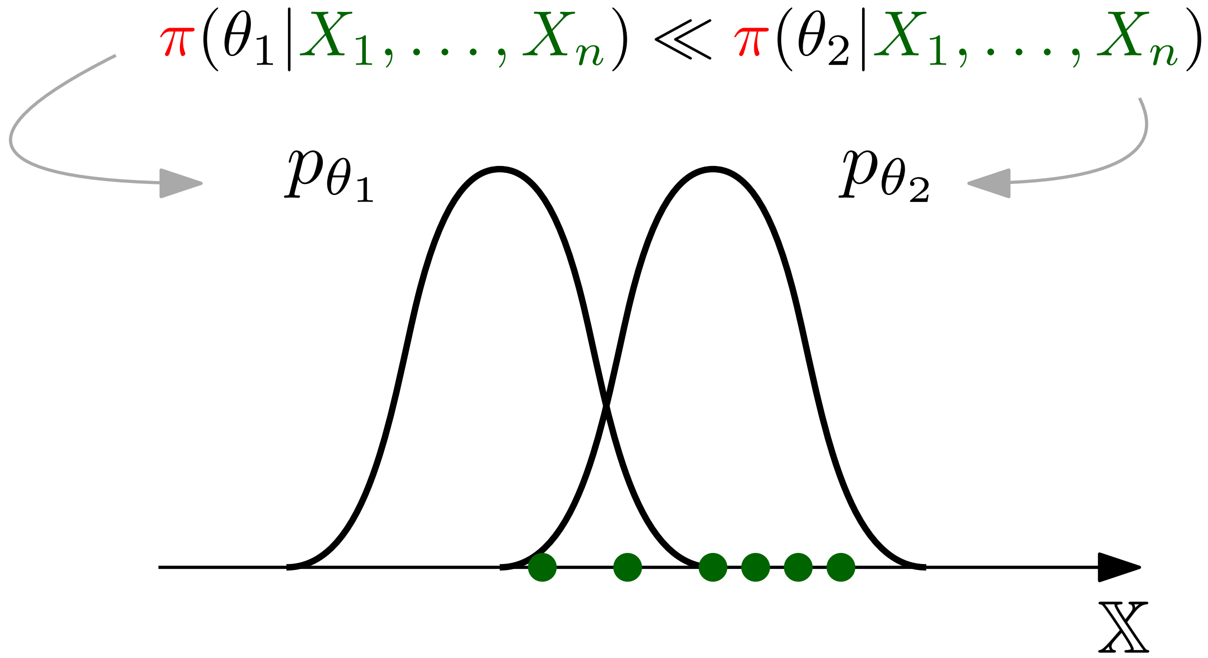
Goal: infer θ from data.

prior in $\mathcal{P}(\Theta)$

$$\theta \sim \pi$$

data in $\mathbb{X}$

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



Inference gives posterior $\theta | X_1, \ldots, X_n$.

Remark: p_θ with $\theta \sim \pi$ is a random probability: $\mathbb{Q} = (\theta \mapsto p_\theta) \# \pi$.

Bayesian NonParametrics: define directly $\mathbb{Q}$ (that is a random probability $\tilde{P}$) instead of p_θ and π.

Merging of opinions

Question. Different priors π^1, π^2 but same data $X_1, \ldots, X_n$.

Does the **distance** between the posteriors $\pi^1(\cdot|X_1, \ldots, X_n)$ and $\pi^2(\cdot|X_1, \ldots, X_n)$ converge to zero as $n \to +\infty$? At which rate in n?

Blackwell & Dubins (1962). Merging of opinions with increasing information.

Merging of opinions

Question. Different priors π^1, π^2 but same data $X_1, \ldots, X_n$.

Does the **distance** between the posteriors $\pi^1(\cdot|X_1, \ldots, X_n)$ and $\pi^2(\cdot|X_1, \ldots, X_n)$ converge to zero as $n \to +\infty$? At which rate in n?

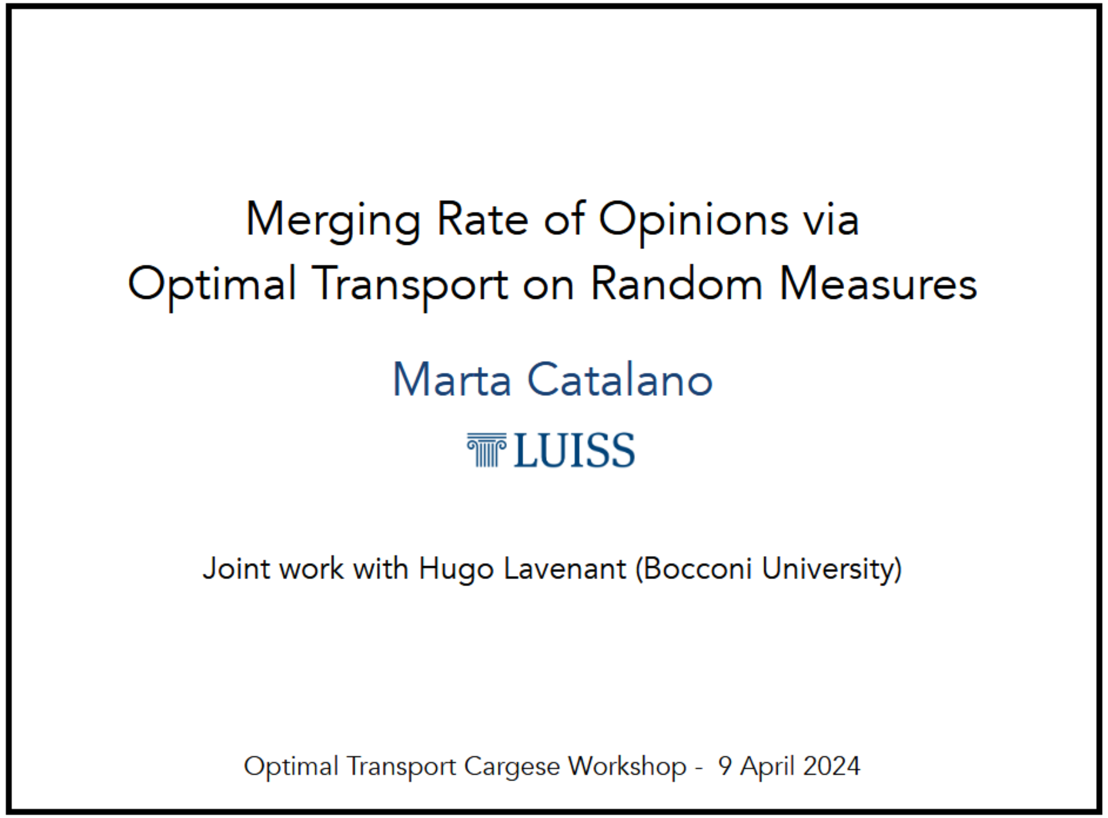
In Bayesian Nonparametrics, need for a distance between laws of random probabilities.

Blackwell & Dubins (1962). Merging of opinions with increasing information.
Catalano & Lavenant (2023). Merging Rate of Opinions via Optimal Transport on Random Measures.

Merging of opinions

Questi

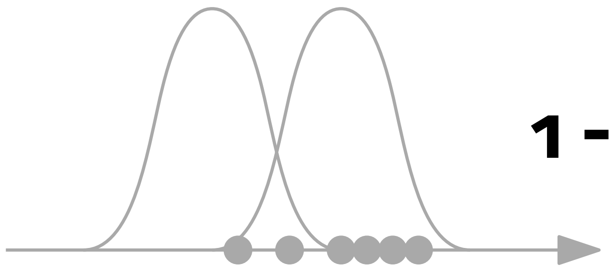
Does t ... and
$\pi^2(\cdot|X$ ... in n?



More about this with Marta in a few minutes!

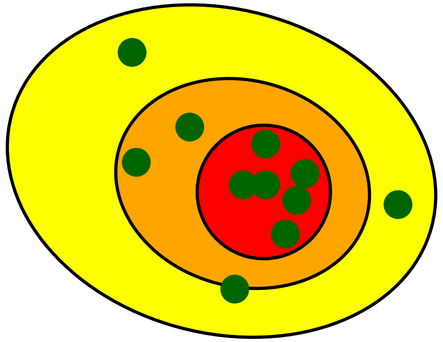
In Bayesian ... f random
probabilitie

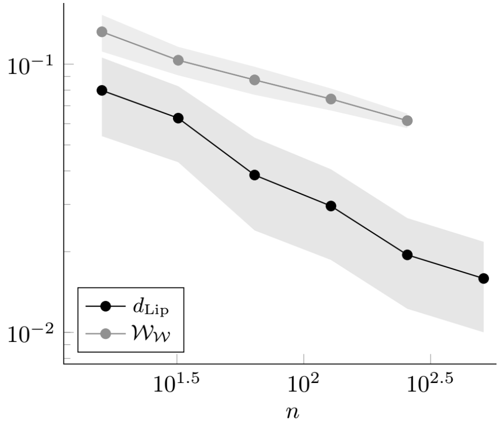
Blackwell & Dubins (1962). Merging of opinions with increasing information.
Catalano & Lavenant (2023). Merging Rate of Opinions via Optimal Transport on Random Measures.



1 - Why? Bayesian Nonparametric Statistics

2 - Wasserstein over Wasserstein and its sample complexity





3 - A new distance with a better sample complexity

Wasserstein over Wasserstein distance

$\mathbb{X}$ metric space, $\mathcal{W}$ Wasserstein distance of order 1 on $\mathcal{P}(\mathbb{X})$.

Definition. If $\mathbb{Q}_1, \mathbb{Q}_2 \in \mathcal{P}(\mathcal{P}(\mathbb{X}))$, the "Wasserstein over Wasserstein" distance is:

$$\mathcal{W}_{\mathcal{W}}(\mathbb{Q}_1, \mathbb{Q}_2) = \inf_{\gamma \in \Gamma(\mathbb{Q}_1, \mathbb{Q}_2)} \mathbb{E}_{(\tilde{P}_1, \tilde{P}_2) \sim \gamma} \left[\mathcal{W}(\tilde{P}_1, \tilde{P}_2) \right].$$

Couplings between $\mathbb{Q}_1$ and $\mathbb{Q}_2$

Nguyen (2016). Borrowing strengh in hierarchical Bayes: Posterior concentration of the Dirichlet base measure.
Yurochkin et al (2019) Hierarchical optimal transport for document representation.
Bing et al (2016).The sketched Wasserstein distance for mixture distributions.

Wasserstein over Wasserstein distance

$\mathbb{X}$ metric space, $\mathcal{W}$ Wasserstein distance of order 1 on $\mathcal{P}(\mathbb{X})$.

Definition. If $\mathbb{Q}_1, \mathbb{Q}_2 \in \mathcal{P}(\mathcal{P}(\mathbb{X}))$, the "Wasserstein over Wasserstein" distance is:

$$\mathcal{W}_{\mathcal{W}}(\mathbb{Q}_1, \mathbb{Q}_2) = \inf_{\gamma \in \Gamma(\mathbb{Q}_1, \mathbb{Q}_2)} \mathbb{E}_{(\tilde{P}_1, \tilde{P}_2) \sim \gamma} \left[\mathcal{W}(\tilde{P}_1, \tilde{P}_2) \right].$$

Couplings between $\mathbb{Q}_1$ and $\mathbb{Q}_2$

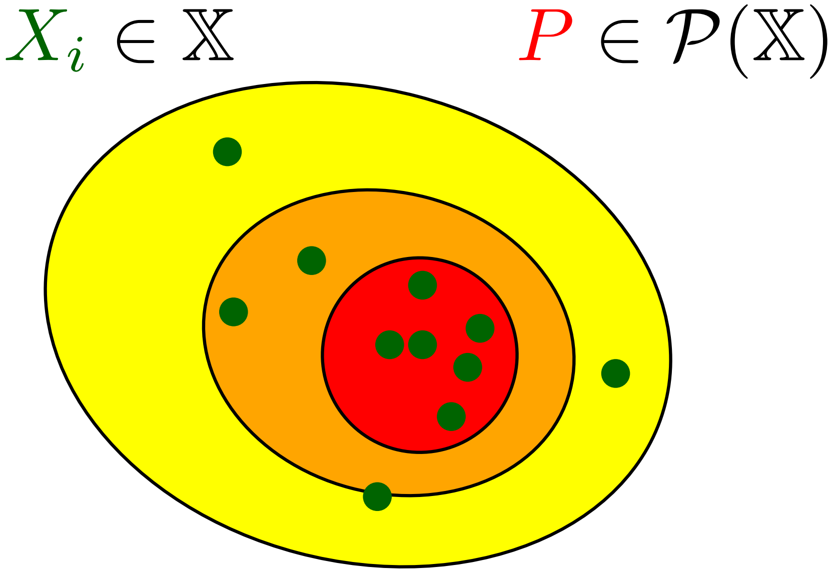
Weak convergence over weak convergence

Theorem. If $\mathbb{X}$ is bounded, then $\mathcal{W}_{\mathcal{W}}$ metrizes the weak convergence over $\mathcal{P}(\mathcal{P}(\mathbb{X}))$.

Nguyen (2016). Borrowing strengh in hierarchical Bayes: Posterior concentration of the Dirichlet base measure.
Yurochkin et al (2019) Hierarchical optimal transport for document representation.
Bing et al (2016).The sketched Wasserstein distance for mixture distributions.

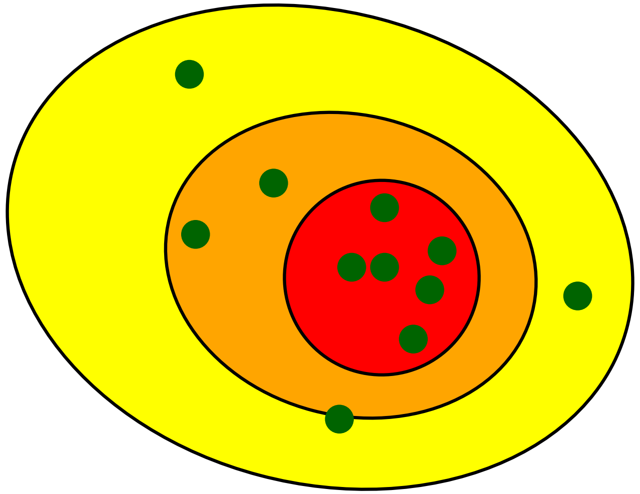
Sample complexity: reminder

- $P \in \mathcal{P}(\mathbb{X})$.
- $X_1, \dots X_n \overset{\text{i.i.d.}}{\sim} P$, build $\tilde{P}_{(n)} = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}$.



How close is $\tilde{P}_{(n)}$ from P?

Sample complexity: reminder



- $P \in \mathcal{P}(\mathbb{X})$.
- $X_1, \ldots X_n \overset{\text{i.i.d.}}{\sim} P$, build $\tilde{P}_{(n)} = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}$.

How close is $\tilde{P}_{(n)}$ from P?

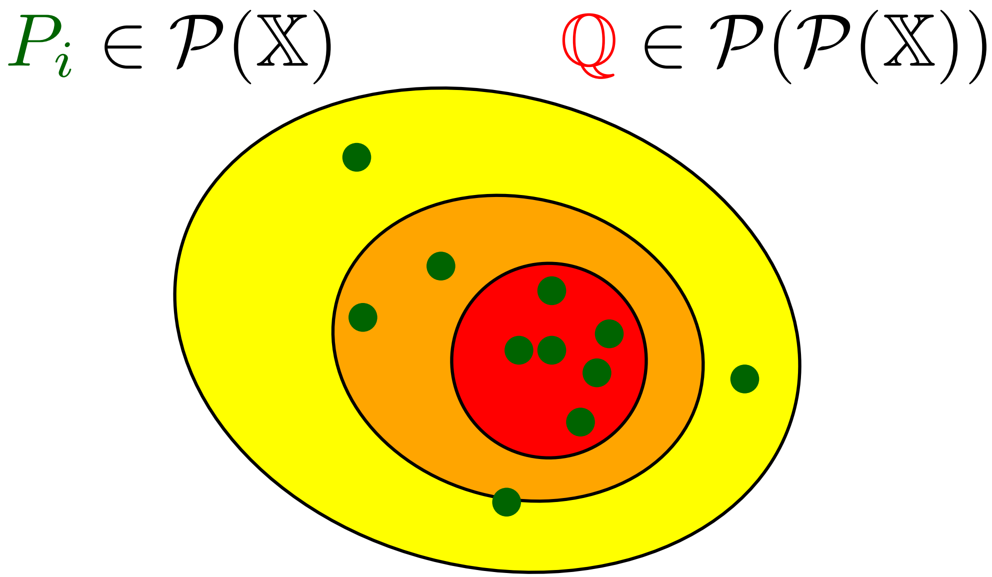
Theorem. If P is "d-dimensional", then:

$$\mathbb{E}\left[\mathcal{W}(\tilde{P}_{(n)}, P)\right] \asymp \begin{cases} n^{-1/2} & \text{if } d = 1, \\ n^{-1/2}\log(n) & \text{if } d = 2, \\ n^{-1/d} & \text{if } d \geq 3. \end{cases}$$

Dudley (1969). The speed of mean Glivenko-Cantelli Convergence.
Weed & Bach (2019). Sharp asymptotic and finite-sample rates of convergence of empirical measures in Wasserstein distance.

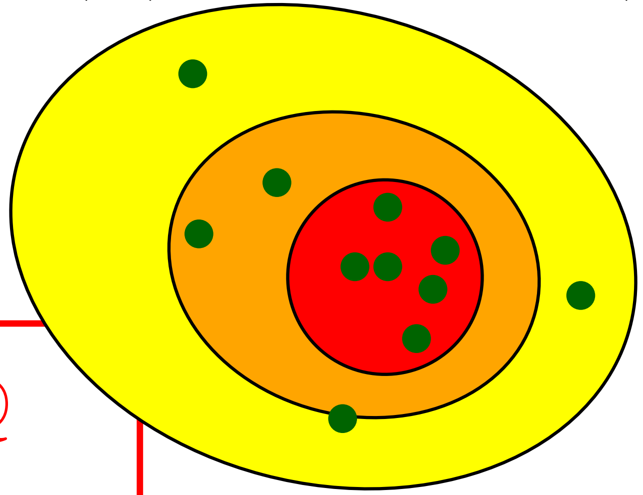
Sample complexity for Wasserstein over Wasserstein

- $\mathbb{Q} \in \mathcal{P}(\mathcal{P}(\mathbb{X}))$.
- $P_1, \ldots P_n \overset{\text{i.i.d.}}{\sim} \mathbb{Q}$, build $\tilde{\mathbb{Q}}_{(n)} = \frac{1}{n} \sum_{i=1}^{n} \delta_{P_i}$.



Sample complexity for Wasserstein over Wasserstein

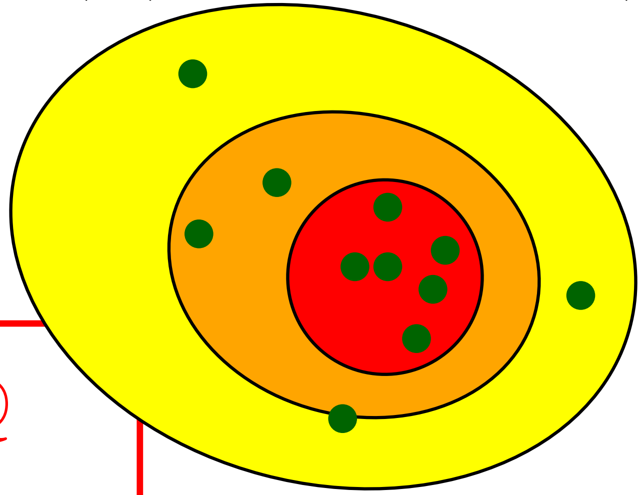
- $\mathbb{Q} \in \mathcal{P}(\mathcal{P}(\mathbb{X}))$.
- $P_1, \ldots P_n \overset{\text{i.i.d.}}{\sim} \mathbb{Q}$, build $\tilde{\mathbb{Q}}_{(n)} = \frac{1}{n}\sum_{i=1}^{n} \delta_{P_i}$.



Theorem. Take $\mathbb{X} \subset \mathbb{R}^d$ bounded. Then for any $\mathbb{Q}$

$$\mathbb{E}\left[\mathcal{W}_{\mathcal{W}}\left(\tilde{\mathbb{Q}}_{(n)}, \mathbb{Q}\right)\right] \leq C_{\mathbb{X}} \frac{\log(\log(n))}{\log(n)},$$

Sample complexity for Wasserstein over Wasserstein



- $\mathbb{Q} \in \mathcal{P}(\mathcal{P}(\mathbb{X}))$.
- $P_1, \ldots P_n \overset{\text{i.i.d.}}{\sim} \mathbb{Q}$, build $\tilde{\mathbb{Q}}_{(n)} = \frac{1}{n}\sum_{i=1}^{n} \delta_{P_i}$.

Theorem. Take $\mathbb{X} \subset \mathbb{R}^d$ bounded. Then for any $\mathbb{Q}$

$$\mathbb{E}\left[\mathcal{W}_{\mathcal{W}}\left(\tilde{\mathbb{Q}}_{(n)}, \mathbb{Q}\right)\right] \leq C_{\mathbb{X}} \frac{\log(\log(n))}{\log(n)},$$

and taking for $\mathbb{Q}$ a **Dirichlet process**, for any $\gamma > 0$,

$$\mathbb{E}\left[\mathcal{W}_{\mathcal{W}}\left(\tilde{\mathbb{Q}}_{(n)}, \mathbb{Q}\right)\right] \geq \frac{c_\gamma}{n^\gamma}.$$

What is this Dirichlet process giving a lower bound?

Parameters: base measure $P_0 \in \mathcal{P}(\mathbb{X})$ and concentration parameter $\alpha > 0$.

To draw $\tilde{P}$ according to a Dirichlet process:

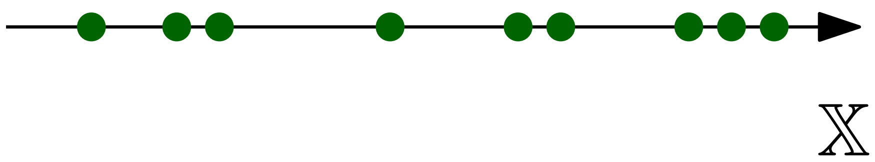
Sethuraman (1994). A constructive definition of Dirichlet priors.

What is this Dirichlet process giving a lower bound?

Parameters: base measure $P_0 \in \mathcal{P}(\mathbb{X})$ and concentration parameter $\alpha > 0$.

To draw $\tilde{P}$ according to a Dirichlet process:

1. Draw $X_1, \ldots, X_n, \ldots \overset{\text{i.i.d.}}{\sim} P_0$.



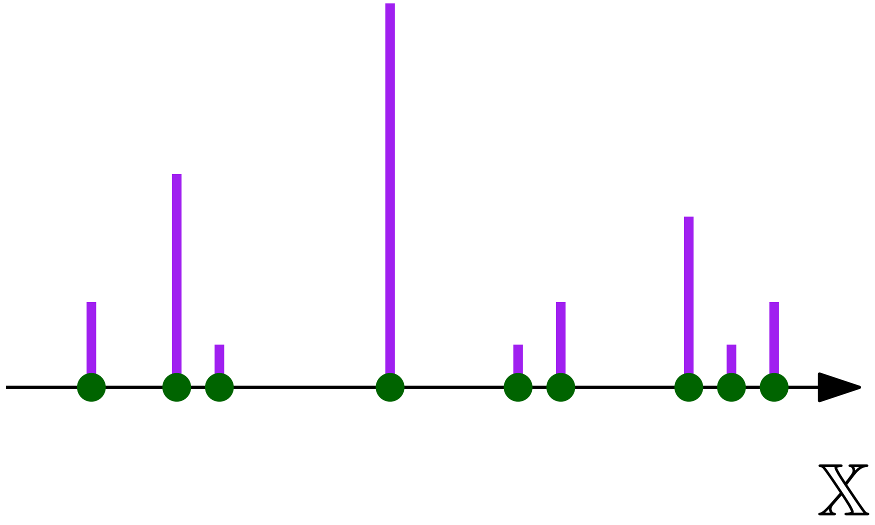
Sethuraman (1994). A constructive definition of Dirichlet priors.

What is this Dirichlet process giving a lower bound?

Parameters: base measure $P_0 \in \mathcal{P}(\mathbb{X})$ and concentration parameter $\alpha > 0$.

To draw $\tilde{P}$ according to a Dirichlet process:

1. Draw $X_1, \dots, X_n, \dots \overset{\text{i.i.d.}}{\sim} P_0$.
2. Draw independently weights $J_1, \dots, J_n, \dots$ which sum to 1 (law depending on α).
3. Define $\tilde{P} = \sum_{n=1}^{+\infty} J_n \delta_{X_n}$.



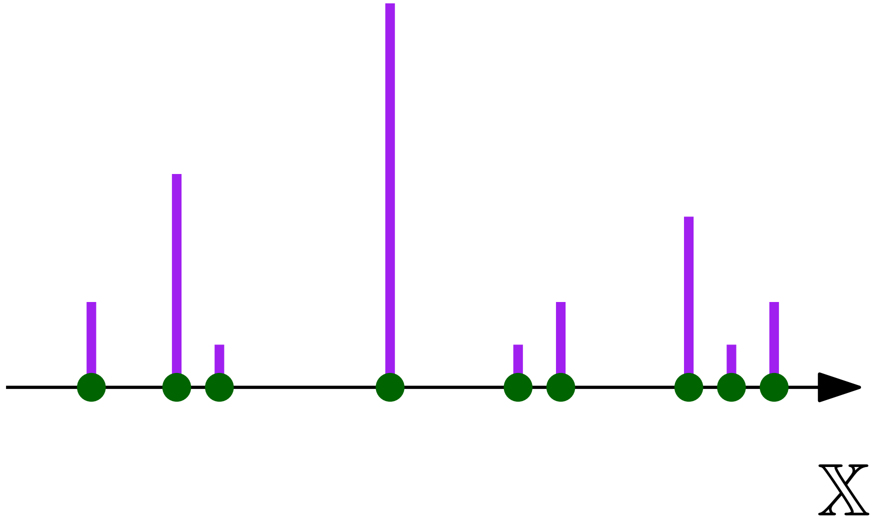
Sethuraman (1994). A constructive definition of Dirichlet priors.

What is this Dirichlet process giving a lower bound?

Parameters: base measure $P_0 \in \mathcal{P}(\mathbb{X})$ and concentration parameter $\alpha > 0$.

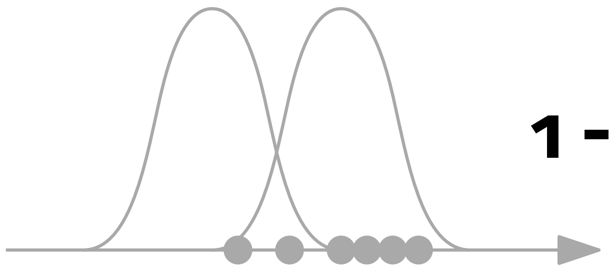
To draw $\tilde{P}$ according to a Dirichlet process:

1. Draw $X_1, \ldots, X_n, \ldots \overset{\text{i.i.d.}}{\sim} P_0$.
2. Draw independently weights $J_1, \ldots, J_n, \ldots$ which sum to 1 (law depending on α).
3. Define $\tilde{P} = \sum_{n=1}^{+\infty} J_n \delta_{X_n}$.



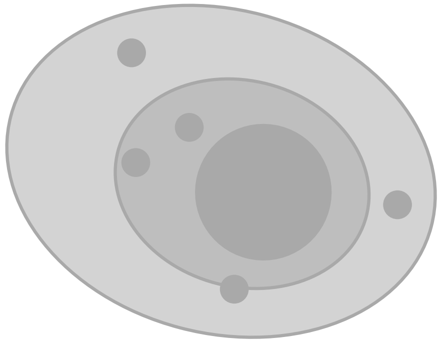
Remark. If the support of P_0 is $\mathbb{X}$, the topological support of the Dirichlet process is $\mathcal{P}(\mathbb{X})$.

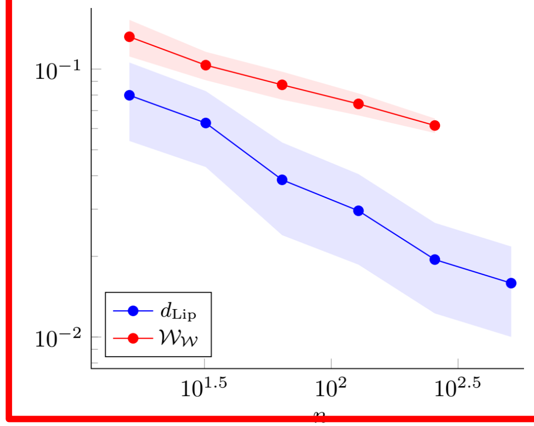
Sethuraman (1994). A constructive definition of Dirichlet priors.



1 - Why? Bayesian Nonparametric Statistics

2 - Wasserstein over Wasserstein and its sample complexity





3 - A new distance with a better sample complexity

A new distance

$\mathbb{Q}_1, \mathbb{Q}_2 \in \mathcal{P}(\mathcal{P}(\mathbb{X}))$, recall:

$$\mathcal{W}_{\mathcal{W}}(\mathbb{Q}_1, \mathbb{Q}_2) = \inf_{\gamma \in \Gamma(\mathbb{Q}_1, \mathbb{Q}_2)} \sup_{f \in \mathrm{Lip}_1(\mathbb{X})} \mathbb{E}_{(\tilde{P}_1, \tilde{P}_2) \sim \gamma} \left[\left| \int_{\mathbb{X}} f \, \mathrm{d}\tilde{P}_1 - \int_{\mathbb{X}} f \, \mathrm{d}\tilde{P}_2 \right| \right].$$

A new distance

$\mathbb{Q}_1, \mathbb{Q}_2 \in \mathcal{P}(\mathcal{P}(\mathbb{X}))$, recall:

$$\mathcal{W}_{\mathcal{W}}(\mathbb{Q}_1, \mathbb{Q}_2) = \inf_{\gamma \in \Gamma(\mathbb{Q}_1, \mathbb{Q}_2)} \sup_{f \in \mathrm{Lip}_1(\mathbb{X})} \mathbb{E}_{(\tilde{P}_1, \tilde{P}_2) \sim \gamma} \left[\left| \int_{\mathbb{X}} f \, \mathrm{d}\tilde{P}_1 - \int_{\mathbb{X}} f \, \mathrm{d}\tilde{P}_2 \right| \right].$$

Definition.

$$d_{\mathrm{Lip}}(\mathbb{Q}_1, \mathbb{Q}_2) = \sup_{f \in \mathrm{Lip}_1(\mathbb{X})} \inf_{\gamma \in \Gamma(\mathbb{Q}_1, \mathbb{Q}_2)} \mathbb{E}_{(\tilde{P}_1, \tilde{P}_2) \sim \gamma} \left[\left| \int_{\mathbb{X}} f \, \mathrm{d}\tilde{P}_1 - \int_{\mathbb{X}} f \, \mathrm{d}\tilde{P}_2 \right| \right]$$

A new distance

$\mathbb{Q}_1, \mathbb{Q}_2 \in \mathcal{P}(\mathcal{P}(\mathbb{X}))$, recall:

$$\mathcal{W}_{\mathcal{W}}(\mathbb{Q}_1, \mathbb{Q}_2) = \inf_{\gamma \in \Gamma(\mathbb{Q}_1, \mathbb{Q}_2)} \sup_{f \in \mathrm{Lip}_1(\mathbb{X})} \mathbb{E}_{(\tilde{P}_1, \tilde{P}_2) \sim \gamma} \left[\left| \int_{\mathbb{X}} f \, \mathrm{d}\tilde{P}_1 - \int_{\mathbb{X}} f \, \mathrm{d}\tilde{P}_2 \right| \right].$$

Definition.

$$d_{\mathrm{Lip}}(\mathbb{Q}_1, \mathbb{Q}_2) = \sup_{f \in \mathrm{Lip}_1(\mathbb{X})} \inf_{\gamma \in \Gamma(\mathbb{Q}_1, \mathbb{Q}_2)} \mathbb{E}_{(\tilde{P}_1, \tilde{P}_2) \sim \gamma} \left[\left| \int_{\mathbb{X}} f \, \mathrm{d}\tilde{P}_1 - \int_{\mathbb{X}} f \, \mathrm{d}\tilde{P}_2 \right| \right]$$

$$= \sup_{f \in \mathrm{Lip}_1(\mathbb{X})} \mathcal{W}\left(\int_{\mathbb{X}} f \, \mathrm{d}\tilde{P}_1, \int_{\mathbb{X}} f \, \mathrm{d}\tilde{P}_2 \right) \qquad \tilde{P}_1 \sim \mathbb{Q}_1, \; \tilde{P}_2 \sim \mathbb{Q}_2.$$

Idea. Project $\mathcal{P}(\mathbb{X})$ on $\mathbb{R}$ via $P \mapsto \int f \, \mathrm{d}P$ for $f \in \mathrm{Lip}_1(\mathbb{X})$, then measure Wasserstein distance of projections.

A new distance

Remark. Replace $\mathrm{Lip}_1(\mathbb{X})$ by $\mathcal{F}$ class of function $f : \mathbb{X} \to \mathbb{R}$ generating an *Integral Probability Metric*. We call the distance **Hierarchical IPM**.

Definition.

$$d_{\mathrm{Lip}}(\mathbb{Q}_1, \mathbb{Q}_2) = \sup_{f \in \mathrm{Lip}_1(\mathbb{X})} \inf_{\gamma \in \Gamma(\mathbb{Q}_1, \mathbb{Q}_2)} \mathbb{E}_{(\tilde{P}_1, \tilde{P}_2) \sim \gamma} \left[\left| \int_{\mathbb{X}} f \, \mathrm{d}\tilde{P}_1 - \int_{\mathbb{X}} f \, \mathrm{d}\tilde{P}_2 \right| \right]$$

$$= \sup_{f \in \mathrm{Lip}_1(\mathbb{X})} \mathcal{W} \left(\int_{\mathbb{X}} f \, \mathrm{d}\tilde{P}_1, \int_{\mathbb{X}} f \, \mathrm{d}\tilde{P}_2 \right) \qquad \tilde{P}_1 \sim \mathbb{Q}_1, \; \tilde{P}_2 \sim \mathbb{Q}_2.$$

Idea. Project $\mathcal{P}(\mathbb{X})$ on $\mathbb{R}$ via $P \mapsto \int f \, \mathrm{d}P$ for $f \in \mathrm{Lip}_1(\mathbb{X})$, then measure Wasserstein distance of projections.

Properties of this new distance

Theorem. There holds $d_{\text{Lip}} \leq \mathcal{W}_{\mathcal{W}}$.
If $\mathbb{X}$ compact, d_{Lip} is a distance metrizing weak convergence over $\mathcal{P}(\mathcal{P}(\mathbb{X}))$.

Properties of this new distance

Theorem. There holds $d_{\mathrm{Lip}} \leq \mathcal{W}_{\mathcal{W}}$.
If $\mathbb{X}$ compact, d_{Lip} is a distance metrizing weak convergence over $\mathcal{P}(\mathcal{P}(\mathbb{X}))$.

Theorem (sample complexity).

- $\mathbb{Q} \in \mathcal{P}(\mathcal{P}(\mathbb{X}))$ with $\mathbb{X}$ bounded subset of $\mathbb{R}^d$.
- $P_1, \ldots P_n \overset{\text{i.i.d.}}{\sim} \mathbb{Q}$, build $\tilde{\mathbb{Q}}_{(n)} = \dfrac{1}{n} \sum_{i=1}^{n} \delta_{P_i}$.

Properties of this new distance

Theorem. There holds $d_{\mathrm{Lip}} \leq \mathcal{W}_{\mathcal{W}}$.
If $\mathbb{X}$ compact, d_{Lip} is a distance metrizing weak convergence over $\mathcal{P}(\mathcal{P}(\mathbb{X}))$.

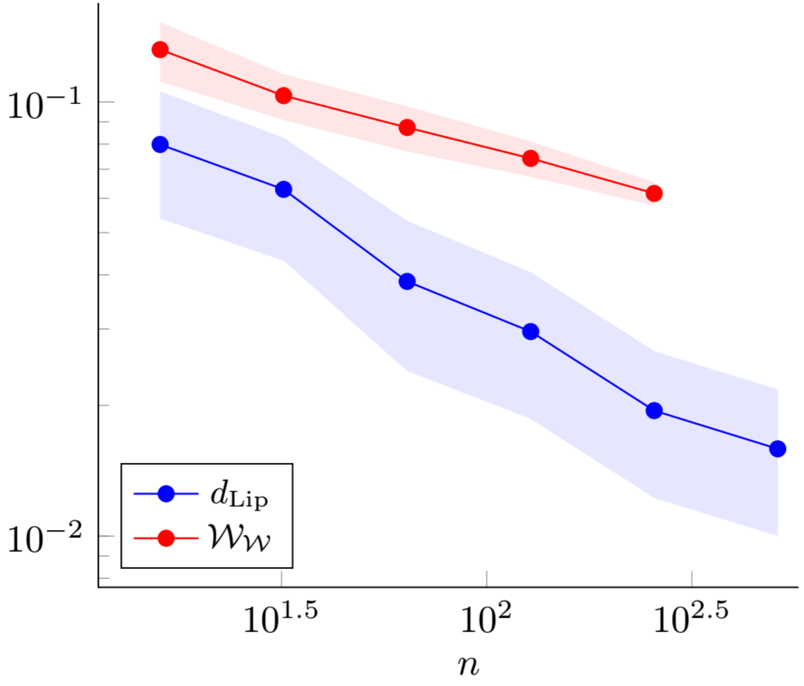
Theorem (sample complexity).

- $\mathbb{Q} \in \mathcal{P}(\mathcal{P}(\mathbb{X}))$ with $\mathbb{X}$ bounded subset of $\mathbb{R}^d$.
- $P_1, \ldots P_n \overset{\text{i.i.d.}}{\sim} \mathbb{Q}$, build $\tilde{\mathbb{Q}}_{(n)} = \frac{1}{n}\sum_{i=1}^{n} \delta_{P_i}$.

Then

$$\mathbb{E}\left[d_{\mathrm{Lip}}\left(\tilde{\mathbb{Q}}_{(n)}, \mathbb{Q}\right)\right] \lesssim \begin{cases} n^{-1/2} & \text{if } d = 1, \\ n^{-1/2}\log(n) & \text{if } d = 2, \\ n^{-1/d} & \text{if } d \geq 3. \end{cases}$$

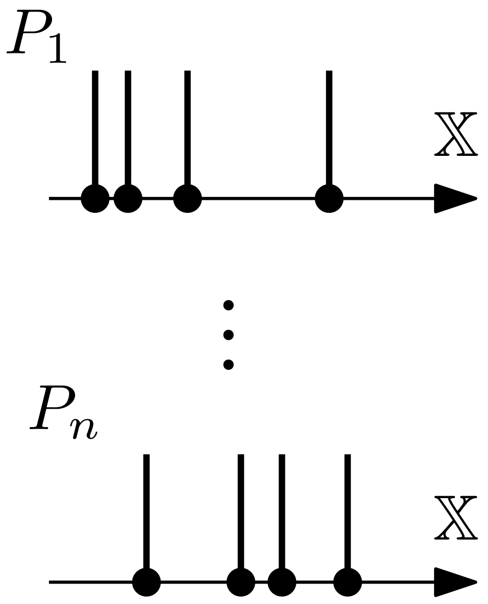
$d(\tilde{\mathbb{Q}}_{(n)}, \tilde{\mathbb{Q}}'_{(n)})$



Rate for *classical* Wasserstein distance in $\mathbb{R}^d$, $\ll \log(\log(n))/\log(n)$.

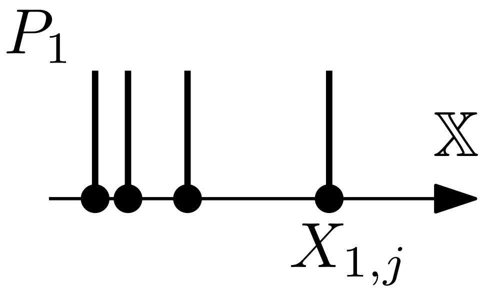
A word on Numerics

$$\mathbb{Q} = \frac{1}{n}\sum_{i=1}^{n}\delta_{P_i} \text{ discrete}$$



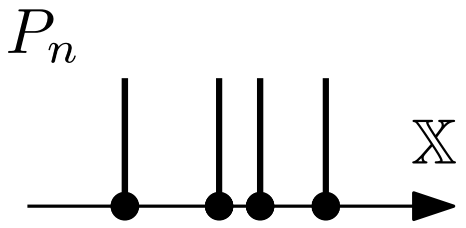
A word on Numerics

$$\mathbb{Q} = \frac{1}{n}\sum_{i=1}^{n} \delta_{P_i} \text{ discrete}$$



$$\rightsquigarrow \quad P_1 = \frac{1}{m}\sum_{j=1}^{m} \delta_{X_{1,j}} \text{ discrete}$$

$$\vdots \qquad\qquad \vdots$$

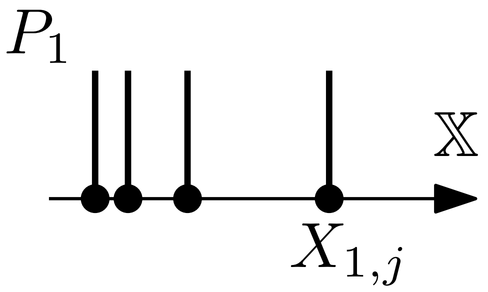


$$\rightsquigarrow \quad P_n = \frac{1}{m}\sum_{j=1}^{m} \delta_{X_{n,j}} \text{ discrete}$$

Each element of $\mathcal{P}(\mathcal{P}(\mathbb{X}))$ is stored as a $n \times m$ array of atoms (and weights).

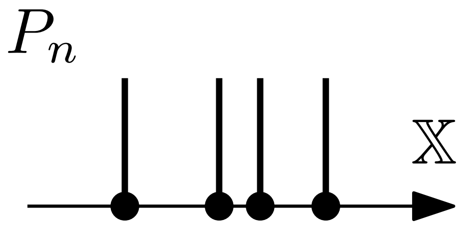
A word on Numerics

$\mathbb{Q} = \frac{1}{n}\sum_{i=1}^{n}\delta_{P_i}$ discrete



$\rightsquigarrow \quad P_1 = \frac{1}{m}\sum_{j=1}^{m}\delta_{X_{1,j}}$ discrete

$\vdots$



$\rightsquigarrow \quad P_n = \frac{1}{m}\sum_{j=1}^{m}\delta_{X_{n,j}}$ discrete

Each element of $\mathcal{P}(\mathcal{P}(\mathbb{X}))$ is stored as a $n \times m$ array of atoms (and weights).

Computing d_{Lip} is finding the supremum of $f \mapsto \mathcal{W}(\int f \,\mathrm{d}\tilde{P}_1, \int f \,\mathrm{d}\tilde{P}_2)$ among $\mathrm{Lip}_1(\mathbb{X})$.

Non convex, non concave. We propose a gradient ascent when $\mathbb{X} \subset \mathbb{R}$.

Thank you for your attention