Quantifying the merging of opinions in Bayesian nonparametrics via optimal transport

Hugo Lavenant

Bocconi University

CIRM Conference “Approximation Methods in Bayesian Analysis”,
Marseille (France), June 22, 2023

Joint work with:

Marta Catalano

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

Joint work with:

WARWICK
THE UNIVERSITY OF WARWICK

Marta Catalano

Disclaimers

I am not a (Bayesian) statistician.
My background: mathematical analysis, optimal transport.

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

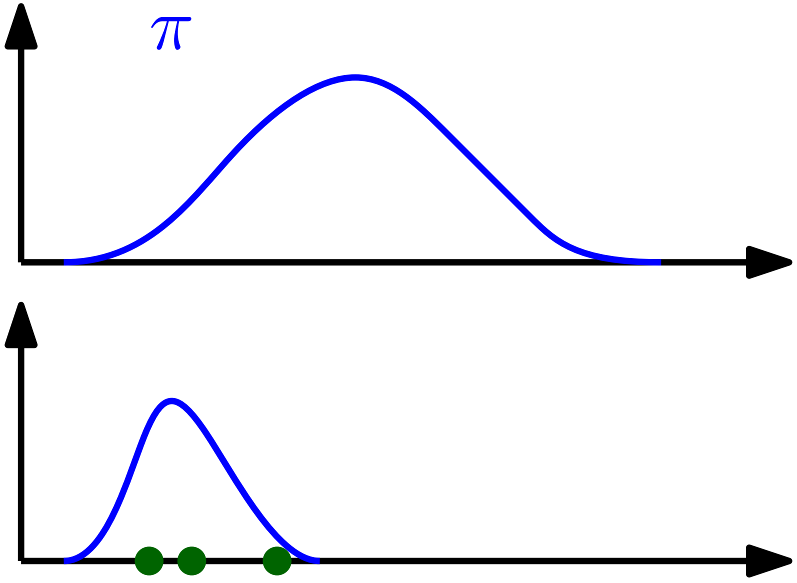
Merging of opinions

$$\theta \sim \pi$$

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

Posterior

$$\theta | X_1, \ldots, X_n$$



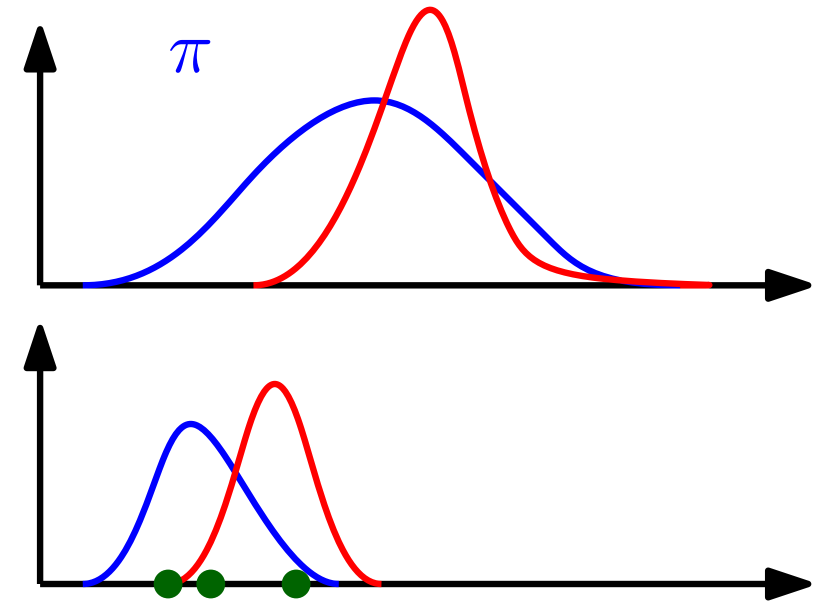
Merging of opinions

$$\theta \sim \pi$$

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

Posterior

$$\theta | X_1, \ldots, X_n$$



Question: different priors π^1, π^2, same data $X_1, \ldots X_n$.

- Does $\theta^1 | X_1, \ldots X_n \overset{\text{d}}{\simeq} \theta^2 | X_1, \ldots X_n$?
- At which rate in n?

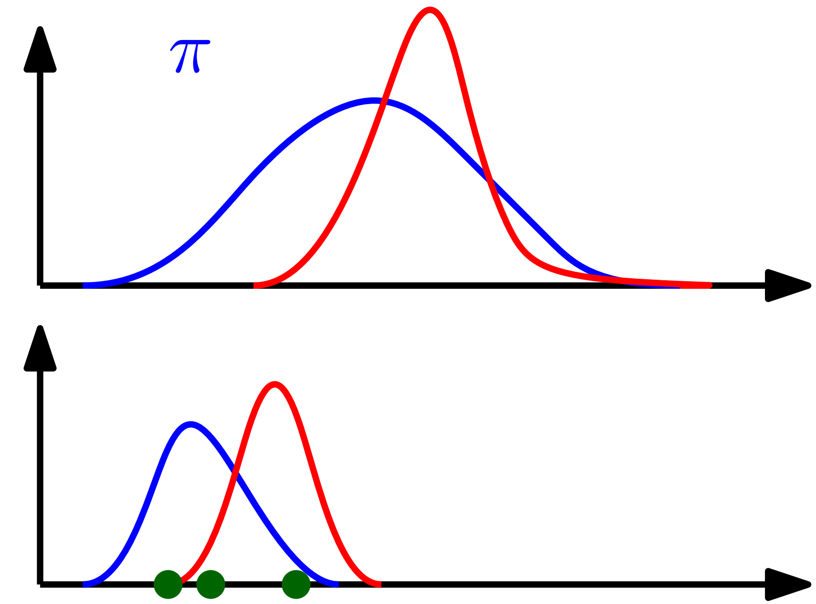
Merging of opinions

$$\theta \sim \pi$$

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

Posterior

$$\theta | X_1, \ldots, X_n$$



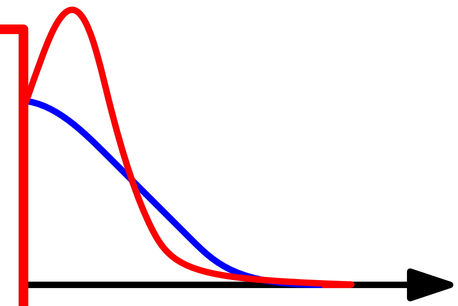
Question: different priors π^1, π^2, same data $X_1, \ldots X_n$.

- Does $\theta^1 | X_1, \ldots X_n \overset{\text{d}}{\simeq} \theta^2 | X_1, \ldots X_n$?
- At which rate in n?

distance between posteriors

Merging of opinions

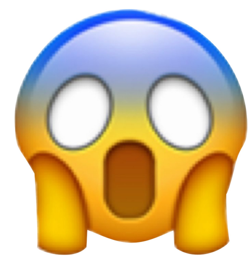
$X_1, \ldots,$



$\theta | X_1$

What about in Bayesian Nonparametrics?

Posterior: distribution over infinite dimensional spaces



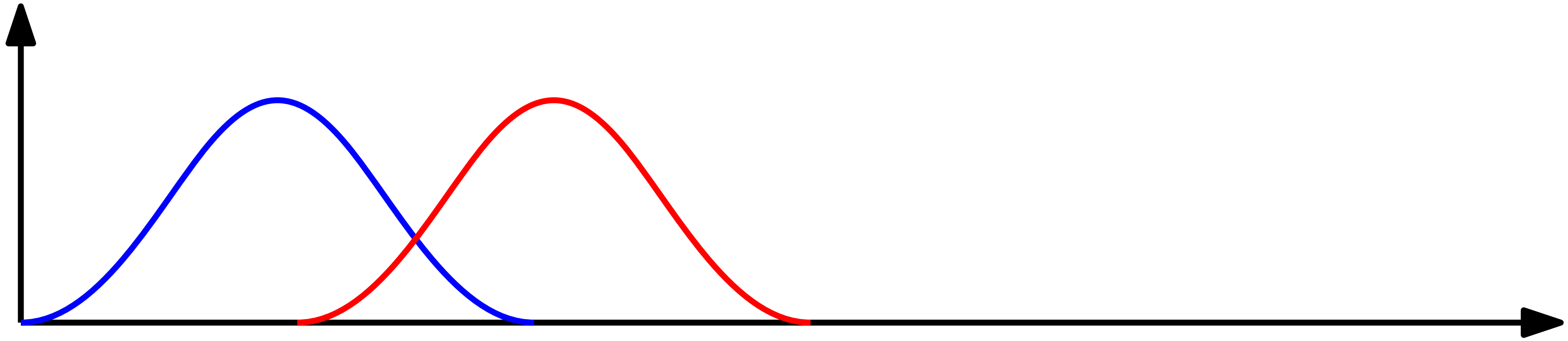
Question: different priors π^1, π^2, same data $X_1, \ldots X_n$.

- Does $\theta^1 | X_1, \ldots X_n \overset{d}{\simeq} \theta^2 | X_1, \ldots X_n$?
- At which rate in n?

distance between posteriors

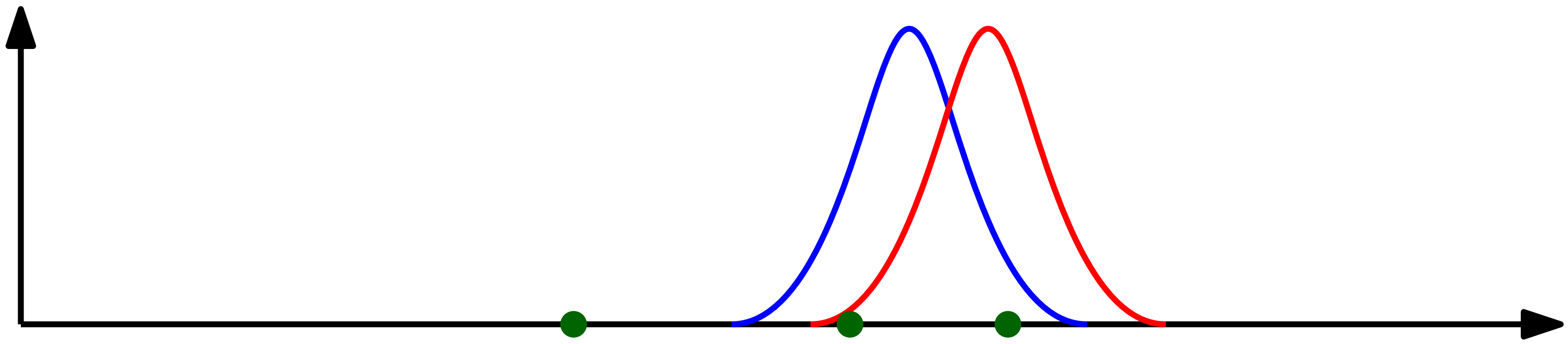
Clarifications

No need to converge to a truth



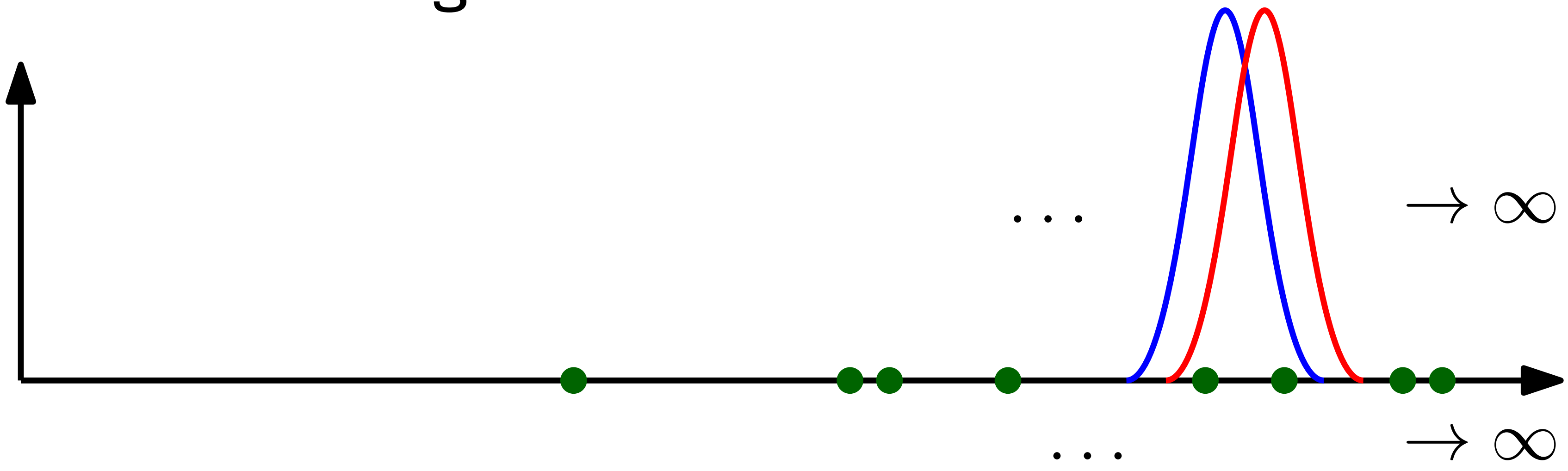
Clarifications

No need to converge to a truth



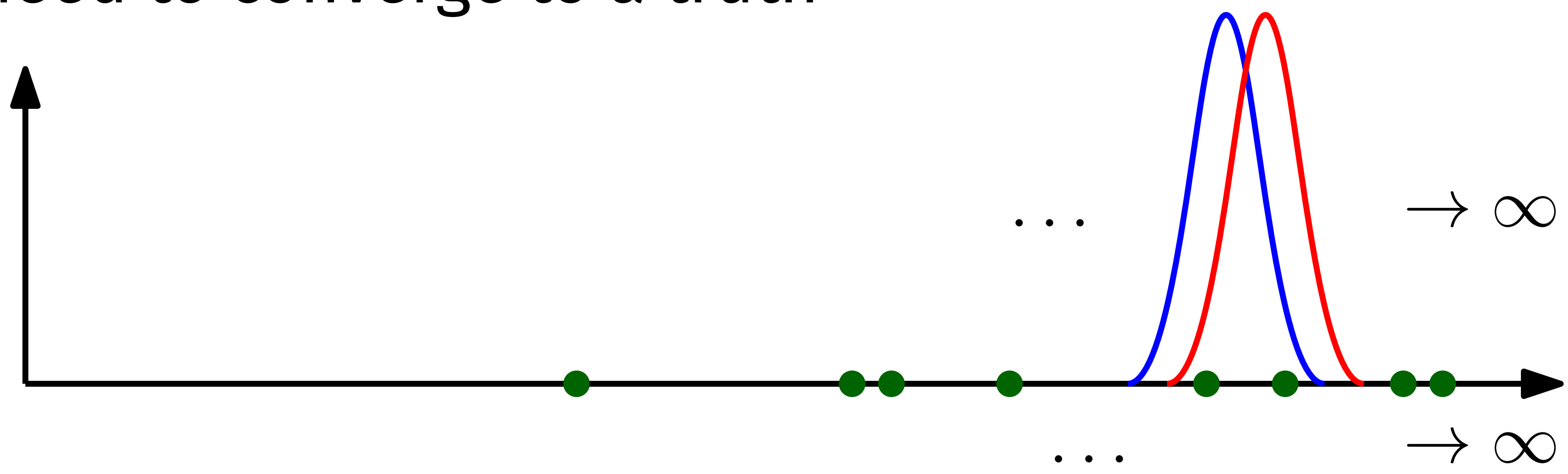
Clarifications

No need to converge to a truth

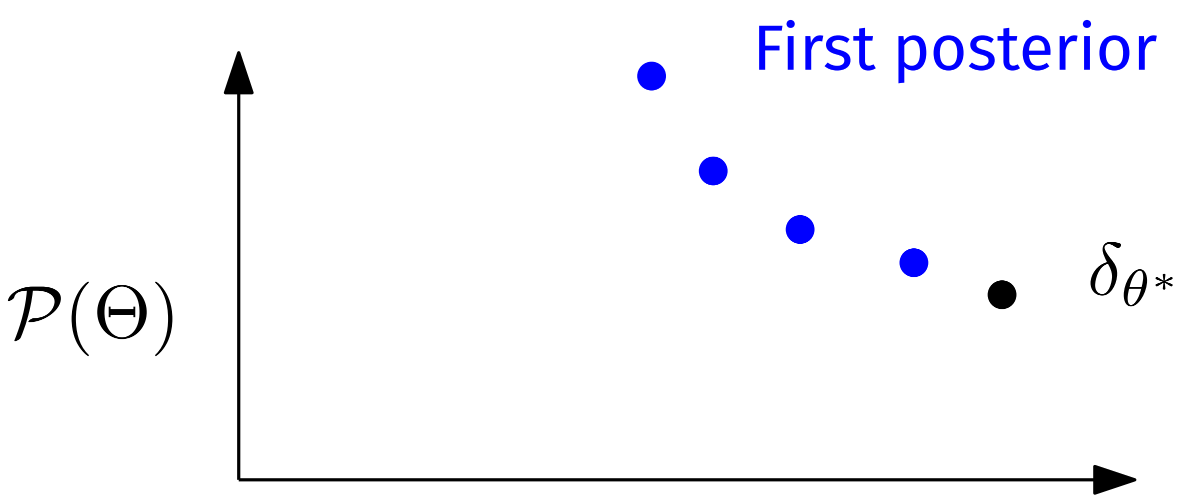


Clarifications

No need to converge to a truth

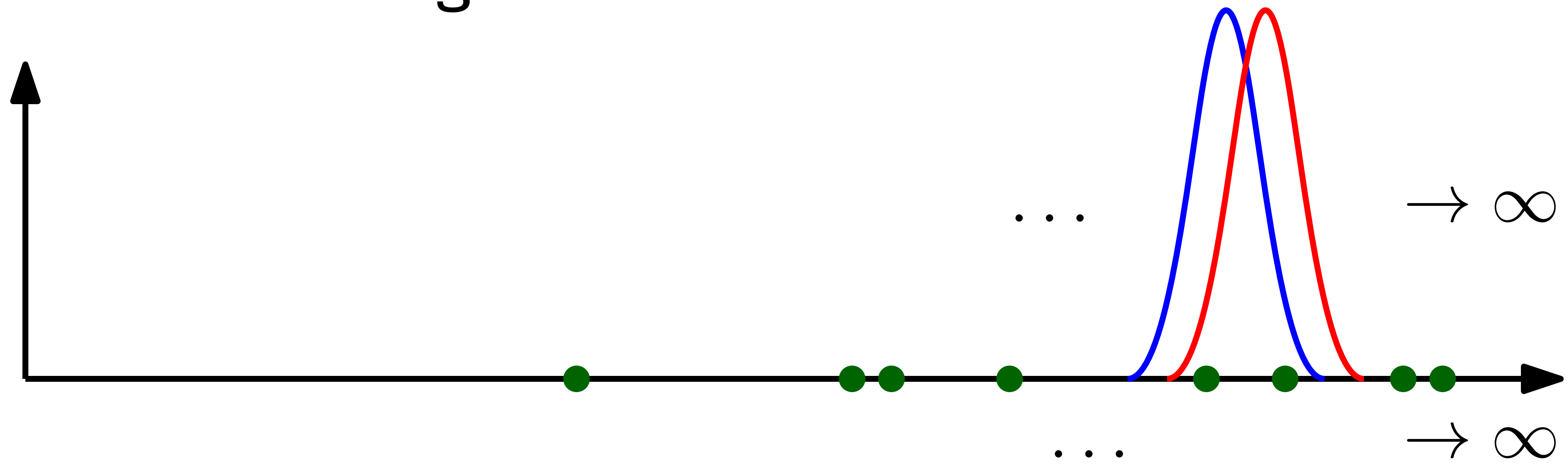


If converges to a truth, different than posterior consistency and contraction rate

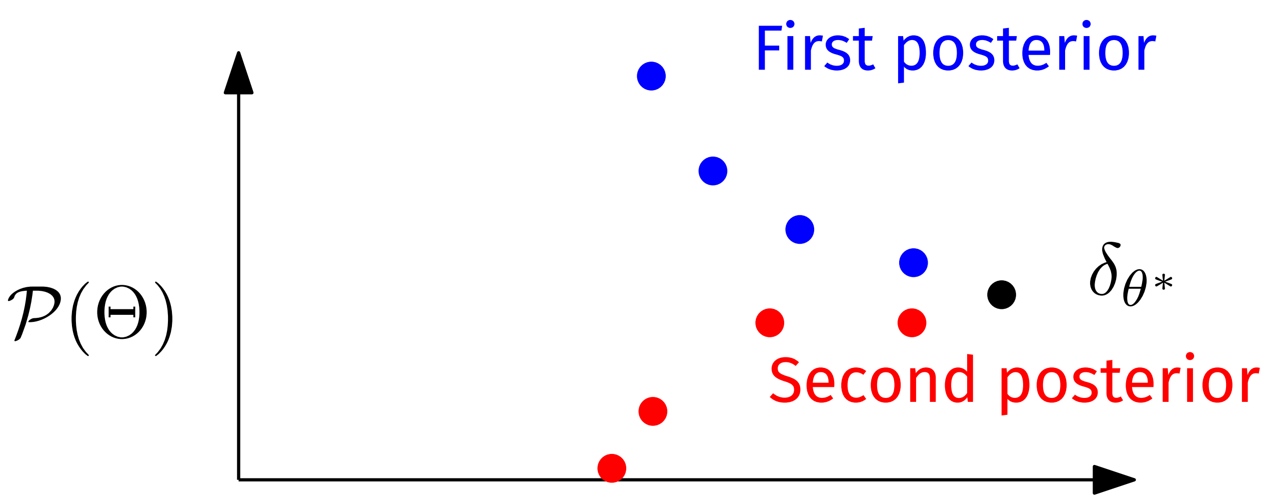


Clarifications

No need to converge to a truth

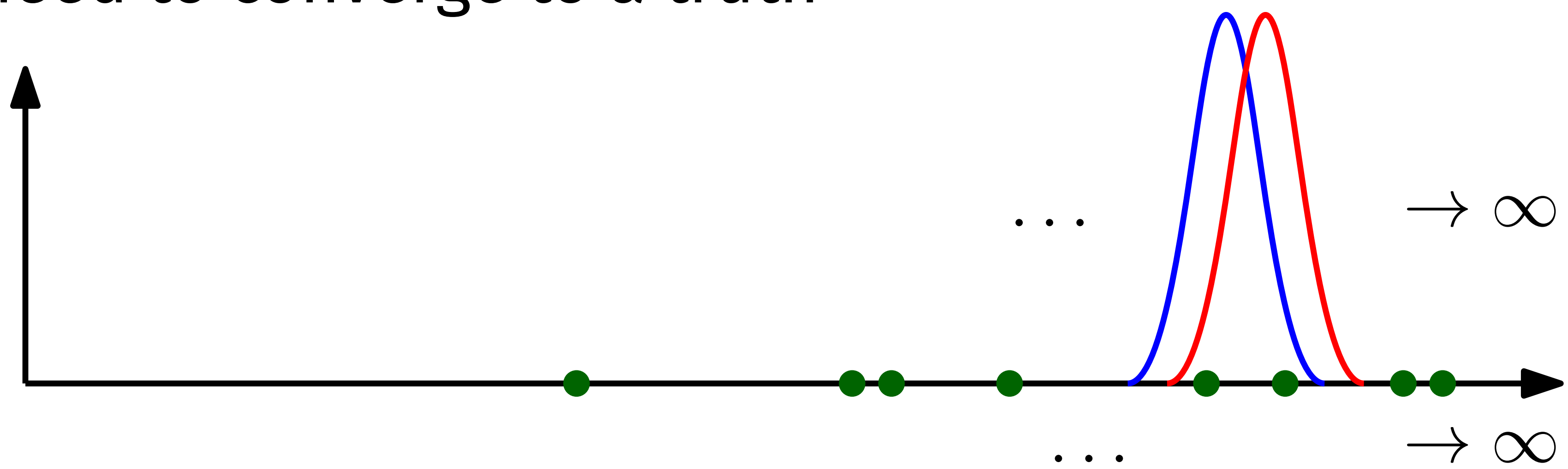


If converges to a truth, different than posterior consistency and contraction rate

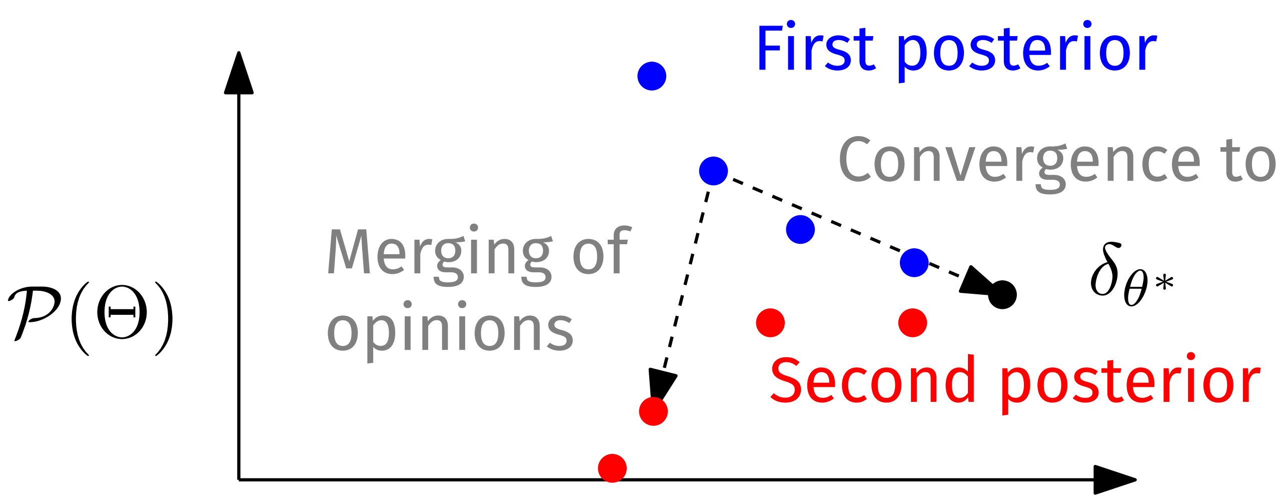


Clarifications

No need to converge to a truth



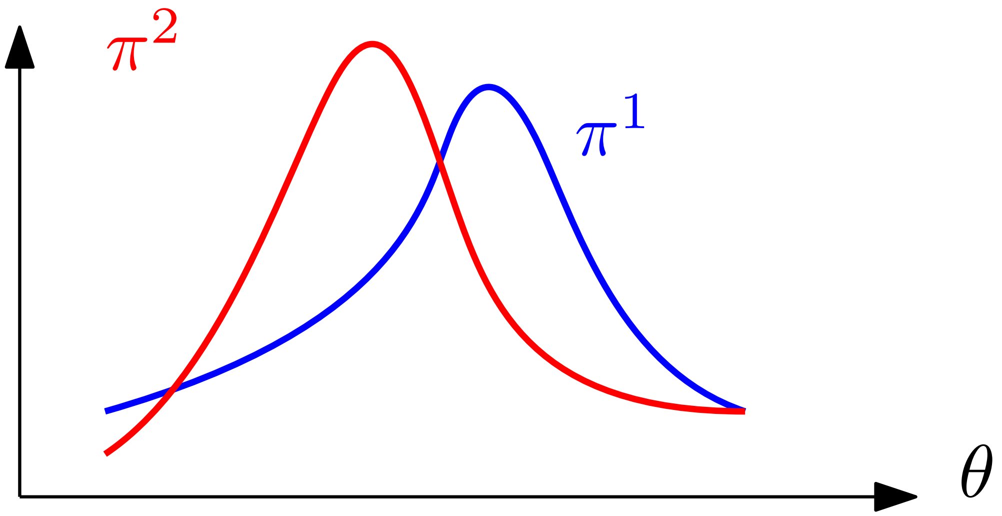
If converges to a truth, different than posterior consistency and contraction rate



Previous works and today's setting

Blackwell and Dubins: yes if $\pi^1 \ll \pi^2$ and data generated from the model.

Ley, Reinart, Swan: rates of convergence in optimal transport distance with $\pi^1 \ll \pi^2$ in 1d.



Blackwell & Dubins (1962). Merging of opinions with increasing information.
Ley, Reinart, Swan (2017). Distances between nested densities and a measure of the impact of the prior in Bayesian statistics.

Previous works and today's setting

Blackwell and Dubins: yes if $\pi^1 \ll \pi^2$ and data generated from the model.

Not valid in BNP

Ley, Reinart, Swan: rates of convergence in optimal transport distance with $\pi^1 \ll \pi^2$ in 1d.

Today:

- NonParametrics: focus on (normalized) **Completely Random Measures (CRM)** as prior.
- Optimal transport distance.
- Rates for merging of opinions.

Blackwell & Dubins (1962). Merging of opinions with increasing information.
Ley, Reinart, Swan (2017). Distances between nested densities and a measure of the impact of the prior in Bayesian statistics.

Previous works and today's setting

Blackwell and Dubins: yes if $\pi^1 \ll \pi^2$ and data generated from the model.

Not valid in BNP

Ley, Reinart, Swan: rates of convergence in optimal transport distance with $\pi^1 \ll \pi^2$ in 1d.

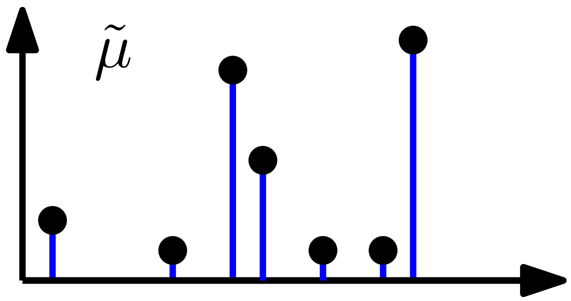
Today:

- NonParametrics: focus on (normalized) **Completely Random Measures (CRM)** as prior.
- Optimal transport distance.
- Rates for merging of opinions.

Side result: identifiability of normalization in CRM.

Side result: Asymptotic of the U latent variable.

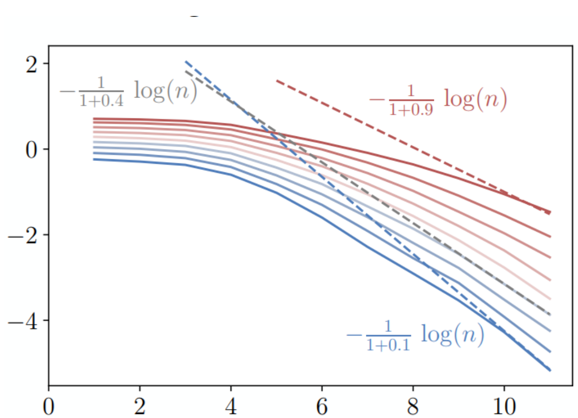
Blackwell & Dubins (1962). Merging of opinions with increasing information.
Ley, Reinart, Swan (2017). Distances between nested densities and a measure of the impact of the prior in Bayesian statistics.



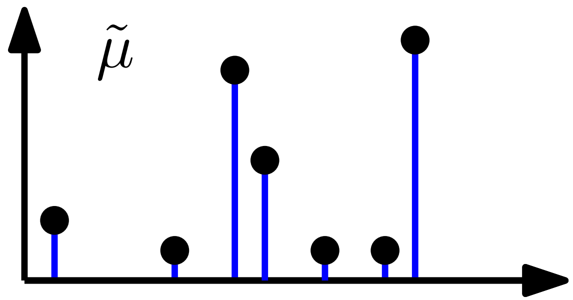
1 - Completely Random Measures a priori and posteriori

2 - Distance between CRMs

$$\inf_{(X,Y)} \{[\ldots],\ X \sim P^1, Y \sim P^2\}$$



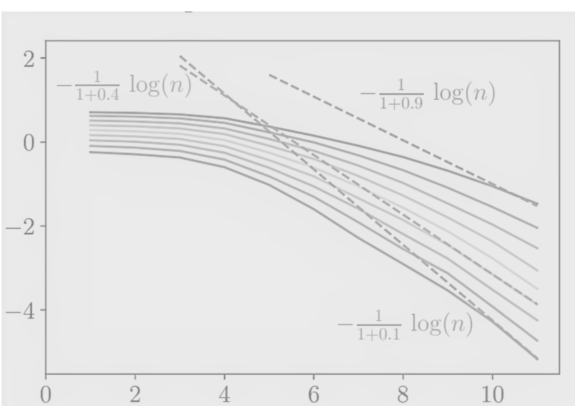
3 - Merging of opinions with CRMs



1 - Completely Random Measures a priori and posteriori

2 - Distance between CRMs

$$\inf_{(X,Y)} \{[\ldots], X \sim P^1, Y \sim P^2\}$$



3 - Merging of opinions with CRMs

Bayesian NonParametrics and normalized CRM

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

$$X_1, X_2, \ldots, X_n | \tilde{p} \overset{\text{i.i.d.}}{\sim} \tilde{p}$$ (Why? More flexibility)

(justified by exchangeability)

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

Bayesian NonParametrics and normalized CRM

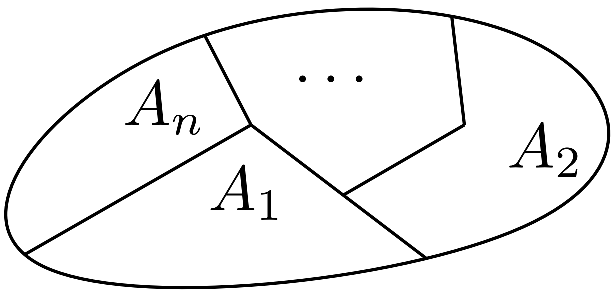
$\tilde{p}$ random probability measure on $\mathbb{X}$

$$X_1, X_2, \ldots, X_n \,|\, \tilde{p} \overset{\text{i.i.d.}}{\sim} \tilde{p}$$ (Why? More flexibility)

(justified by exchangeability)

Normalization: $\tilde{p} = \dfrac{\tilde{\mu}}{\tilde{\mu}(\mathbb{X})}$, $\tilde{\mu}$ random measure

Definition. $\tilde{\mu}$ is a **Completely Random Measure** if $\tilde{\mu}(A_1)$, $\ldots$, $\tilde{\mu}(A_n)$ independent for disjoint $A_1, \ldots, A_n$.



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

Bayesian NonParametrics and normalized CRM

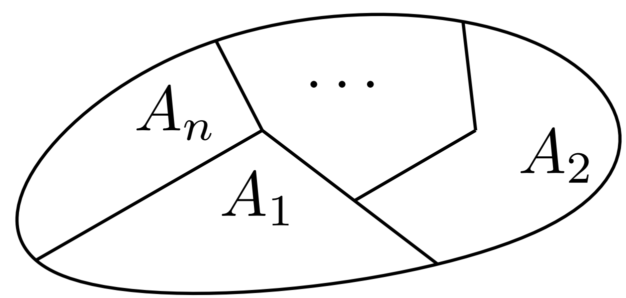
$\tilde{p}$ random probability measure on $\mathbb{X}$

$$X_1, X_2, \ldots, X_n | \tilde{p} \overset{\text{i.i.d.}}{\sim} \tilde{p}$$ (Why? More flexibility)

(justified by exchangeability)

Normalization: $\tilde{p} = \dfrac{\tilde{\mu}}{\tilde{\mu}(\mathbb{X})}$, $\tilde{\mu}$ random measure

Definition. $\tilde{\mu}$ is a **Completely Random Measure** if $\tilde{\mu}(A_1), \ldots,$ $\tilde{\mu}(A_n)$ independent for disjoint $A_1, \ldots, A_n$.



Example. $\tilde{\mu}$ is (α, b) Gamma CRM with base measure P_0.

Then $\tilde{p}$ is (α, P_0) Dirichlet process

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

The problem of identifiability

When do we have means $\tilde{\mu}^1$ and $\tilde{\mu}^2$ define the same prior

$$\tilde{p}^1 = \frac{\tilde{\mu}^1}{\tilde{\mu}^1(\mathbb{X})} \overset{d}{=} \frac{\tilde{\mu}^2}{\tilde{\mu}^2(\mathbb{X})} = \tilde{p}^2$$

The problem of identifiability

When do we have means $\tilde{\mu}^1$ and $\tilde{\mu}^2$ define the same prior

$$\tilde{p}^1 = \frac{\tilde{\mu}^1}{\tilde{\mu}^1(\mathbb{X})} \stackrel{\mathsf{d}}{=} \frac{\tilde{\mu}^2}{\tilde{\mu}^2(\mathbb{X})} = \tilde{p}^2$$

Example

- $\tilde{\mu}^1$ is (α, b^1) Gamma.
- $\tilde{\mu}^2$ is (α, b^2) Gamma.

$\rightarrow$ Dirichlet process

The problem of identifiability

When do we have

means $\tilde{\mu}^1$ and $\tilde{\mu}^2$ define the same prior

$$\tilde{p}^1 = \frac{\tilde{\mu}^1}{\tilde{\mu}^1(\mathbb{X})} \overset{\mathrm{d}}{=} \frac{\tilde{\mu}^2}{\tilde{\mu}^2(\mathbb{X})} = \tilde{p}^2$$

Example

- $\tilde{\mu}^1$ is (α, b^1) Gamma.
- $\tilde{\mu}^2$ is (α, b^2) Gamma.

$\rightarrow$ Dirichlet process

Example

- $\tilde{\mu}^1$ is $(\alpha^1, \tau^1, \sigma)$ gen. Gamma.
- $\tilde{\mu}^2$ is $(\alpha^2, \tau^2, \sigma)$ gen. Gamma.

and $\quad \frac{\alpha^1}{\alpha^2} = \left(\frac{\tau^2}{\tau^1}\right)^{\sigma}$

The problem of identifiability

When do we have — means $\tilde{\mu}^1$ and $\tilde{\mu}^2$ define the same prior

$$\tilde{p}^1 = \frac{\tilde{\mu}^1}{\tilde{\mu}^1(\mathbb{X})} \stackrel{\mathrm{d}}{=} \frac{\tilde{\mu}^2}{\tilde{\mu}^2(\mathbb{X})} = \tilde{p}^2$$

Example

- $\tilde{\mu}^1$ is (α, b^1) Gamma.
- $\tilde{\mu}^2$ is (α, b^2) Gamma.

$\rightarrow$ Dirichlet process

Example

- $\tilde{\mu}^1$ is $(\alpha^1, \tau^1, \sigma)$ gen. Gamma.
- $\tilde{\mu}^2$ is $(\alpha^2, \tau^2, \sigma)$ gen. Gamma.

and $\frac{\alpha^1}{\alpha^2} = \left(\frac{\tau^2}{\tau^1}\right)^\sigma$

Theorem. For CRM with finite mean and infinite activity,

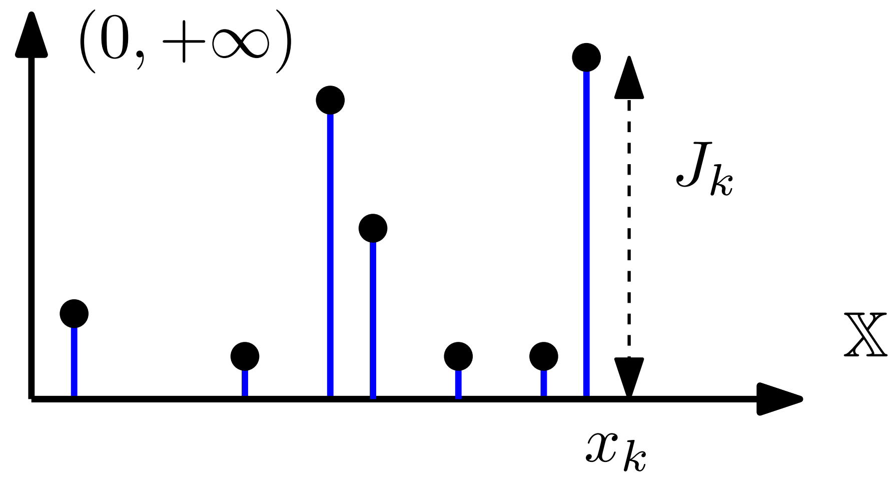
$$\frac{\tilde{\mu}^1}{\tilde{\mu}^1(\mathbb{X})} \stackrel{\mathrm{d}}{=} \frac{\tilde{\mu}^2}{\tilde{\mu}^2(\mathbb{X})} \quad \Leftrightarrow \quad \tilde{\mu}^1 \stackrel{\mathrm{d}}{=} a\tilde{\mu}^2 \text{ for } a > 0.$$

Aldous and Evans in Pitman and Yor (1992). Arcsine Laws and Interval Partitions Derived from a Stable Subordinator.

Structure of CRMs

Theorem. If no deterministic components, no fixed atoms

$$\tilde{\mu} = \sum_k J_k \delta_{x_k}$$



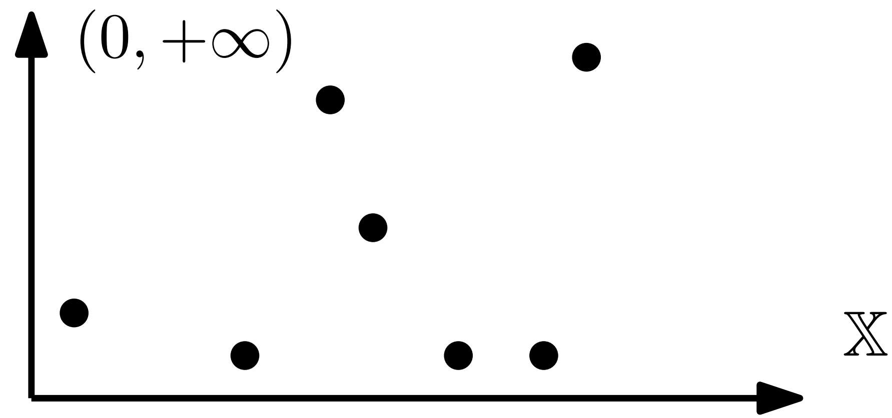
Kingman (1967). Completely random measures.

Structure of CRMs

Theorem. If no deterministic components, no fixed atoms

$$\tilde{\mu} = \sum_k J_k \delta_{x_k}$$

Look at (J_k, x_k)

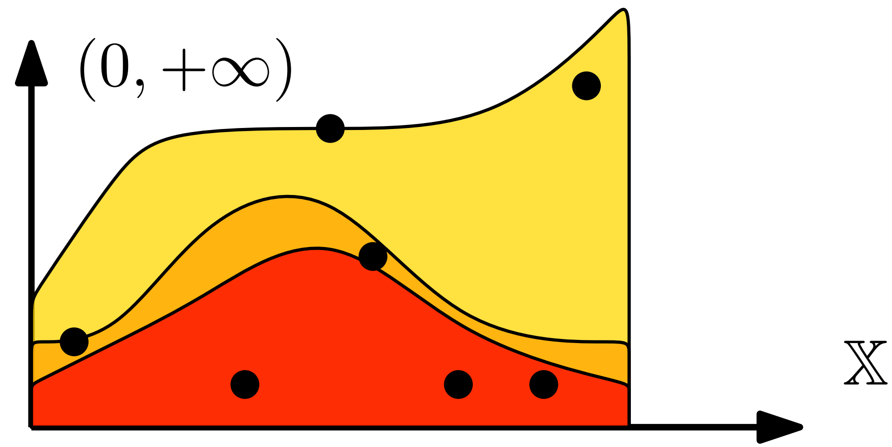


Kingman (1967). Completely random measures.

Structure of CRMs

Theorem. If no deterministic components, no fixed atoms

$$\tilde{\mu} = \sum_k J_k \delta_{x_k}$$



Look at (J_k, x_k)

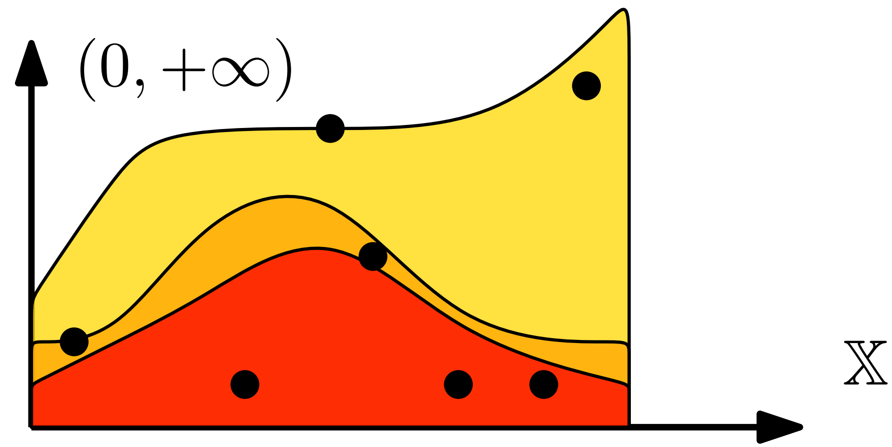
Lévy intensity $\nu(\mathrm{d}s, \mathrm{d}x)$

Kingman (1967). Completely random measures.

Structure of CRMs

Theorem. If no deterministic components, no fixed atoms

$$\tilde{\mu} = \sum_k J_k \delta_{x_k}$$



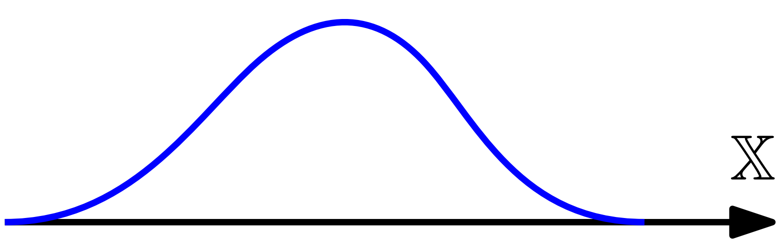
Look at (J_k, x_k)

Lévy intensity $\nu(\mathrm{d}s, \mathrm{d}x) = P_0(\mathrm{d}x)\rho_x(\mathrm{d}s)$

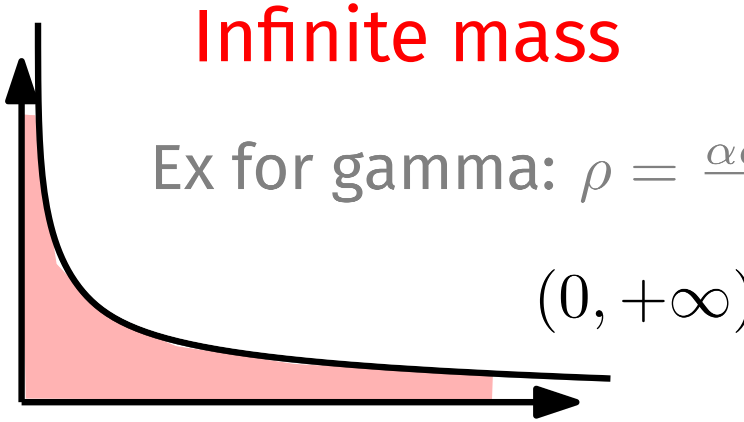
$\rho_x(\mathrm{d}s)$ distribution of jumps

Infinite mass

Distributions of atoms P_0



Ex for gamma: $\rho = \frac{\alpha e^{-s}}{s} \mathrm{d}s$



Kingman (1967). Completely random measures.

A posteriori

Data $X_1, \ldots, X_n$ gives posterior $\tilde{\mu}^* = \tilde{\mu} | X_1 \ldots X_n$

James, Lijoi and Prünster (2009). Posterior analysis for normalized random measures with independent increments.

A posteriori

Data $X_1, \ldots, X_n$ gives posterior $\tilde{\mu}^* = \tilde{\mu} | X_1 \ldots X_n$

Theorem. There exists a latent variable U such that

distinct values

Observations

$$\tilde{\mu}^* | U = \tilde{\mu}_U + \sum_{i=1}^{k} J_i \delta_{X_i}$$

Random jumps

Lévy intensity $\nu_U(\mathrm{d}s, \mathrm{d}x) = e^{-sU} \nu(\mathrm{d}s, \mathrm{d}x)$

James, Lijoi and Prünster (2009). Posterior analysis for normalized random measures with independent increments.

A posteriori

Data $X_1, \ldots, X_n$ gives posterior $\tilde{\mu}^* = \tilde{\mu} | X_1 \ldots X_n$

Theorem. There exists a latent variable U such that

distinct values

Observations

$$\tilde{\mu}^* | U = \tilde{\mu}_U + \sum_{i=1}^{k} J_i \delta_{X_i}$$

Random jumps

Lévy intensity $\nu_U(\mathrm{d}s, \mathrm{d}x) = e^{-sU} \nu(\mathrm{d}s, \mathrm{d}x)$

Consequence. $\tilde{\mu}^*$ is a **Cox CRM**, a "CRM with random Lévy intensity".

James, Lijoi and Prünster (2009). Posterior analysis for normalized random measures with independent increments.

Identifiability a posteriori

Recall for CRMs

def. $\rightarrow = \tilde{\mu}^1_{\mathcal{S}}$ "scaled CRM"

$$\frac{\tilde{\mu}^1}{\tilde{\mu}^1(\mathbb{X})} \overset{\mathrm{d}}{=} \frac{\tilde{\mu}^2}{\tilde{\mu}^2(\mathbb{X})} \quad \Leftrightarrow \quad \frac{\tilde{\mu}^1}{\mathbb{E}(\tilde{\mu}^1(\mathbb{X}))} \overset{\mathrm{d}}{=} \frac{\tilde{\mu}^2}{\mathbb{E}(\tilde{\mu}^2(\mathbb{X}))}.$$

$\Leftrightarrow \tilde{\mu}^1 \overset{\mathrm{d}}{=} a\tilde{\mu}^2$ for $a > 0$

Identifiability a posteriori

Recall for CRMs

def. $= \tilde\mu^1_{\mathcal{S}}$ "scaled CRM"

$$\frac{\tilde\mu^1}{\tilde\mu^1(\mathbb{X})} \overset{\mathrm{d}}{=} \frac{\tilde\mu^2}{\tilde\mu^2(\mathbb{X})} \quad \Leftrightarrow \quad \frac{\tilde\mu^1}{\mathbb{E}(\tilde\mu^1(\mathbb{X}))} \overset{\mathrm{d}}{=} \frac{\tilde\mu^2}{\mathbb{E}(\tilde\mu^2(\mathbb{X}))}.$$

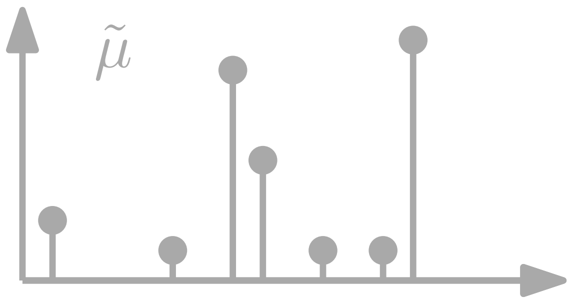
$\Leftrightarrow \tilde\mu^1 \overset{\mathrm{d}}{=} a\tilde\mu^2$ for $a > 0$

Definition. If $\tilde\mu$ conditionnaly a CRM w.r.t. U, define $\tilde\mu_{\mathcal{S}}$:

$$\tilde\mu_{\mathcal{S}}|U = \left.\frac{\tilde\mu|U}{\mathbb{E}(\tilde\mu(\mathbb{X})|U)}\right| U$$

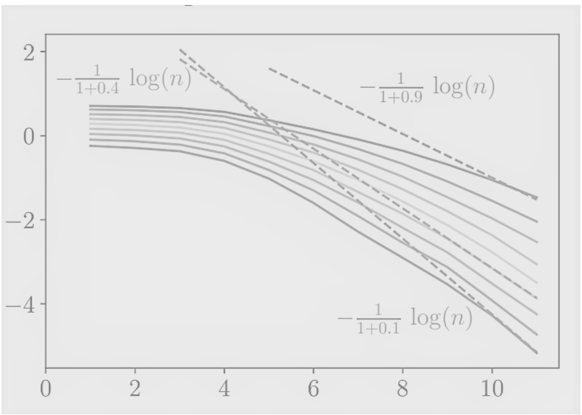
Theorem. If $\tilde\mu^1$, $\tilde\mu^2$ are both Cox CRM then

$$\frac{\tilde\mu^1}{\tilde\mu^1(\mathbb{X})} \overset{\mathrm{d}}{=} \frac{\tilde\mu^2}{\tilde\mu^2(\mathbb{X})} \quad \Leftrightarrow \quad \tilde\mu^1_{\mathcal{S}} \overset{\mathrm{d}}{=} \tilde\mu^2_{\mathcal{S}}.$$



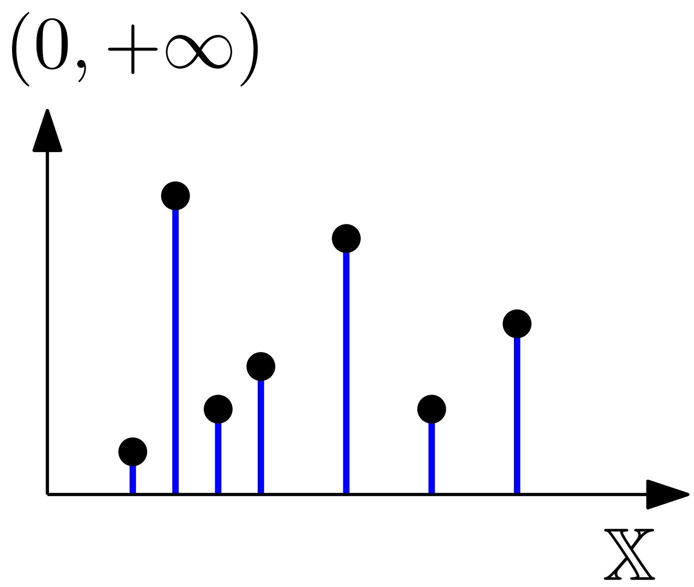
1 - Completely Random Measures a priori and posteriori

2 - Distance between CRMs $\inf_{(X,Y)} \{[\ldots], X \sim P^1, Y \sim P^2\}$



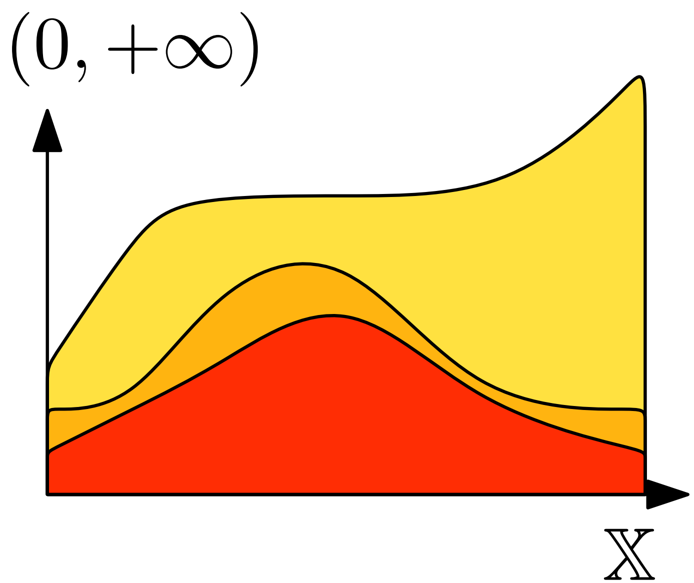
3 - Merging of opinions with CRMs

Distance between CRMs



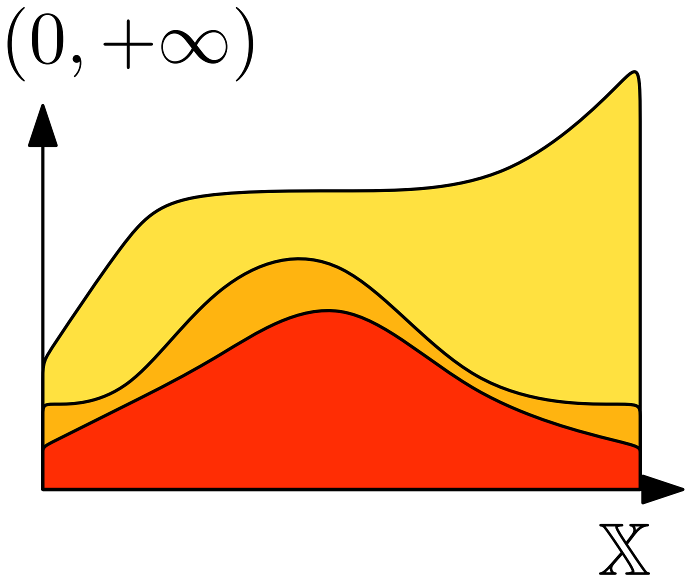
$\tilde{\mu}^1 \quad \leftrightsquigarrow \quad \tilde{\mu}^2$

Distance between CRMs



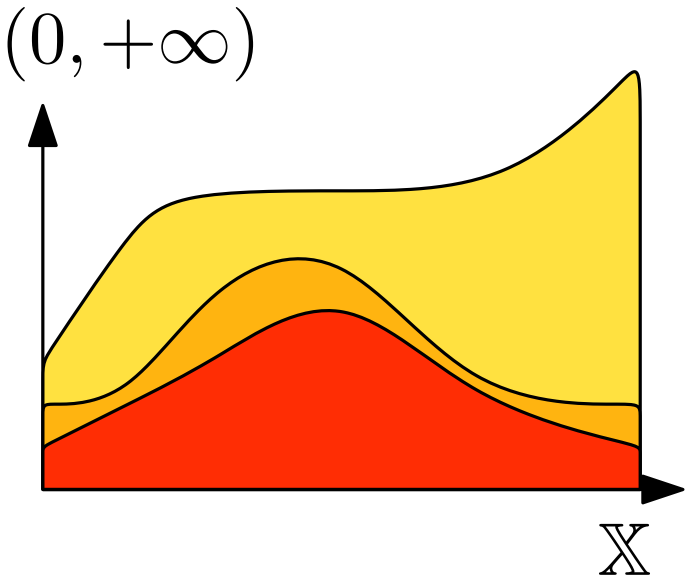
$$\begin{array}{ccc} \tilde{\mu}^1 & \leftrightsquigarrow & \tilde{\mu}^2 \\ \downarrow & & \downarrow \\ \nu^1(\mathrm{d}s, \mathrm{d}x) & \leftrightsquigarrow & \nu^2(\mathrm{d}s, \mathrm{d}x) \end{array}$$

Distance between CRMs



$$
\begin{array}{ccc}
\tilde{\mu}^1 & \leftrightsquigarrow & \tilde{\mu}^2 \\
\downarrow & & \downarrow \\
\nu^1(\mathrm{d}s, \mathrm{d}x) & \leftrightsquigarrow & \nu^2(\mathrm{d}s, \mathrm{d}x) \\
\downarrow & & \downarrow \\
P_0^1(\mathrm{d}x)\rho_x^2(\mathrm{d}s) & \leftrightsquigarrow & P_0^2(\mathrm{d}x)\rho_x^2(\mathrm{d}s)
\end{array}
$$

Distance between CRMs

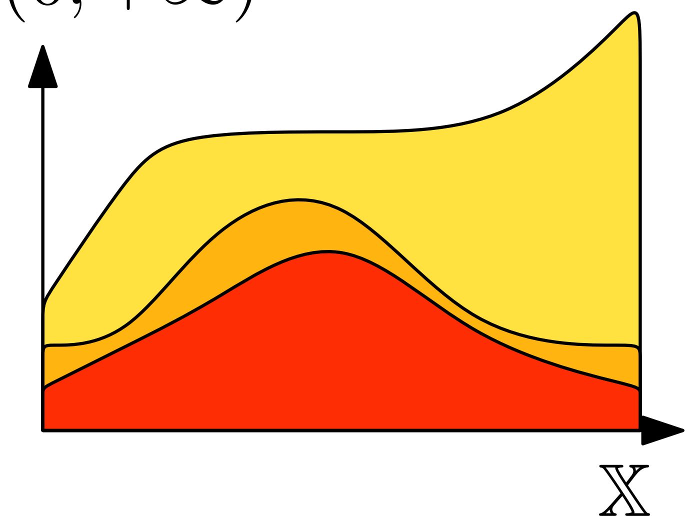


$$
\begin{array}{ccc}
\tilde{\mu}^1 & \leftrightsquigarrow & \tilde{\mu}^2 \\
\downarrow & & \downarrow \\
\nu^1(\mathrm{d}s,\mathrm{d}x) & \leftrightsquigarrow & \nu^2(\mathrm{d}s,\mathrm{d}x) \\
\downarrow & & \downarrow \\
P_0^1(\mathrm{d}x)\rho_x^2(\mathrm{d}s) & \leftrightsquigarrow & P_0^2(\mathrm{d}x)\rho_x^2(\mathrm{d}s)
\end{array}
$$

Definition.

$$d_{\mathrm{W}}(\nu^1,\nu^2) = \inf_{(X,Y)} \{\mathbb{E}(d_{\mathbb{X}}(X,Y) \quad) \quad X \sim P_0^1,\, Y \sim P_0^2\}$$

Distance between CRMs



$$\begin{array}{ccc} \tilde{\mu}^1 & \leftrightsquigarrow & \tilde{\mu}^2 \\ \downarrow & & \downarrow \\ \nu^1(\mathrm{d}s, \mathrm{d}x) & \leftrightsquigarrow & \nu^2(\mathrm{d}s, \mathrm{d}x) \\ \downarrow & & \downarrow \\ P_0^1(\mathrm{d}x)\rho_x^2(\mathrm{d}s) & \leftrightsquigarrow & P_0^2(\mathrm{d}x)\rho_x^2(\mathrm{d}s) \end{array}$$

$W_*(\rho^1, \rho^2) = \int_0^{+\infty} |\rho^1(t, +\infty) - \rho^2(t, +\infty)| \mathrm{d}t$

"Extended Wasserstein distance"

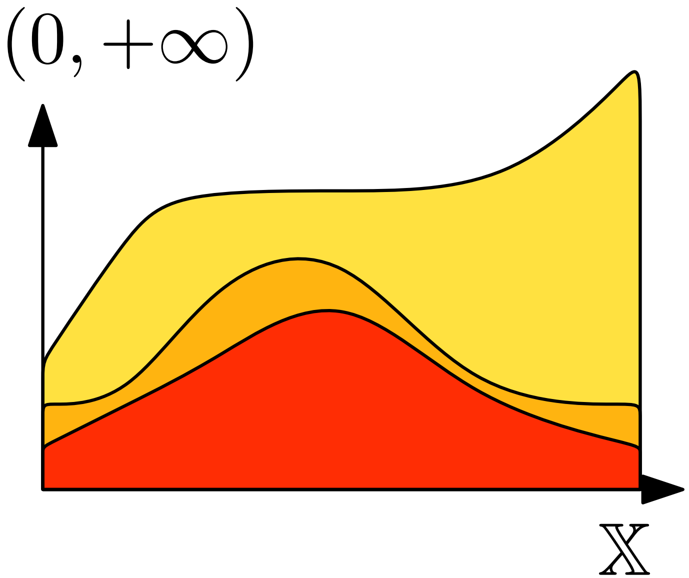
Definition.

$$d_{\mathrm{W}}(\nu^1, \nu^2) = \inf_{(X,Y)} \{\mathbb{E}(d_{\mathbb{X}}(X, Y) + W_*(\rho_X^1(\mathrm{d}s), \rho_Y^2(\mathrm{d}s))\ ) \quad X \sim P_0^1, Y \sim P_0^2\}$$

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.

Distance between CRMs



$$\tilde{\mu}^1 \quad \leftrightsquigarrow \quad \tilde{\mu}^2$$

$$\downarrow \qquad\qquad \downarrow$$

$$\nu^1(\mathrm{d}s, \mathrm{d}x) \quad \leftrightsquigarrow \quad \nu^2(\mathrm{d}s, \mathrm{d}x)$$

$$\downarrow \qquad\qquad \downarrow$$

$$P_0^1(\mathrm{d}x)\rho_x^2(\mathrm{d}s) \quad \leftrightsquigarrow \quad P_0^2(\mathrm{d}x)\rho_x^2(\mathrm{d}s)$$

$$W_*(\rho^1, \rho^2) = \int_0^{+\infty} |\rho^1(t, +\infty) - \rho^2(t, +\infty)| \mathrm{d}t$$

"Extended Wasserstein distance"

Definition.

$$d_{\mathrm{W}}(\nu^1, \nu^2) = \inf_{(X,Y)} \{\mathbb{E}(d_{\mathbb{X}}(X, Y) + W_*(\rho_X^1(\mathrm{d}s), \rho_Y^2(\mathrm{d}s))\ ) \quad X \sim P_0^1, Y \sim P_0^2\}$$

If $\tilde{\mu}^1$, $\tilde{\mu}^2$ have random Lévy intensity $\tilde{\nu}^1, \tilde{\nu}^2$

$$d_{\mathrm{WoW}}(\tilde{\nu}^1, \tilde{\nu}^2) = \inf_{(\tilde{\nu}^1, \tilde{\nu}^2)} \mathbb{E}(d_{\mathrm{W}}(\tilde{\nu}^1, \tilde{\nu}^2))$$

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.

Comments on the distance

- Interpretable.
- Analytically tractable. ✓

Comments on the distance

- Interpretable.
- Analytically tractable. ✓

Topology:

Theorem. If scaled homogeneous CRMs with same P_0:

$$W_1(\tilde{\mu}^1(A), \tilde{\mu}^2(A)) \leq d_{\mathrm{W}}(\nu^1, \nu^2).$$

Consequence. Convergence in our new distance implies weak convergence of the random measures.

Comments on the distance

- Interpretable.
- Analytically tractable. ✓

Topology:

Theorem. If scaled homogeneous CRMs with same P_0:

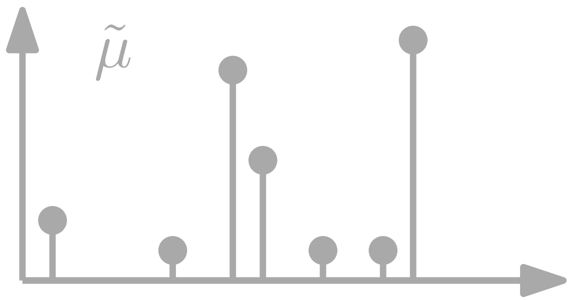
$$W_1(\tilde{\mu}^1(A), \tilde{\mu}^2(A)) \leq d_{\mathrm{W}}(\nu^1, \nu^2).$$

In General:

$$W_1\left(\int_{\mathbb{X}} f\,\mathrm{d}\tilde{\mu}^1, \int_{\mathbb{X}} f\,\mathrm{d}\tilde{\mu}^2\right) \leq c_f d_{\mathrm{W}}(\nu^1, \nu^2)$$

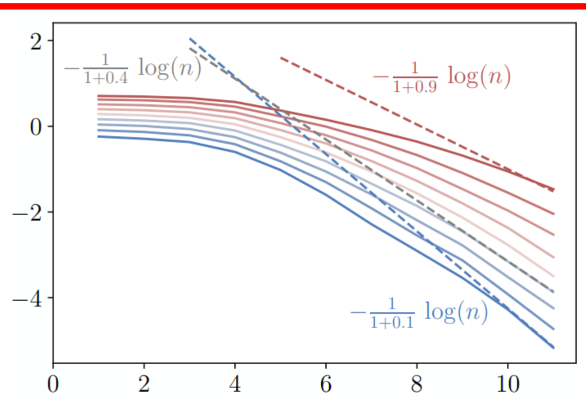
with $c_f = \max(\|f\|_\infty, \mathrm{Lip}(f))$.

Consequence. Convergence in our new distance implies weak convergence of the random measures.



1 - Completely Random Measures a priori and posteriori

2 - Distance between CRMs



3 - Merging of opinions with CRMs

Dirichlet process (a.k.a. normalized Gamma CRM)

$$\tilde{\mu}^i \sim \text{Gamma}(\alpha^i, P_0^i), \qquad i = 1, 2$$

$$\nu^i = \alpha^i \frac{e^{-s}}{s} \mathrm{d}s P_0^i(\mathrm{d}x)$$

Total base measure

Base probability measure

Dirichlet process (a.k.a. normalized Gamma CRM)

$$\tilde{\mu}^i \sim \mathrm{Gamma}(\alpha^i, P_0^i), \qquad i = 1, 2$$

$\nu^i = \alpha^i \frac{e^{-s}}{s} \mathrm{d}s P_0^i(\mathrm{d}x)$

Total base measure

Base probability measure

Data $X_1, \ldots, X_n, \ldots$, **no "truth"**, only $\sup_n \frac{1}{n} \sum_{k=1}^n |X_k| < +\infty$.

Dirichlet process (a.k.a. normalized Gamma CRM)

$$\tilde{\mu}^i \sim \mathrm{Gamma}(\alpha^i, P_0^i), \qquad i = 1, 2$$

$$\nu^i = \alpha^i \frac{e^{-s}}{s} \mathrm{d}s P_0^i(\mathrm{d}x)$$

Total base measure

Base probability measure

Data $X_1, \ldots, X_n, \ldots$, **no "truth"**, only $\sup_n \frac{1}{n} \sum_{k=1}^n |X_k| < +\infty$.

Then

$$d_{\mathrm{W}}(\tilde{\mu}^{1,*}, \tilde{\mu}^{2,*}) = J + A \asymp \frac{1}{n}$$

Jumps

Atoms

Dirichlet process (a.k.a. normalized Gamma CRM)

$$\tilde{\mu}^i \sim \text{Gamma}(\alpha^i, P_0^i), \qquad i = 1, 2$$

$$\nu^i = \alpha^i \frac{e^{-s}}{s} \mathrm{d}s P_0^i(\mathrm{d}x)$$

Base probability measure

Total base measure

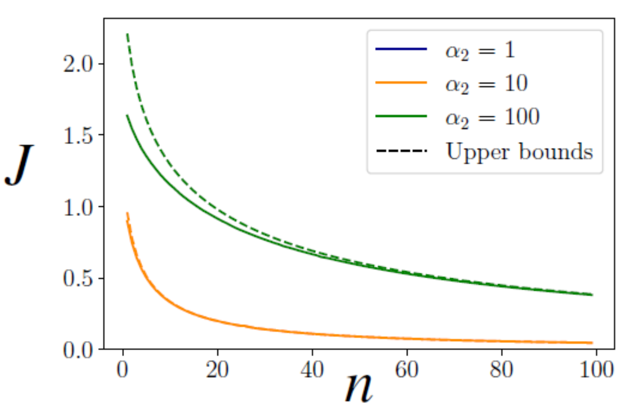
Data $X_1, \ldots, X_n, \ldots$, **no "truth"**, only $\sup_n \frac{1}{n} \sum_{k=1}^{n} |X_k| < +\infty$.

Then

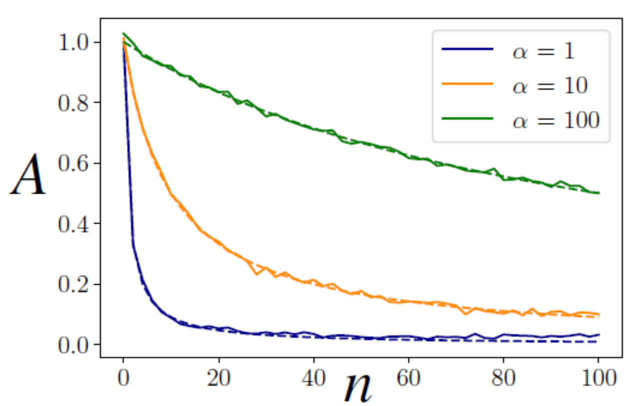
$$d_{\mathrm{W}}(\tilde{\mu}^{1,*}, \tilde{\mu}^{2,*}) = J + A \asymp \frac{1}{n}$$

Jumps

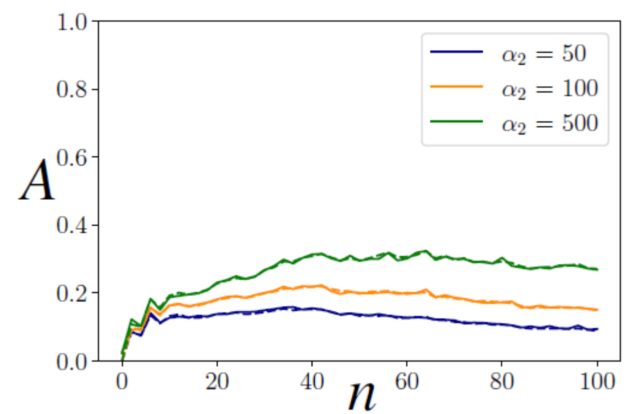
Atoms



J decreasing



$\alpha^1 = \alpha^2$ decreasing



$P_0^1 = P_0^2$: max at $\sqrt{\alpha^1 \alpha^2}$

Dirichlet process: intuition

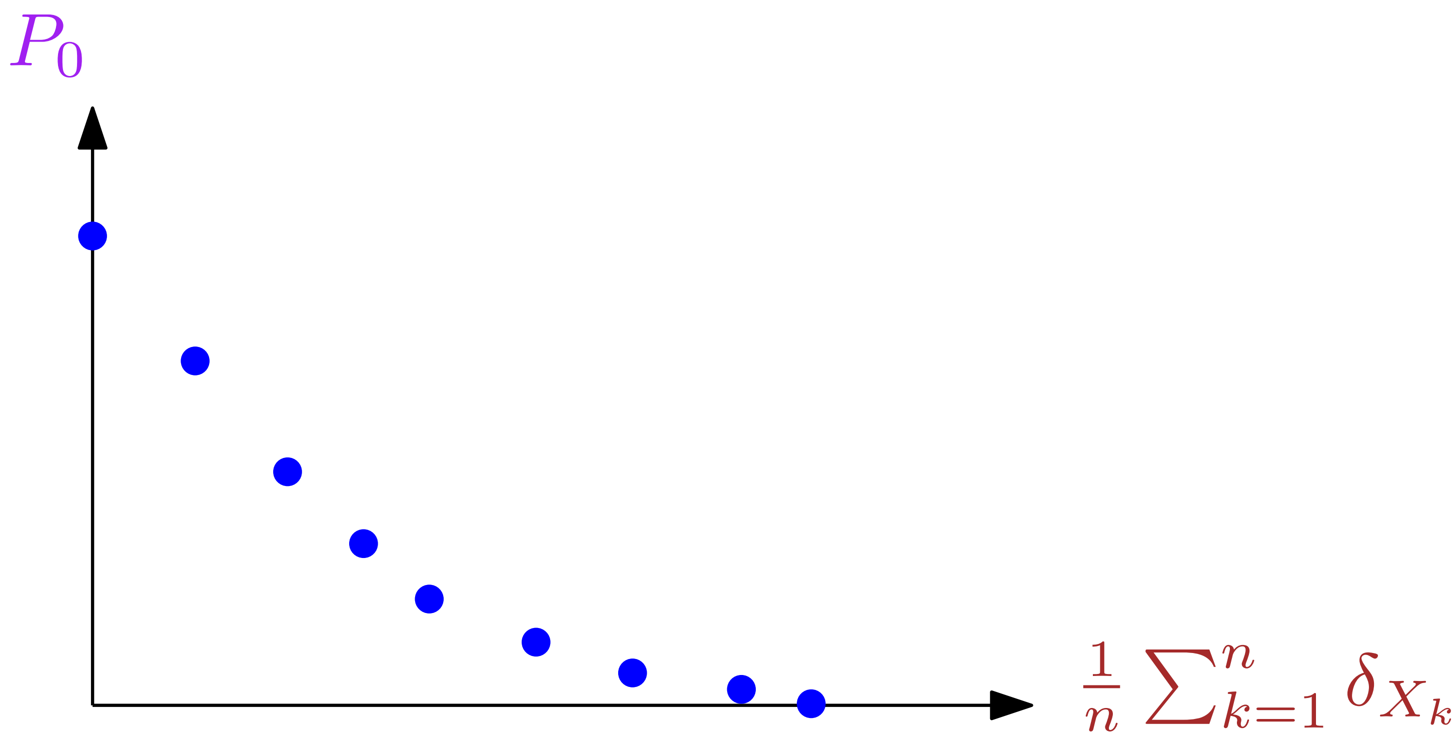
Predictive distribution

$$X_{n+1}|X_1,\ldots X_n \sim \frac{\alpha}{\alpha+n}P_0 + \frac{n}{\alpha+n}\frac{1}{n}\sum_{k=1}^{n}\delta_{X_k}$$

Dirichlet process: intuition

Predictive distribution

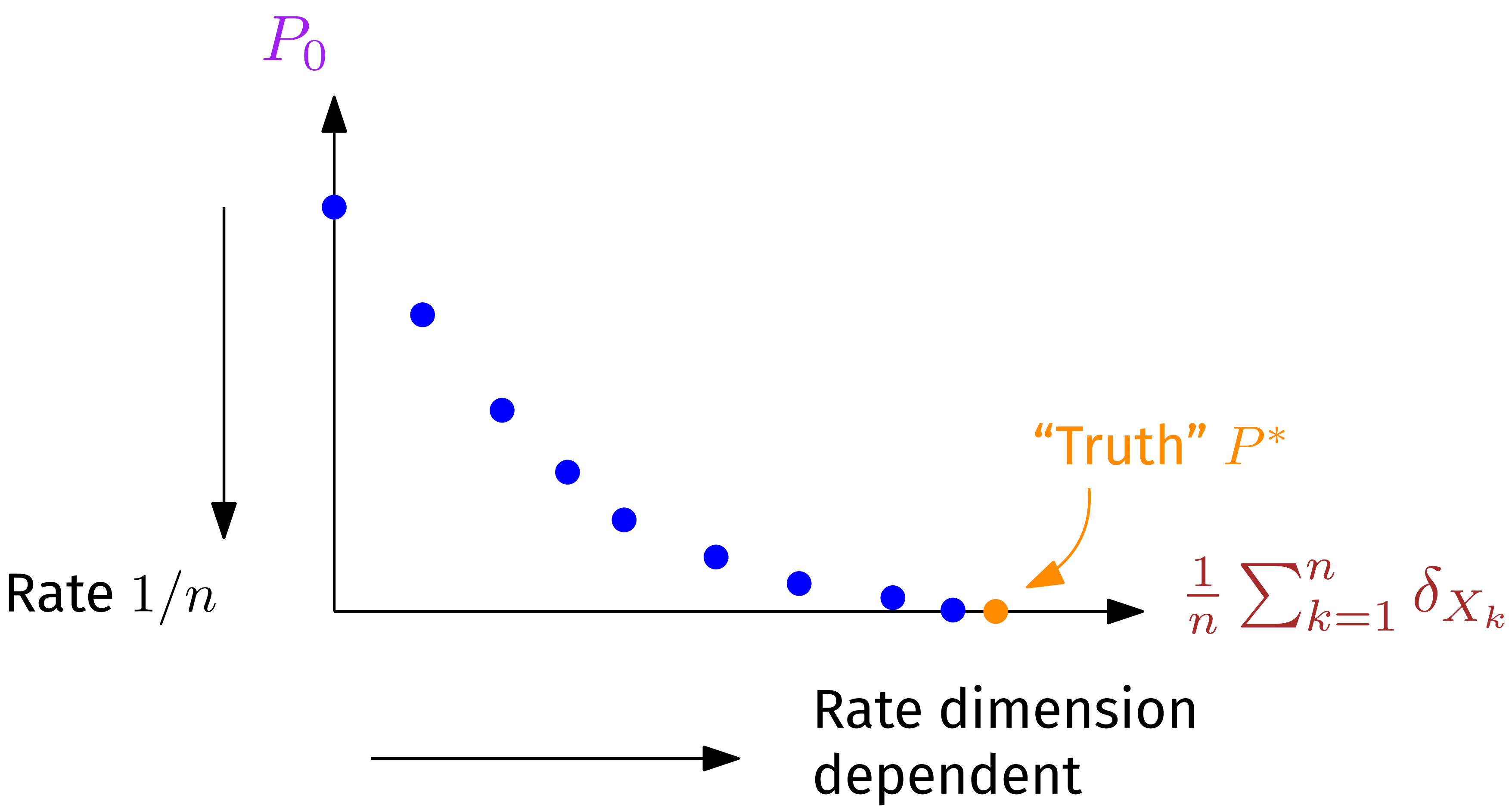
$$X_{n+1}|X_1, \ldots X_n \sim \frac{\alpha}{\alpha+n} P_0 + \frac{n}{\alpha+n} \frac{1}{n} \sum_{k=1}^{n} \delta_{X_k}$$



Dirichlet process: intuition

Predictive distribution

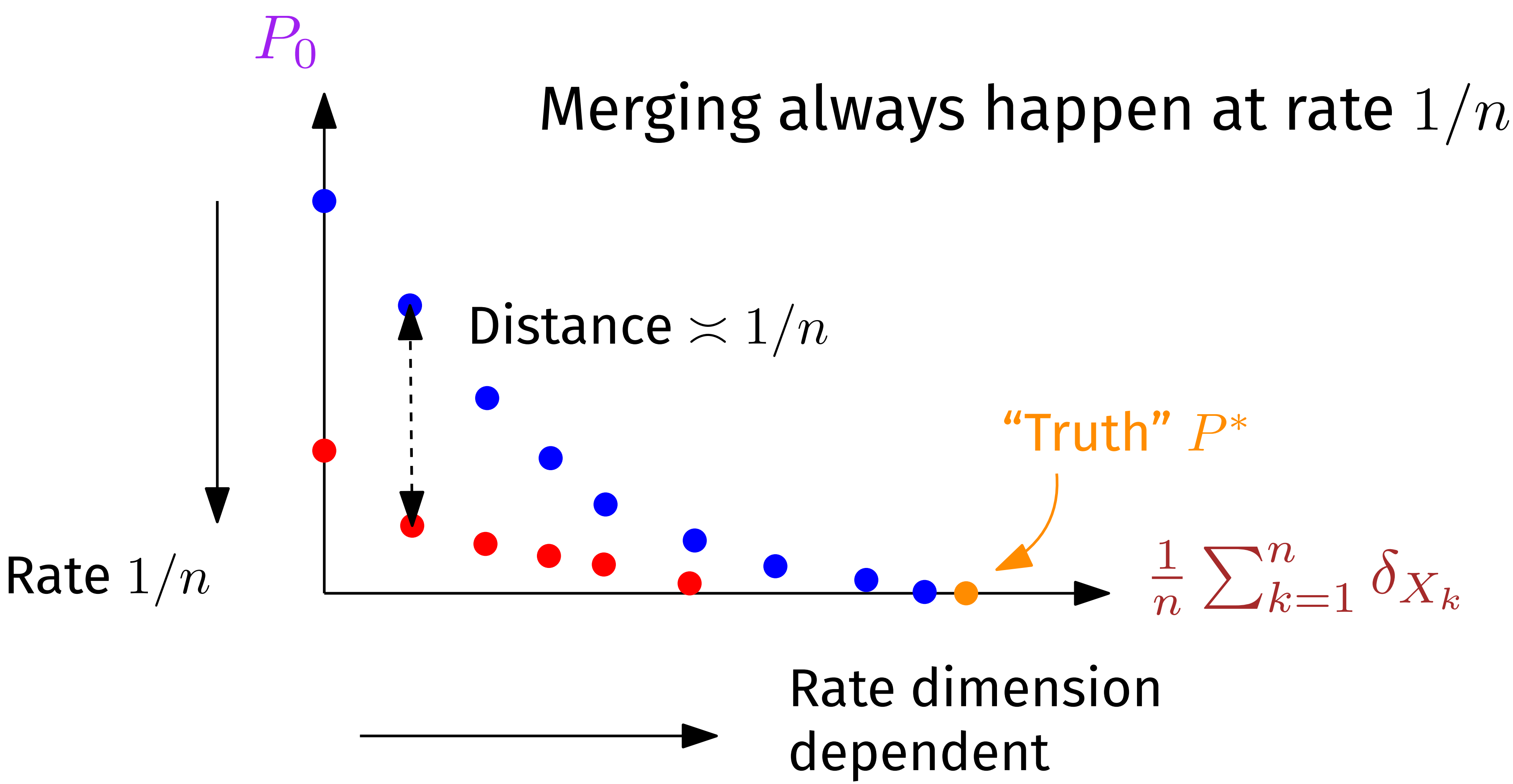
$$X_{n+1}|X_1,\dots X_n \sim \frac{\alpha}{\alpha+n}P_0 + \frac{n}{\alpha+n}\frac{1}{n}\sum_{k=1}^{n}\delta_{X_k}$$



Dirichlet process: intuition

Predictive distribution

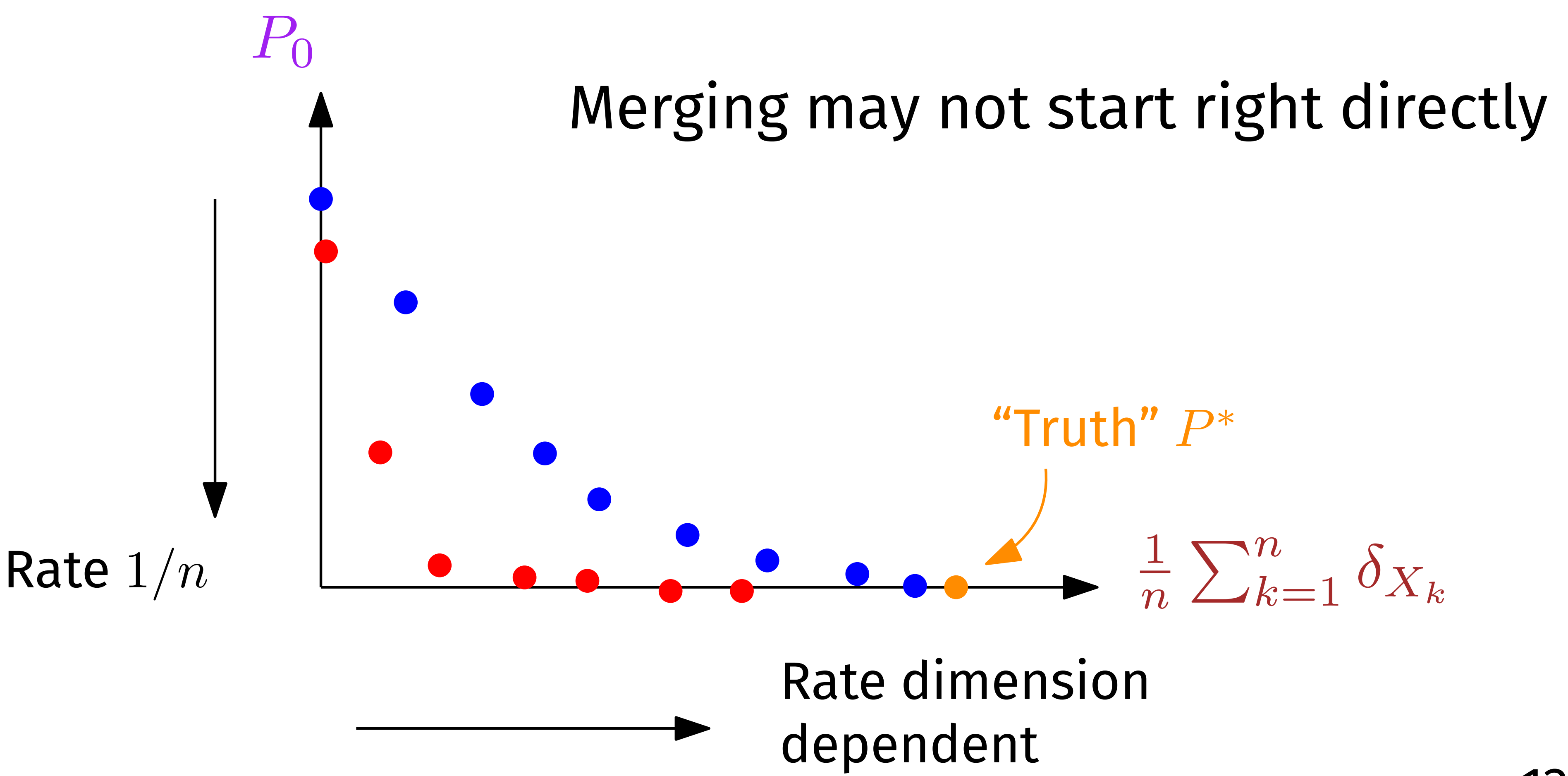
$$X_{n+1}|X_1,\dots X_n \sim \frac{\alpha}{\alpha+n}P_0 + \frac{n}{\alpha+n}\frac{1}{n}\sum_{k=1}^{n}\delta_{X_k}$$



Dirichlet process: intuition

Predictive distribution

$$X_{n+1}|X_1, \dots X_n \sim \frac{\alpha}{\alpha+n} P_0 + \frac{n}{\alpha+n} \frac{1}{n} \sum_{k=1}^{n} \delta_{X_k}$$



Generalized Gamma: setting

Generalized Gamma CRM with parameters α, P_0 and $\sigma \in [0, 1)$

$$\mathrm{d}\nu(s,x) = \frac{\alpha}{\Gamma(1-\sigma)} \frac{e^{-s}}{s^{1+\sigma}} \mathrm{d}s \mathrm{d}P_0(x) \xrightarrow{\sigma = 0} \mathrm{Gamma}(\alpha, P_0)$$

$\tilde{\mu}^1$ $\tilde{\mu}^2$

Lijoi, Mena and Prünster (2007). Controlling the reinforcement in Bayesian nonparametric mixture models.

Generalized Gamma: setting

Generalized Gamma CRM with parameters α, P_0 and $\sigma \in [0,1)$

$$\mathrm{d}\nu(s,x) = \frac{\alpha}{\Gamma(1-\sigma)} \frac{e^{-s}}{s^{1+\sigma}} \mathrm{d}s \mathrm{d}P_0(x) \xrightarrow{\sigma = 0} \mathrm{Gamma}(\alpha, P_0)$$

$\tilde{\mu}^1$ $\tilde{\mu}^2$

A posteriori, latent variable U such that

$$\mathrm{d}\nu^*|U(s,x) = \frac{(1+U)^\sigma}{c^\sigma} \frac{\alpha}{\Gamma(1-\sigma)} \frac{e^{-cs}}{s^{1+\sigma}} \mathrm{d}s \mathrm{d}P_0(x) + \sum_{i=1}^{k} (n_i - \sigma) \frac{e^{-cs}}{s} \mathrm{d}s \delta_{X_i^*}(\mathrm{d}x)$$

distinct values

with $c = \alpha(1+U)^\sigma + n - k\sigma$

Number observatiions

Lijoi, Mena and Prünster (2007). Controlling the reinforcement in Bayesian nonparametric mixture models.

Generalized Gamma: results

U density proportional to $u^{n-1}(1+u)^{k-n\sigma}e^{-\alpha/\sigma(1+u)^{\sigma}}$.

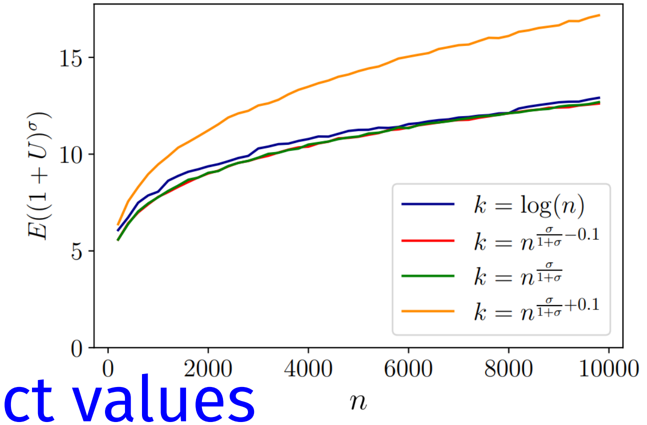
Generalized Gamma: results

U density proportional to $u^{n-1}(1+u)^{k-n\sigma}e^{-\alpha/\sigma(1+u)^{\sigma}}$.

Theorem. There holds $(1+U)^{\sigma} \sim r_n$ in L^1 with

$$r_n \asymp \begin{cases} n^{\sigma/(1+\sigma)} & \text{if } k \lesssim n^{\sigma/(1+\sigma)} \\ k & \text{if } k \gg n^{\sigma/(1+\sigma)} \end{cases}$$

distinct values

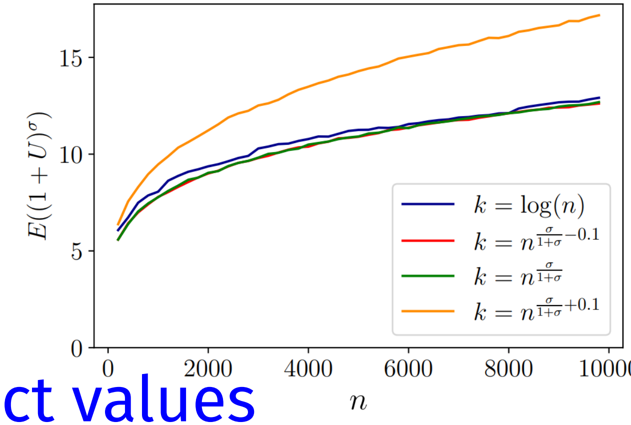


Generalized Gamma: results

U density proportional to $u^{n-1}(1+u)^{k-n\sigma}e^{-\alpha/\sigma(1+u)^\sigma}$.

Theorem. There holds $(1+U)^\sigma \sim r_n$ in L^1 with

$$r_n \asymp \begin{cases} n^{\sigma/(1+\sigma)} & \text{if } k \lesssim n^{\sigma/(1+\sigma)} \\ k & \text{if } k \gg n^{\sigma/(1+\sigma)} \end{cases}$$



and $(1+U)^\sigma$ is log-concave.

distinct values

Consequence. $$d_{\mathrm{WoW}}(\tilde{\mu}^{1,*}, \tilde{\mu}^{2,*}) \asymp \max\left(\frac{1}{n^{1/(1+\sigma)}}, \frac{k}{n}\right)$$

Gen. Gamma $\leftrightarrow$ Gamma

Generalized Gamma: results

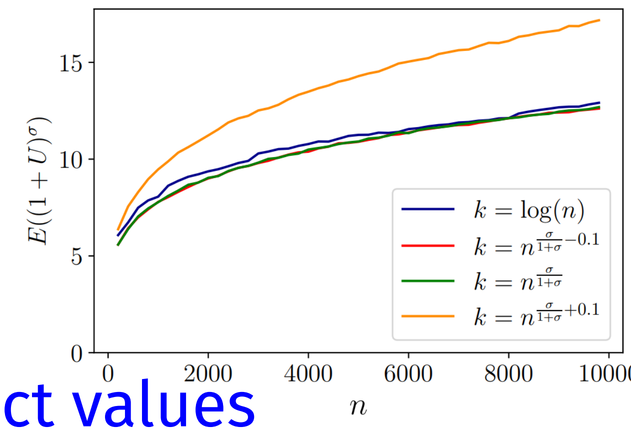
U density proportional to $u^{n-1}(1+u)^{k-n\sigma}e^{-\alpha/\sigma(1+u)^\sigma}$.

Theorem. There holds $(1+U)^\sigma \sim r_n$ in L^1 with

$$r_n \asymp \begin{cases} n^{\sigma/(1+\sigma)} & \text{if } k \lesssim n^{\sigma/(1+\sigma)} \\ k & \text{if } k \gg n^{\sigma/(1+\sigma)} \end{cases}$$

and $(1+U)^\sigma$ is log-concave.

distinct values



Consequence. $d_{\mathrm{WoW}}(\tilde\mu^{1,*}, \tilde\mu^{2,*}) \asymp \max\left(\frac{1}{n^{1/(1+\sigma)}}, \frac{k}{n}\right)$

Gen. Gamma $\leftrightarrow$ Gamma

- Merging if $k \ll n$
- Merging rate depends on k, n and σ.
- **Different** outcomes if n small or k large.

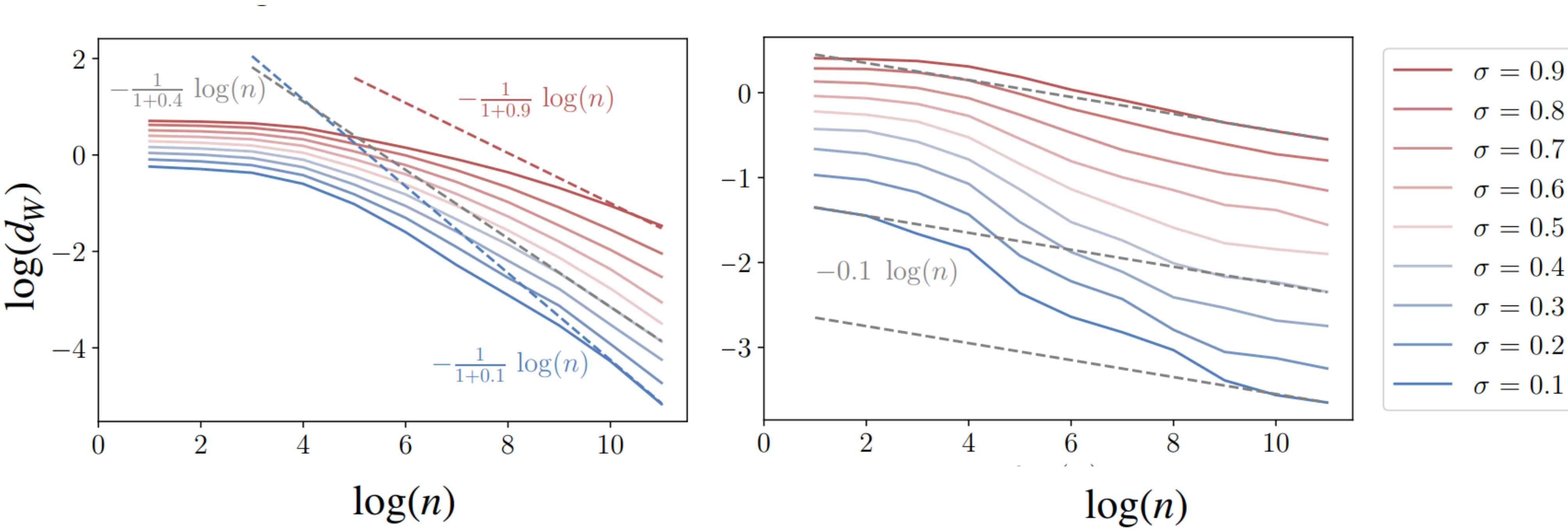
Generalized Gamma: simulations

When $k \ll n^{\sigma/(1+\sigma)}$

(Ground truth: Dirichlet)

When $k \gg n^{\sigma/(1+\sigma)}$

(Ground truth: Pitman-Yor)



Conclusion

What we did

- Distance between CRMs and quantitative study of merging of opinions.
- (Previous work: use of the distance to measure dependence)

Conclusion

What we did

- Distance between CRMs and quantitative study of merging of opinions.
- (Previous work: use of the distance to measure dependence)

To be explored

- Other use of our distance in BNP,
- What is a good distance between random measures?

Catalano, Lavenant, Lijoi and Prünster (2021). A Wasserstein index of dependence for random measures .

Conclusion

What we did

- Distance between CRMs and quantitative study of merging of opinions.
- (Previous work: use of the distance to measure dependence)

To be explored

- Other use of our distance in BNP,
- What is a good distance between random measures?

Thank you for your attention