Merging Rate of Opinions via Optimal Transport on Random Measures

M. Catalano, H. Lavenant, Bernoulli, 2026

Gianluca Covini

Presentation for the exam of *Statistical Theory*

Motivation

Let $(X_n)_{n\geq 1}$ be a sequence of exchangeable observations. We model them as

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

(Philosophical) question

How does the subjective opinion of the statistician (the choice of the prior π) influence the final inference?

Motivation

Let $(X_n)_{n\geq 1}$ be a sequence of exchangeable observations. We model them as

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

(Philosophical) question

How does the subjective opinion of the statistician (the choice of the prior π) influence the final inference?

Starting from different priors π^1, π^2, does seeing the same data induce **merging of opinions**?

Merging of opinions

Convergence of opinions after seeing the same data?

(Mathematical) question

Given $\theta^1 \sim \pi^1$ and $\theta^2 \sim \pi^2$, do we have that, as $n \to \infty$,

$$\text{distance}(\mathcal{L}(\theta^1 \mid X_1, \ldots, X_n), \mathcal{L}(\theta^2 \mid X_1, \ldots, X_n))$$

Merging of opinions

Convergence of opinions after seeing the same data?

(Mathematical) question

Given $\theta^1 \sim \pi^1$ and $\theta^2 \sim \pi^2$, do we have that, as $n \to \infty$,

$$\text{distance}(\mathcal{L}(\theta^1 \mid X_1, \ldots, X_n), \mathcal{L}(\theta^2 \mid X_1, \ldots, X_n))$$

$\to 0$?

Merging of opinions

Convergence of opinions after seeing the same data?

(Mathematical) question

Given $\theta^1 \sim \pi^1$ and $\theta^2 \sim \pi^2$, do we have that, as $n \to \infty$,

$$\text{distance}(\mathcal{L}(\theta^1 \mid X_1, \dots, X_n), \mathcal{L}(\theta^2 \mid X_1, \dots, X_n))$$

$\to 0$?

at which rate?

Merging of opinions

Convergence of opinions after seeing the same data?

(Mathematical) question

Given $\theta^1 \sim \pi^1$ and $\theta^2 \sim \pi^2$, do we have that, as $n \to \infty$,

$$\text{distance}(\mathcal{L}(\theta^1 \mid X_1, \dots, X_n), \mathcal{L}(\theta^2 \mid X_1, \dots, X_n))$$

$\to 0$?

at which rate?

monotonically?

Analogy with posterior consistency

$\rightarrow$ Similar in spirit to **posterior consistency** and **posterior contraction rates**

Analogy with posterior consistency

- → Similar in spirit to **posterior consistency** and **posterior contraction rates**
- → No convergence to a real distribution P_{θ_0}, but convergence of two distributions to each other.
- → No well-specified true parameter θ_0 is assumed; the data sequence can be arbitrary!

Analogy with posterior consistency

- → Similar in spirit to **posterior consistency** and **posterior contraction rates**
- → No convergence to a real distribution P_{θ_0}, but convergence of two distributions to each other.
- → No well-specified true parameter θ_0 is assumed; the data sequence can be arbitrary!
- → In general, posterior consistency entails merging of opinions but no information on rates.

Outline

1. Which model?
2. Which distance?
3. How to use it?
4. Application to Dirichlet process

Which model?

Setting

We consider the **Bayesian nonparametric** setting

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

where $\tilde{p}$ is a random probability measure on $\mathbb{X}$.

Setting

In particular, we consider a **completely random measure** $\tilde{\mu}$.

$$X_1, \ldots, X_n | \tilde{\mu} \overset{\text{i.i.d.}}{\sim} \frac{\tilde{\mu}}{\tilde{\mu}(\mathbb{X})}$$

$$\tilde{\mu} \sim \text{CRM}(\nu)$$

Setting

In particular, we consider a **completely random measure** $\tilde{\mu}$.

$$X_1, \dots, X_n | \tilde{\mu} \overset{\text{i.i.d.}}{\sim} \frac{\tilde{\mu}}{\tilde{\mu}(\mathbb{X})}$$
$$\tilde{\mu} \sim \text{CRM}(\nu)$$

Random measures

$\tilde{\mu}$ is a **random measure** on $\mathbb{X}$ if it is a measurable function

$$\tilde{\mu} : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathcal{M}_B(\mathbb{X}), \mathcal{B}(\mathcal{M}_B(\mathbb{X})))$$

Setting

We consider the **nonparametric** setting

$$X_1, \ldots, X_n | \tilde{\mu} \overset{\text{i.i.d.}}{\sim} \frac{\tilde{\mu}}{\tilde{\mu}(\mathbb{X})}$$

$$\tilde{\mu} \sim \text{CRM}(\nu)$$

Completely random measures

$\tilde{\mu}$ is a **completely random measure** if

- $\tilde{\mu}$ is a random measure on $\mathbb{X}$;
- for any disjoint $A_1, \ldots, A_k \subseteq \mathbb{X}$, the random variables $\tilde{\mu}(A_1), \ldots, \tilde{\mu}(A_k)$ are independent.

Why CRMs?

$$X_1, \dots, X_n | \tilde{\mu} \overset{\text{i.i.d.}}{\sim} \frac{\tilde{\mu}}{\tilde{\mu}(\mathbb{X})}, \quad \tilde{\mu} \sim \text{CRM}(\nu)$$

Completely Random Measure

With little abuse and under well-posedness assumptions, a random measure $\tilde{\mu}$ is a CRM if it can be represented as

$$\tilde{\mu}(A) \overset{d}{=} \int_{(0,+\infty)\times A} s \mathrm{d}\tilde{\mathcal{N}}(s,x) \overset{d}{\approx} \sum_{j=0}^{N} S_j \delta_{X_j}$$

with

- $\tilde{\mathcal{N}}(s,x)$ is a Poisson r. m. with intensity $\nu \in \mathcal{M}((0,+\infty) \times \mathbb{X})$;
- $N \sim$ Poisson$(\nu((\varepsilon,+\infty) \times A))$, $(S_j, X_j) \overset{\text{i.i.d.}}{\sim} \frac{\nu_{|(\varepsilon,+\infty)\times A)}}{\nu((\varepsilon,+\infty)\times A)}$

A note on identifiability

$$X_1, \ldots, X_n | \tilde{\mu} \overset{\text{i.i.d.}}{\sim} \frac{\tilde{\mu}}{\tilde{\mu}(\mathbb{X})}, \quad \tilde{\mu} \sim \text{CRM}(\nu)$$

→ A CRM is characterized by its Lévy intensity $\nu \in \mathcal{M}((0, +\infty) \times \mathbb{X})$;

→ The canonical decomposition of ν is

$$\mathrm{d}\nu(s, x) = \overbrace{\mathrm{d}\rho_x(s)}^{\text{jumps}}\overbrace{\mathrm{d}P_0(x)}^{\text{atoms}} \quad P_0 \in \mathcal{P}(\mathbb{X}), \rho \text{ transition kernel on } \mathbb{X} \times \mathcal{B}(\mathbb{R}_+)$$

A note on identifiability

$$X_1, \dots, X_n | \tilde{\mu} \overset{\text{i.i.d.}}{\sim} \frac{\tilde{\mu}}{\tilde{\mu}(\mathbb{X})}, \qquad \tilde{\mu} \sim \text{CRM}(\nu)$$

→ A CRM is characterized by its Lévy intensity $\nu \in \mathcal{M}((0, +\infty) \times \mathbb{X})$;

→ The canonical decomposition of ν is

$$\mathrm{d}\nu(s, x) = \overbrace{\mathrm{d}\rho_x(s)}^{\text{jumps}} \overbrace{\mathrm{d}P_0(x)}^{\text{atoms}} \quad P_0 \in \mathcal{P}(\mathbb{X}), \rho \text{ transition kernel on } \mathbb{X} \times \mathcal{B}(\mathbb{R}_+)$$

Moreover, if well-defined,

$$\frac{\tilde{\mu}^1}{\tilde{\mu}^1(\mathbb{X})} \overset{d}{=} \frac{\tilde{\mu}^2}{\tilde{\mu}^2(\mathbb{X})} \quad \Longleftrightarrow \quad \tilde{\mu}^1 = \alpha \tilde{\mu}^2 \text{ for some } \alpha > 0$$

Rescaling

From now on we will consider CRMs through their rescaled representative $\tilde{\mu}_S = \frac{\tilde{\mu}}{\mathbb{E}[\tilde{\mu}(\mathbb{X})]}$.

Properties

$$\tilde{\mu} = \sum_j S_j \delta_{X_j}, \quad \mathrm{d}\nu(s,x) = \mathrm{d}\rho_x(s)\mathrm{d}P_0(x)$$

Flexibility

DP as Gamma CRM:

$$\mathrm{d}\nu(s,x) = \alpha\frac{e^{-bs}}{s}\mathrm{d}s\mathrm{d}P_0(x) \implies \tilde{\mu}(A) \sim \mathrm{Gamma}(\alpha P_0(A), b) \implies \frac{\tilde{\mu}}{\tilde{\mu}(\mathbb{X})} \sim DP(\alpha, P_0)$$

Properties

$$\tilde{\mu} = \sum_j S_j \delta_{X_j}, \quad \mathrm{d}\nu(s,x) = \mathrm{d}\rho_x(s)\mathrm{d}P_0(x)$$

Flexibility

DP as Gamma CRM:

$\mathrm{d}\nu(s,x) = \alpha \frac{e^{-bs}}{s} \mathrm{d}s \mathrm{d}P_0(x) \implies \tilde{\mu}(A) \sim \mathrm{Gamma}(\alpha P_0(A), b) \implies \frac{\tilde{\mu}}{\tilde{\mu}(\mathbb{X})} \sim DP(\alpha, P_0)$

An almost-conjugacy result

$\mathcal{L}(\tilde{\mu} | X_1 = x_1, \dots, X_n = x_n) = \mathcal{L}(\tilde{\mu}^*)$ where $\tilde{\mu}^*$ is such that

$$\tilde{\mu}^* | \tilde{\nu} \sim \mathrm{CRM}(\tilde{\nu})$$

$$\tilde{\nu}$$

$\rightarrow$ Heuristically, the posterior can be seen as a CRM with random Lévy intensity.

Which distance?

From $\mathbb{X}$ to $\mathcal{M}_B(\mathbb{X})$

We need a distance between $\mathcal{L}(\tilde{\mu}^1)$ and $\mathcal{L}(\tilde{\mu}^2)$, i.e. on $\mathcal{P}(\mathcal{M}_B(\mathbb{X}))$.

From $\mathbb{X}$ to $\mathcal{M}_B(\mathbb{X})$

We need a distance between $\mathcal{L}(\tilde{\mu}^1)$ and $\mathcal{L}(\tilde{\mu}^2)$, i.e. on $\mathcal{P}(\mathcal{M}_B(\mathbb{X}))$.

Wasserstein Distance

The **1-Wasserstein distance** is a tool used to lift distances from $(\mathbb{X}, d_{\mathbb{X}})$ to $(\mathcal{P}(\mathbb{X}), W_{d_{\mathbb{X}}})$

$$W_{d_{\mathbb{X}}}(P^1, P^2) := \min_{\pi \in \Pi(P^1, P^2)} \mathbb{E}_{(X,Y)\sim\pi}[d_{\mathbb{X}}(X, Y)]$$

From $\mathbb{X}$ to $\mathcal{M}_B(\mathbb{X})$

We need a distance between $\mathcal{L}(\tilde{\mu}^1)$ and $\mathcal{L}(\tilde{\mu}^2)$, i.e. on $\mathcal{P}(\mathcal{M}_B(\mathbb{X}))$.

Wasserstein Distance

The **1-Wasserstein distance** is a tool used to lift distances from $(\mathbb{X}, d_{\mathbb{X}})$ to $(\mathcal{P}(\mathbb{X}), W_{d_{\mathbb{X}}})$

$$W_{d_{\mathbb{X}}}(P^1, P^2) := \min_{\pi \in \Pi(P^1, P^2)} \mathbb{E}_{(X,Y)\sim\pi}[d_{\mathbb{X}}(X,Y)] \overset{\substack{dual \\ \downarrow}}{=} \sup_{f:\mathbb{X}\to\mathbb{R}} \left\{ \int_{\mathbb{X}} f \,\mathrm{d}P^2 - \int_{\mathbb{X}} f \,\mathrm{d}P^1 : f \text{ 1-Lip} \right\}$$

From $\mathbb{X}$ to $\mathcal{M}_B(\mathbb{X})$

We need a distance between $\mathcal{L}(\tilde{\mu}^1)$ and $\mathcal{L}(\tilde{\mu}^2)$, i.e. on $\mathcal{P}(\mathcal{M}_B(\mathbb{X}))$.

Wasserstein Distance

The **1-Wasserstein distance** is a tool used to lift distances from $(\mathbb{X}, d_{\mathbb{X}})$ to $(\mathcal{P}(\mathbb{X}), W_{d_{\mathbb{X}}})$

$$W_{d_{\mathbb{X}}}(P^1, P^2) := \min_{\pi \in \Pi(P^1, P^2)} \mathbb{E}_{(X,Y)\sim\pi}[d_{\mathbb{X}}(X,Y)] \overset{\overset{dual}{\downarrow}}{=} \sup_{f:\mathbb{X}\to\mathbb{R}} \left\{ \int_{\mathbb{X}} f \mathrm{d}P^2 - \int_{\mathbb{X}} f \mathrm{d}P^1 : f \text{ 1-Lip} \right\}$$

BL distance

BL distance lifts distance from $(\mathbb{X}, d_{\mathbb{X}})$ to $(\mathcal{M}_B(\mathbb{X}), \mathrm{BL})$

$$\mathrm{BL}(\mu^1, \mu^2) = \sup_{f:\mathbb{X}\to\mathbb{R}} \left\{ \int_{\mathbb{X}} f \mathrm{d}\mu^2 - \int_{\mathbb{X}} f \mathrm{d}\mu^1 : f \text{ 1-Lip and 1-bounded} \right\} \quad \mu^1, \mu^2 \in \mathcal{M}_B(\mathbb{X})$$

From $\mathcal{M}_B(\mathbb{X})$ to $\mathcal{P}(\mathcal{M}_B(\mathbb{X}))$

Lifting BL

We can lift BL on $\mathcal{M}_B(\mathbb{X})$ to $(\mathcal{P}(\mathcal{M}_B(\mathbb{X})), W_{\mathsf{BL}})$ using the 1-Wasserstein distance

$$W_{\mathsf{BL}}(\mathtt{Q}^1, \mathtt{Q}^2) := \min_{\pi \in \Pi(\mathtt{Q}^1, \mathtt{Q}^2)} \mathbb{E}_{(\tilde{\mu}^1, \tilde{\mu}^2) \sim \pi}[\mathsf{BL}(\tilde{\mu}^1, \tilde{\mu}^2)] \quad \mathtt{Q}^1, \mathtt{Q}^2 \in \mathcal{P}(\mathcal{M}_B(\mathbb{X}))$$

From $\mathcal{M}_B(\mathbb{X})$ to $\mathcal{P}(\mathcal{M}_B(\mathbb{X}))$

Lifting BL

We can lift BL on $\mathcal{M}_B(\mathbb{X})$ to $(\mathcal{P}(\mathcal{M}_B(\mathbb{X})), W_{\mathsf{BL}})$ using the 1-Wasserstein distance

$$W_{\mathsf{BL}}(\mathtt{Q}^1, \mathtt{Q}^2) := \min_{\pi \in \Pi(\mathtt{Q}^1, \mathtt{Q}^2)} \mathbb{E}_{(\tilde{\mu}^1, \tilde{\mu}^2) \sim \pi}[\mathsf{BL}(\tilde{\mu}^1, \tilde{\mu}^2)] \quad \mathtt{Q}^1, \mathtt{Q}^2 \in \mathcal{P}(\mathcal{M}_B(\mathbb{X}))$$

Properties

- On the set of laws of random measures with finite mean, W_{BL} is complete;

From $\mathcal{M}_B(\mathbb{X})$ to $\mathcal{P}(\mathcal{M}_B(\mathbb{X}))$

Lifting BL

We can lift BL on $\mathcal{M}_B(\mathbb{X})$ to $(\mathcal{P}(\mathcal{M}_B(\mathbb{X})), W_{\mathsf{BL}})$ using the 1-Wasserstein distance

$$W_{\mathsf{BL}}(\mathtt{Q}^1, \mathtt{Q}^2) := \min_{\pi \in \Pi(\mathtt{Q}^1, \mathtt{Q}^2)} \mathbb{E}_{(\tilde{\mu}^1, \tilde{\mu}^2) \sim \pi} [\mathsf{BL}(\tilde{\mu}^1, \tilde{\mu}^2)] \quad \mathtt{Q}^1, \mathtt{Q}^2 \in \mathcal{P}(\mathcal{M}_B(\mathbb{X}))$$

Properties

- On the set of laws of random measures with finite mean, W_{BL} is complete;
- $W_{\mathsf{BL}}(\mathcal{L}(\tilde{\mu}_n), \mathcal{L}(\tilde{\mu})) \to 0 \iff \tilde{\mu}_n \xrightarrow{d} \tilde{\mu}, \mathbb{E}[\tilde{\mu}_n(\mathbb{X})] \to \mathbb{E}[\tilde{\mu}(\mathbb{X})]$

From $\mathcal{M}_B(\mathbb{X})$ to $\mathcal{P}(\mathcal{M}_B(\mathbb{X}))$

Lifting BL

We can lift BL on $\mathcal{M}_B(\mathbb{X})$ to $(\mathcal{P}(\mathcal{M}_B(\mathbb{X})), W_{\mathsf{BL}})$ using the 1-Wasserstein distance

$$W_{\mathsf{BL}}(\mathtt{Q}^1, \mathtt{Q}^2) := \min_{\pi \in \Pi(\mathtt{Q}^1, \mathtt{Q}^2)} \mathbb{E}_{(\tilde{\mu}^1, \tilde{\mu}^2) \sim \pi} [\mathsf{BL}(\tilde{\mu}^1, \tilde{\mu}^2)] \quad \mathtt{Q}^1, \mathtt{Q}^2 \in \mathcal{P}(\mathcal{M}_B(\mathbb{X}))$$

Properties

- On the set of laws of random measures with finite mean, W_{BL} is complete;
- $W_{\mathsf{BL}}(\mathcal{L}(\tilde{\mu}_n), \mathcal{L}(\tilde{\mu})) \to 0 \iff \tilde{\mu}_n \xrightarrow{d} \tilde{\mu}, \mathbb{E}[\tilde{\mu}_n(\mathbb{X})] \to \mathbb{E}[\tilde{\mu}(\mathbb{X})]$
- $\mathsf{BL}(\mathbb{E}(\tilde{\mu}^1), \mathbb{E}(\tilde{\mu}^2)) \leq W_{\mathsf{BL}}(\mathcal{L}(\tilde{\mu}^1), \mathcal{L}(\tilde{\mu}^2))$

From $\mathcal{M}_B(\mathbb{X})$ to $\mathcal{P}(\mathcal{M}_B(\mathbb{X}))$

Lifting BL

We can lift BL on $\mathcal{M}_B(\mathbb{X})$ to $(\mathcal{P}(\mathcal{M}_B(\mathbb{X})), W_{\mathsf{BL}})$ using the 1-Wasserstein distance

$$W_{\mathsf{BL}}(\mathtt{Q}^1, \mathtt{Q}^2) := \min_{\pi \in \Pi(\mathtt{Q}^1, \mathtt{Q}^2)} \mathbb{E}_{(\tilde{\mu}^1, \tilde{\mu}^2) \sim \pi}[\mathsf{BL}(\tilde{\mu}^1, \tilde{\mu}^2)] \quad \mathtt{Q}^1, \mathtt{Q}^2 \in \mathcal{P}(\mathcal{M}_B(\mathbb{X}))$$

Properties

- On the set of laws of random measures with finite mean, W_{BL} is complete;
- $W_{\mathsf{BL}}(\mathcal{L}(\tilde{\mu}_n), \mathcal{L}(\tilde{\mu})) \to 0 \iff \tilde{\mu}_n \xrightarrow{d} \tilde{\mu}, \mathbb{E}[\tilde{\mu}_n(\mathbb{X})] \to \mathbb{E}[\tilde{\mu}(\mathbb{X})]$
- $\mathsf{BL}(\mathbb{E}(\tilde{\mu}^1), \mathbb{E}(\tilde{\mu}^2)) \leq W_{\mathsf{BL}}(\mathcal{L}(\tilde{\mu}^1), \mathcal{L}(\tilde{\mu}^2))$

Idea: use W_{BL} to study merging rate of opinions $\to$, i.e., study

$$W_{\mathsf{BL}}(\mathcal{L}(\tilde{\mu}^1 | X_1, \dots, X_n), \mathcal{L}(\tilde{\mu}^2 | X_1, \dots, X_n)) \text{ as } n \to \infty$$

How to use it?

Distance between CRMs

Idea 1: bound W_{BL} from above using an analytically tractable object.

Distance between CRMs

Idea 1: bound W_{BL} from above using an analytically tractable object.

Idea 2: exploit the description of CRMs through their intensity ν .

Distance between CRMs

Idea 1: bound W_{BL} from above using an analytically tractable object.

Idea 2: exploit the description of CRMs through their intensity ν .

We consider $\tilde{\mu}^1 \sim \text{CRM}(\nu^1)$ and $\tilde{\mu}^2 \sim \text{CRM}(\nu^2)$ **rescaled** with $\mathrm{d}\nu^i(s,x) = \mathrm{d}\rho^i_x(s)\mathrm{d}P^i_0(x)$.

$\to$ **Problem:** ρ^1_x and ρ^2_y may have infinite mass.

Distance between CRMs

Idea 1: bound W_{BL} from above using an analytically tractable object.

Idea 2: exploit the description of CRMs through their intensity ν .

We consider $\tilde{\mu}^1 \sim \text{CRM}(\nu^1)$ and $\tilde{\mu}^2 \sim \text{CRM}(\nu^2)$ **rescaled** with $\mathrm{d}\nu^i(s,x) = \mathrm{d}\rho^i_x(s)\mathrm{d}P^i_0(x)$.

$\rightarrow$ **Problem:** ρ^1_x and ρ^2_y may have infinite mass.

Lifted extended Wasserstein

$$W_*(\rho^1,\rho^2) := \int_0^{+\infty} |\rho^1(s,+\infty) - \rho^2(s,+\infty)|\mathrm{d}s \quad \rho^1,\rho^2 \in \mathcal{M}((0,+\infty))$$

Distance between CRMs

Idea 1: bound W_{BL} from above using an analytically tractable object.

Idea 2: exploit the description of CRMs through their intensity ν .

We consider $\tilde{\mu}^1 \sim \text{CRM}(\nu^1)$ and $\tilde{\mu}^2 \sim \text{CRM}(\nu^2)$ **rescaled** with $\mathrm{d}\nu^i(s,x) = \mathrm{d}\rho^i_x(s)\mathrm{d}P^i_0(x)$.

$\rightarrow$ **Problem:** ρ^1_x and ρ^2_y may have infinite mass.

Lifted extended Wasserstein

$$W_*(\rho^1,\rho^2) := \int_0^{+\infty} |\rho^1(s,+\infty) - \rho^2(s,+\infty)| \mathrm{d}s \quad \rho^1,\rho^2 \in \mathcal{M}((0,+\infty))$$

$$\inf_{\pi \in \Pi(P^1_0, P^2_0)} \mathbb{E}_{(X,Y)\sim\pi}[W_*(\rho^1_X, \rho^2_Y)]$$

Distance between CRMs

Idea 1: bound W_{BL} from above using an analytically tractable object.

Idea 2: exploit the description of CRMs through their intensity ν .

We consider $\tilde{\mu}^1 \sim \mathsf{CRM}(\nu^1)$ and $\tilde{\mu}^2 \sim \mathsf{CRM}(\nu^2)$ **rescaled** with $\mathrm{d}\nu^i(s,x) = \mathrm{d}\rho^i_x(s)\mathrm{d}P^i_0(x)$.

→ **Problem:** ρ^1_x and ρ^2_y may have infinite mass.

Lifted extended Wasserstein

$$W_*(\rho^1,\rho^2) := \int_0^{+\infty} |\rho^1(s,+\infty) - \rho^2(s,+\infty)|\mathrm{d}s \quad \rho^1,\rho^2 \in \mathcal{M}((0,+\infty))$$

$$\inf_{\pi\in\Pi(P^1_0,P^2_0)} \mathbb{E}_{(X,Y)\sim\pi}[W_*(\rho^1_X,\rho^2_Y)]$$

$$\mathsf{AD}(\nu^1,\nu^2) = \inf_{\pi\in\Pi(P^1_0,P^2_0)} \mathbb{E}_{(X,Y)\sim\pi}\left[(d_{\mathbb{X}}(X,Y)\wedge 2) + W_*(\rho^1_X,\rho^2_Y)\right]$$

Upper bound

Adapted extended Wasserstein

$$\mathrm{AD}(\nu^1, \nu^2) = \inf_{\pi \in \Pi(P_0^1, P_0^2)} \mathbb{E}_{(X,Y)\sim\pi} \left[(d_{\mathbb{X}}(X,Y) \wedge 2) + W_*(\rho_X^1, \rho_Y^2) \right]$$

Upper bound

For CRMs $\tilde{\mu}^i$ with finite mean,

$$W_{\mathrm{BL}}(\mathcal{L}(\tilde{\mu}^1), \mathcal{L}(\tilde{\mu}^2)) \leq \mathrm{AD}(\nu^1, \nu^2)$$

Posteriors of CRMs can be seen as CRMs with random intensity $\tilde{\nu}^i$ for which

$$W_{\mathrm{BL}}(\mathcal{L}(\tilde{\mu}^1), \mathcal{L}(\tilde{\mu}^2)) \leq W_{\mathrm{AD}}(\mathcal{L}(\tilde{\nu}^1), \mathcal{L}(\tilde{\nu}^2))$$

Application to Dirichlet process

Dirichlet process

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

$$\tilde{p} \sim DP(P_0, \alpha)$$

Question

Given two Bayesians who model the data as DP but with different parameters, do their posterior beliefs converge after they see the same data?

Dirichlet process

Model: $X_1, \ldots, X_n | \tilde{\mu} \overset{\text{i.i.d.}}{\sim} \dfrac{\tilde{\mu}}{\tilde{\mu}(\mathbb{X})}$

$$\tilde{\mu} \sim \text{CRM}(\nu) = \text{GammaCRM}(\alpha, P_0)$$

by choosing ν s.t. $\mathrm{d}\nu(s, x) = \alpha \dfrac{e^{-bs}}{s} \mathrm{d}s \mathrm{d}P_0(x)$

Note: (After rescaling) the posterior is a Gamma CRM with ν^* of the form

$$\rho^*(s) = (\alpha + n) \frac{e^{-(\alpha+n)s}}{s}$$

$$P_0^* = \frac{\alpha}{\alpha + n} P_0 + \frac{1}{\alpha + n} \sum_{j=1}^{n} \delta_{x_j}$$

Dirichlet process

$$\textbf{Model: } X_1, \dots, X_n | \tilde{\mu} \overset{\text{i.i.d.}}{\sim} \frac{\tilde{\mu}}{\tilde{\mu}(\mathbb{X})}$$

$$\tilde{\mu} \sim \text{CRM}(\nu) = \text{GammaCRM}(\alpha, P_0)$$

$$\text{by choosing } \nu \text{ s.t. } \mathrm{d}\nu(s, x) = \alpha \frac{e^{-bs}}{s} \mathrm{d}s \mathrm{d}P_0(x)$$

Note: (After rescaling) the posterior is a Gamma CRM with ν^* of the form

$$\rho^*(s) = (\alpha + n) \frac{e^{-(\alpha+n)s}}{s}$$

$$P_0^* = \frac{\alpha}{\alpha + n} P_0 + \frac{1}{\alpha + n} \sum_{j=1}^{n} \delta_{x_j}$$

$\rightarrow$ What is the impact of α and P_0 on merging of opinions?

Merging of opinions

Theorem

Let $\tilde{\mu}^1, \tilde{\mu}^2$ Gamma CRMs with parameters (α_i, P_0^i), $i = 1, 2$ and $\tilde{\mu}^{1*}, \tilde{\mu}^{2*}$ the corresponding rescaled posterior after n observations. Then,

$$W_{\mathsf{BL}}(\mathcal{L}(\tilde{\mu}^{1*}), \mathcal{L}(\tilde{\mu}^{2*})) \lesssim \frac{1}{n}$$

Proof sketch

Let $\tilde{\mu}^{1*}, \tilde{\mu}^{2*}$ be the posterior and ν^{1*}, ν^{2*} the corresponding intensities.

Proof sketch

Let $\tilde{\mu}^{1*}, \tilde{\mu}^{2*}$ be the posterior and ν^{1*}, ν^{2*} the corresponding intensities.

Idea: we study a tight bound on $\mathsf{AD}(\nu^{1*}, \nu^{2*})$ and then use $W_{\mathsf{BL}}(\mathcal{L}(\tilde{\mu}^{1*}), \mathcal{L}(\tilde{\mu}^{2*})) \leq \mathsf{AD}(\nu^{1*}, \nu^{2*})$.

Proof sketch

Let $\tilde{\mu}^{1*}, \tilde{\mu}^{2*}$ be the posterior and ν^{1*}, ν^{2*} the corresponding intensities.

Idea: we study a tight bound on $\mathrm{AD}(\nu^{1*}, \nu^{2*})$ and then use $W_{\mathrm{BL}}(\mathcal{L}(\tilde{\mu}^{1*}), \mathcal{L}(\tilde{\mu}^{2*})) \leq \mathrm{AD}(\nu^{1*}, \nu^{2*})$.

Step 1

Decompose $\mathrm{AD}(\nu^{1*}, \nu^{2*})$ as

$$\mathrm{AD}(\nu^{1*}, \nu^{2*}) = \overbrace{\mathcal{J}(\alpha_1, \alpha_2, n)}^{\text{jump term}} + \overbrace{\mathcal{A}(\alpha_1, P_0^1, \alpha_2, P_0^2, n)}^{\text{atom term}}$$

- $\mathcal{J}$ measures discrepancy between the jumps of the posterior;
- $\mathcal{A}$ is the (truncated) Wasserstein distance between the predictive distributions $\mathcal{L}(X_{n+1} | X^{(n)})$.

Proof sketch

$$\mathrm{AD}(\nu^{1*}, \nu^{2*}) = \mathcal{J} + \mathcal{A}$$

Step 2

$$\mathcal{A} = W_{d_{\mathbb{X},t}}\left(\frac{\alpha_1}{\alpha_1+n}P_0^1 + \frac{1}{\alpha_1+n}\sum_{i=1}^n \delta_{x_i}, \frac{\alpha_2}{\alpha_2+n}P_0^2 + \frac{1}{\alpha_2+n}\sum_{i=1}^n \delta_{x_i}\right)$$

Then, using convexity of the Wasserstein distance,

$$\mathcal{A} \leq \frac{\alpha_1}{\alpha_1+n}W_{d_{\mathbb{X},t}}(P_0^1, P_0^2) + \left(\frac{\alpha_2}{\alpha_2+n} - \frac{\alpha_1}{\alpha_1+n}\right)W_{d_{\mathbb{X},t}}\left(\frac{1}{n}\sum_{i=1}^n \delta_{x_i}, P_0^2\right) \lesssim \frac{1}{n}$$

Proof sketch

$$\mathsf{AD}(\nu^{1*}, \nu^{2*}) = \mathcal{J} + \mathcal{A}$$

Step 3

$$\begin{aligned}
\mathcal{J} &= W_*\left((\alpha_1 + n)\frac{e^{-(\alpha_1+n)s}}{s}\mathrm{d}s, (\alpha_2 + n)\frac{e^{-(\alpha_2+n)s}}{s}\mathrm{d}s\right) \\
&= \int_0^{+\infty} |(\alpha_1 + n)\Gamma(0, (\alpha_1 + n)s) - (\alpha_2 + n)\Gamma(0, (\alpha_2 + n)s)|\mathrm{d}s
\end{aligned}$$

It holds,

$$\mathcal{J} = \left(\int_0^{+\infty} |\Gamma(0, t) - e^{-t}|\mathrm{d}t\right)\frac{\alpha_2 - \alpha_1}{n} + o\left(\frac{1}{n}\right)$$

$$\implies \mathcal{J} \asymp \frac{1}{n} \quad (\text{if } \alpha_1 \neq \alpha_2)$$

Proof sketch

Therefore...

It follows

$$\mathsf{AD}(\nu^{1*}, \nu^{2*}) = \mathcal{J} + \mathcal{A} \asymp \frac{1}{n}$$

and therefore

$$W_{\mathsf{BL}}(\mathcal{L}(\tilde{\mu}^{1*}), \mathcal{L}(\tilde{\mu}^{2*})) \leq \mathsf{AD}(\nu^{1*}, \nu^{2*}) \implies W_{\mathsf{BL}}(\mathcal{L}(\tilde{\mu}^{1*}), \mathcal{L}(\tilde{\mu}^{2*})) \lesssim \frac{1}{n}$$

□

Proof sketch

Therefore...

It follows

$$\mathsf{AD}(\nu^{1*}, \nu^{2*}) = \mathcal{J} + \mathcal{A} \asymp \frac{1}{n}$$

and therefore

$$W_{\mathsf{BL}}(\mathcal{L}(\tilde{\mu}^{1*}), \mathcal{L}(\tilde{\mu}^{2*})) \leq \mathsf{AD}(\nu^{1*}, \nu^{2*}) \implies W_{\mathsf{BL}}(\mathcal{L}(\tilde{\mu}^{1*}), \mathcal{L}(\tilde{\mu}^{2*})) \lesssim \frac{1}{n}$$

□

Non-asymptotic results

- If $\alpha_1 = \alpha_2$, AD decrease is monotone;
- If $P_0^1 = P_0^2$, then AD can increase before merging of opinions.

Conclusion

Summary

$$X_1, \ldots, X_n | \tilde{\mu} \overset{\text{i.i.d.}}{\sim} \frac{\tilde{\mu}}{\tilde{\mu}(\mathbb{X})}, \quad \tilde{\mu} \sim \text{CRM}(\nu)$$

Conclusion

Summary

$$X_1, \dots, X_n | \tilde{\mu} \overset{\text{i.i.d.}}{\sim} \frac{\tilde{\mu}}{\tilde{\mu}(\mathbb{X})}, \quad \tilde{\mu} \sim \mathrm{CRM}(\nu)$$

$\rightarrow$ Distance between laws of CRMs:

$$W_{\mathrm{BL}}(\mathtt{Q}^1, \mathtt{Q}^2) := \min_{\pi \in \Pi(\mathtt{Q}^1, \mathtt{Q}^2)} \mathbb{E}_{(\tilde{\mu}^1, \tilde{\mu}^2) \sim \pi}[\mathrm{BL}(\tilde{\mu}^1, \tilde{\mu}^2)] \quad \mathtt{Q}^1, \mathtt{Q}^2 \in \mathcal{P}(\mathcal{M}_B(\mathbb{X}))$$

Conclusion

Summary

$$X_1, \dots, X_n | \tilde{\mu} \overset{\text{i.i.d.}}{\sim} \frac{\tilde{\mu}}{\tilde{\mu}(\mathbb{X})}, \quad \tilde{\mu} \sim \mathsf{CRM}(\nu)$$

→ Distance between laws of CRMs:

$$W_{\mathsf{BL}}(\mathtt{Q}^1, \mathtt{Q}^2) := \min_{\pi \in \Pi(\mathtt{Q}^1, \mathtt{Q}^2)} \mathbb{E}_{(\tilde{\mu}^1, \tilde{\mu}^2) \sim \pi}[\mathsf{BL}(\tilde{\mu}^1, \tilde{\mu}^2)] \quad \mathtt{Q}^1, \mathtt{Q}^2 \in \mathcal{P}(\mathcal{M}_B(\mathbb{X}))$$

→ Use W_{BL} to measure the merging of opinions between posteriors as $n \to +\infty$.

Conclusion

Summary

$$X_1, \dots, X_n | \tilde{\mu} \overset{\text{i.i.d.}}{\sim} \frac{\tilde{\mu}}{\tilde{\mu}(\mathbb{X})}, \quad \tilde{\mu} \sim \mathsf{CRM}(\nu)$$

→ Distance between laws of CRMs:

$$W_{\mathsf{BL}}(\mathtt{Q}^1, \mathtt{Q}^2) := \min_{\pi \in \Pi(\mathtt{Q}^1, \mathtt{Q}^2)} \mathbb{E}_{(\tilde{\mu}^1, \tilde{\mu}^2) \sim \pi}[\mathsf{BL}(\tilde{\mu}^1, \tilde{\mu}^2)] \quad \mathtt{Q}^1, \mathtt{Q}^2 \in \mathcal{P}(\mathcal{M}_B(\mathbb{X}))$$

→ Use W_{BL} to measure the merging of opinions between posteriors as $n \to +\infty$.

→ Bound by tractable discrepancy between intensities: $W_{\mathsf{BL}}(\mathcal{L}(\tilde{\mu}^1), \mathcal{L}(\tilde{\mu}^2)) \leq \mathsf{AD}(\nu^1, \nu^2)$

Conclusion

Summary

$$X_1, \dots, X_n | \tilde{\mu} \overset{\text{i.i.d.}}{\sim} \frac{\tilde{\mu}}{\tilde{\mu}(\mathbb{X})}, \quad \tilde{\mu} \sim \mathrm{CRM}(\nu)$$

→ Distance between laws of CRMs:

$$W_{\mathrm{BL}}(\mathtt{Q}^1, \mathtt{Q}^2) := \min_{\pi \in \Pi(\mathtt{Q}^1, \mathtt{Q}^2)} \mathbb{E}_{(\tilde{\mu}^1, \tilde{\mu}^2) \sim \pi}[\mathrm{BL}(\tilde{\mu}^1, \tilde{\mu}^2)] \quad \mathtt{Q}^1, \mathtt{Q}^2 \in \mathcal{P}(\mathcal{M}_B(\mathbb{X}))$$

→ Use W_{BL} to measure the merging of opinions between posteriors as $n \to +\infty$.

→ Bound by tractable discrepancy between intensities: $W_{\mathrm{BL}}(\mathcal{L}(\tilde{\mu}^1), \mathcal{L}(\tilde{\mu}^2)) \leq \mathrm{AD}(\nu^1, \nu^2)$

→ For the Dirichlet process prior: $W_{\mathrm{BL}}(\mathcal{L}(\tilde{\mu}^{1*}), \mathcal{L}(\tilde{\mu}^{2*})) \lesssim 1/n$

Conclusion

Summary

$$X_1, \dots, X_n | \tilde{\mu} \overset{\text{i.i.d.}}{\sim} \frac{\tilde{\mu}}{\tilde{\mu}(\mathbb{X})}, \quad \tilde{\mu} \sim \mathrm{CRM}(\nu)$$

→ Distance between laws of CRMs:

$$W_{\mathrm{BL}}(\mathtt{Q}^1, \mathtt{Q}^2) := \min_{\pi \in \Pi(\mathtt{Q}^1, \mathtt{Q}^2)} \mathbb{E}_{(\tilde{\mu}^1, \tilde{\mu}^2) \sim \pi}[\mathrm{BL}(\tilde{\mu}^1, \tilde{\mu}^2)] \quad \mathtt{Q}^1, \mathtt{Q}^2 \in \mathcal{P}(\mathcal{M}_B(\mathbb{X}))$$

→ Use W_{BL} to measure the merging of opinions between posteriors as $n \to +\infty$.

→ Bound by tractable discrepancy between intensities: $W_{\mathrm{BL}}(\mathcal{L}(\tilde{\mu}^1), \mathcal{L}(\tilde{\mu}^2)) \le \mathrm{AD}(\nu^1, \nu^2)$

→ For the Dirichlet process prior: $W_{\mathrm{BL}}(\mathcal{L}(\tilde{\mu}^{1*}), \mathcal{L}(\tilde{\mu}^{2*})) \lesssim 1/n$

Thank you for your attention!

Generalized Gamma Processes

Definition

Given $\alpha > 0, \sigma \in (0,1)$ and $P_0 \in \mathcal{P}(\mathbb{X})$, a *generalized gamma CRM* is the CRM with Lévy intensity

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

Theorem

Given $\tilde{\mu}^1$ generalized gamma CRM with parameters $(\alpha^1, \sigma, P_0^1)$ and $\tilde{\mu}^2 \sim \mathrm{GammaCRM}(\alpha^2, P_0^2)$ then, for $k = k(n) \ll n$,

$$W_{\mathrm{BL}}(\mathcal{L}(\tilde{\mu}^{1*}), \mathcal{L}(\tilde{\mu}^{2*})) \lesssim \max\left(n^{-1/(1+\sigma)}, k/n\right)$$

where $k = k(n)$ is the number of distinct observations.

Generalized Gamma Processes

Definition

Given $\alpha > 0, \sigma \in (0,1)$ and $P_0 \in \mathcal{P}(\mathbb{X})$, a *generalized gamma CRM* is the CRM with Lévy intensity

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

Proposition

Given $\tilde{\mu}^1$ generalized gamma CRM with parameters $(\alpha^1, \sigma, P_0^1)$ and $\tilde{\mu}^2 \sim \mathrm{GammaCRM}(\alpha^2, P_0^2)$. Let also $x_i \overset{\text{i.i.d.}}{\sim} P$, with P continuous distribution. Then, as $n \to +\infty$, for a.e. realization of $(x_n)_{n\geq 1}$,

$$W_{\mathrm{AD}}(\mathcal{L}(\tilde{\nu}^{1*}), \mathcal{L}(\nu^{2*})) \to \sigma \left(\int_0^{+\infty} \left| \frac{\Gamma(-\sigma,t)}{\Gamma(1-\sigma)} - \Gamma(0,t) \right| \mathrm{d}t + W_{d_{\mathbb{X}} \wedge 2}(P_0, P) \right)$$

Almost-conjugacy of CRMs

Theorem [James et al. (2009)]

Let $\tilde{\mu}$ a CRM with Lévy intensity $\mathrm{d}\nu(s,x) = \mathrm{d}\rho_x(s)\mathrm{d}P_0(x)$. Consider $U = U_n$ latent variable with p.d.f. $f_U(u) \propto u^{n-1}e^{-\psi(u)} \prod_{i=1}^k \tau_{n_i|x_i^*}(u)$ where

$$\psi(u) = \int_{(0,+\infty)\times\mathbb{X}} (1-e^{us})\mathrm{d}\nu(s,x), \quad \tau_{m|x}(u) = \int_{(0,+\infty)} e^{-us}s^m \mathrm{d}\rho_x(s)$$

Then,

$$\tilde{\mu}^* \mid U \stackrel{d}{=} \tilde{\mu}_U + \sum_{i=1}^k J_i^U \delta_{x_i^*}$$

with $\tilde{\mu}_U$ is a CRM with $\mathrm{d}\nu(s,x) = e^{-Us}\mathrm{d}\nu(s,x)$, J_i^U random jump with p.d.f. depending on U and

$$\mathcal{L}(\tilde{\mu} \mid X_1,\dots,X_n) = \mathcal{L}(\tilde{\mu}^*).$$