Measuring dependence in the Wasserstein distance for Bayesian nonparametric models

Marta Catalano

University of Torino

One World YoungStatS - 21 April 2021.

Bayesian nonparametrics

$$X_1, \dots, X_n \mid \tilde P \overset{\text{iid}}{\sim} \tilde P \qquad \tilde P \sim Q$$

- Q prior distribution of unknown parameter $\tilde P \in \mathbb{P}_{\mathbb{X}}$ with large support ⇒ flexible inference through $\tilde P \mid X_1, \dots, X_n$
- de Finetti theorem ⇔ exchangeable observations $(X_1, X_2, X_3) \overset{\text{d}}{=} (X_2, X_3, X_1)$
- Bayesian homogeneity assumption "equivalent" to iid observations

Covariates introduce heterogeneity!

Categorical covariate with d=2 groups (for simplicity)

$$X_1, \dots, X_n, Y_1, \dots, Y_m \mid (\tilde P_1, \tilde P_2) \sim \tilde P_1^n \times \tilde P_2^m \qquad (\tilde P_1, \tilde P_2) \sim Q$$

- partially exchangeable observations $(X_1, Y_1, X_2) \overset{\text{d}}{=} (X_2, Y_1, X_1)$

Dependence of $(\tilde P_1, \tilde P_2)$ regulates borrowing of information across groups

$\tilde P_1 \perp \tilde P_2$ ⇒ No borrowing & independent observations

$\tilde P_1 = \tilde P_2$ a.s. ⇒ Maximal borrowing & exchangeable observations

Measuring dependence

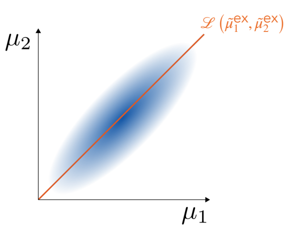
[C., Lijoi, Prünster (2021). AoS, to appear.]

$$(\tilde P_1, \tilde P_2) = (T(\tilde\mu_1), T(\tilde\mu_2)) \text{ with } \tilde\mu_1 \overset{\text{d}}{=} \tilde\mu_2$$

- $(\tilde\mu_1, \tilde\mu_2)$ random measures with ⊥ increments (CRVs) ⇒ bivariate Lévy measure
- T component-wise transformation (normalization, exponential, kernel mixture)

Goal: Measure dependence

1) generic wrt T
2) more than two groups
3) Beyond state-of-the-art $\mathrm{Corr}(\tilde P_1(A), \tilde P_2(A))$



$\tilde\mu_1 = \tilde\mu_2$ a.s. ⇒ Exchangeability

Idea: measure dependence via distance on CRVs

$$d\left(\mathscr{L}(\tilde\mu_1, \tilde\mu_2), \mathscr{L}(\tilde\mu_1^{\text{ex}}, \tilde\mu_2^{\text{ex}})\right) = 0 \iff \tilde\mu_1 = \tilde\mu_2 \text{ a.s.}$$

Results

1. Define 1st informative & tractable distance on CRVs
2. Tight bounds in terms of the Lévy intensities
3. Measure dependence for well-known CRVs in terms of hyperparameters

Distance between CRVs

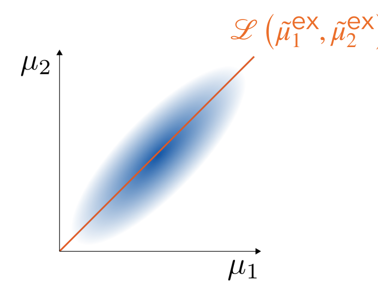
$$d_W\left(\mathscr{L}\begin{pmatrix}\tilde\mu_1\\ \tilde\mu_2\end{pmatrix}, \mathscr{L}\begin{pmatrix}\tilde\mu_1^{\text{ex}}\\ \tilde\mu_2^{\text{ex}}\end{pmatrix}\right) = \sup_{A \in \mathscr{X}} W\left(\mathscr{L}\begin{pmatrix}\tilde\mu_1(A)\\ \tilde\mu_2(A)\end{pmatrix}, \mathscr{L}\begin{pmatrix}\tilde\mu_1^{\text{ex}}(A)\\ \tilde\mu_2^{\text{ex}}(A)\end{pmatrix}\right)$$

where W is the Wasserstein distance (of order 2) between probabilities in $\mathbb{R}^2$.

$$W(P, Q)^2 = \inf_{(X,Y):\, X \overset{\text{d}}{=} P \,\&\, Y \overset{\text{d}}{=} Q} \mathbb{E}\|X - Y\|^2 = \inf_{\Gamma \in C(P,Q)} \int \|x - y\|^2\, \Gamma(dx, dy)$$

Under mild conditions ∃ an optimal coupling of the form $(X, Y) = (X, T(X))$.

- Pros: takes into account the geometry of the space ⇒ Ideal for distributions with different support



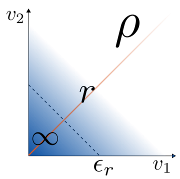
- Cons: difficult to evaluate in $\mathbb{R}^2$:

1. No explicit expression for OT map T
2. If we find T, (difficult) bivariate integral

Multivariate Lévy intensities

Result. Find a tight bound in terms of the Lévy intensities

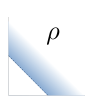
For simplicity we focus on homogeneous $\nu(dv_1, dv_2, dx) = aP_0(dx)\rho(dv_1, dv_2)$

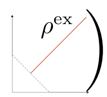


- Exchangeability if Support(ρ) $\subset \{(v_1, v_2) : v_1 = v_2\}$
- In BNP: infinite mass around the origin
- If ρ has finite mass r, $(\tilde\mu_1(A), \tilde\mu_2(A))$ has a compound Poisson distribution

Idea: one can approximate any CRV with a compound Poisson distribution by removing a neighborhood of the origin

Theorem (Bound in terms of Lévy intensity)

$$d_W\left(\begin{pmatrix}\tilde\mu_1\\ \tilde\mu_2\end{pmatrix}, \begin{pmatrix}\tilde\mu_1^{\text{ex}}\\ \tilde\mu_2^{\text{ex}}\end{pmatrix}\right) \leq \lim_{r\to+\infty} \sqrt{r}\, W\Bigg(\frac{1}{r}$$


$$, \frac{1}{r}$$


$$\Bigg)$$

Compound Random Measures

[Griffin & Leisen (2017), JRSSB]

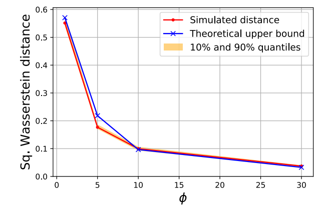
$$\begin{pmatrix}\tilde\mu_1\\ \tilde\mu_2\end{pmatrix} = \sum_{i=1}^{+\infty} \begin{pmatrix}m_{1,i}\\ m_{2,i}\end{pmatrix} J_i\, \delta_{X_i}$$

- $\sum_{i=1}^{+\infty} J_i\, \delta_{X_i} \sim$ CRM(ν_0) $\qquad \nu_0(dv, dx) = a\, P_0(dx)\,(1-v)^{\phi-1} v^{-1} \mathbb{1}_{(0,1)}(v)\, dv$
- $(m_{i,1}, m_{i,2}) \sim h \qquad h = f \otimes f, \qquad f = \text{gamma}(\phi, 1)$

Marginally, $\tilde\mu_1, \tilde\mu_2$ are gamma CRMs & ϕ only adjusts for dependence (how?)

Theorem (Bound for compound random measures)

$$d_W\left(\begin{pmatrix}\tilde\mu_1\\ \tilde\mu_2\end{pmatrix}, \begin{pmatrix}\tilde\mu_1^{\text{ex}}\\ \tilde\mu_2^{\text{ex}}\end{pmatrix}\right)^2 \leq 4\, a\, K(\phi)$$



Current developments

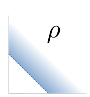
[C., Lavenant, Lijoi, Prünster]

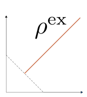
Recap [C., Lijoi, Prünster (2021). AoS, to appear.]

$$d_W\left(\begin{pmatrix}\tilde\mu_1\\ \tilde\mu_2\end{pmatrix}, \begin{pmatrix}\tilde\mu_1^{\text{ex}}\\ \tilde\mu_2^{\text{ex}}\end{pmatrix}\right) = \sup_{A\in\mathcal{X}} W\left(\mathscr{L}\begin{pmatrix}\tilde\mu_1(A)\\ \tilde\mu_2(A)\end{pmatrix}, \mathscr{L}\begin{pmatrix}\tilde\mu_1^{\text{ex}}(A)\\ \tilde\mu_2^{\text{ex}}(A)\end{pmatrix}\right)$$

$\Rightarrow$ tractable principled measure of dependence for generic CRV-BNP models

Current research [C., Lavenant, Lijoi, Prünster. Work in progress]

$$\lim_{r\to+\infty} \sqrt{r}\, W\Bigg(\frac{1}{r}$$


$$, \frac{1}{r}$$


$$\Bigg) \geq d_W\left(\begin{pmatrix}\tilde\mu_1\\ \tilde\mu_2\end{pmatrix}, \begin{pmatrix}\tilde\mu_1^{\text{ex}}\\ \tilde\mu_2^{\text{ex}}\end{pmatrix}\right)$$

Coincides with $d_W(\rho, \rho^{\text{ex}})$ between measures with ∞ mass! [Figalli & Gigli, 2009]

Wasserstein Index of Dependence in [0,1]

$$\text{WID}(\tilde P_1, \tilde P_2) = 1 - \frac{d_W(\rho, \rho^{\text{ex}})}{d_W(\rho^{\text{ind}}, \rho^{\text{ex}})}$$

$0 \iff$ independence

$1 \iff$ exchangeability

References

- C., Lijoi, Prünster (2020). Approximation of Bayesian models for time-to-event data. EJS.
- C., Lijoi, Prünster (2021). Measuring dependence in the Wasserstein distance for Bayesian nonparametric models. Aos, to appear.
- C., Lavenant, Lijoi, Prünster (2021+). Wasserstein index of dependence. Work in progress.
- Epifani & Lijoi (2010). Nonparametric priors for vectors of survival functions. Stat. Sinica.
- Figalli & Gigli (2009). A new transportation distance between non-negative measures, with applications to gradients flows with Dirichlet boundary conditions. Journal de Mathématiques Pures et Appliquées
- Griffin & Leisen (2017). Compound random measures and their use in Bayesian non-parametrics. JRSSB
- Lijoi, Nipoti & Prünster (2014). Bayesian inference with dependent normalized completely random measures. Bernoulli.
- Mariucci & Reiß (2018). Wasserstein and total variation distance between marginals of Lévy processes. EJS.
- Villani (2008). Optimal Transport: Old and New. Springer Berlin Heidelberg.