Smart Statistics for Smart Applications

Book of Short Papers SIS2019

SIS 2019
SMART STATISTICS FOR SMART APPLICATIONS

SIS Società Italiana di Statistica

Editors: Giuseppe Arbia, Stefano Peluso,
Alessia Pini and Giulia Rivellini

Published by Pearson

www.pearson.com

Giugno 2019 ISBN 9788891915108

Bayesian Model Comparison based on Wasserstein Distances

Confronto di Modelli Bayesiani tramite Distanze di Wasserstein

Marta Catalano, Antonio Lijoi and Igor Prünster

Abstract Exchangeable processes are extensively used in Bayesian nonparametrics to model exchangeable data. Most common approaches assign a law to the process through the specification of a random measure. When two processes only differ in the law of the random measure, a distance between random measures provides a natural way to compare them. In this work we propose one by relying on the Wasserstein distance. Moreover, we overcome the analytical difficulties of evaluating the distance by developing sharp upper and lower bounds. The specialization of these bounds to Gamma random measures provides the exact value of the Wasserstein distance in terms of the Kolmogorov distance between the base measures. The results are based on a forthcoming work in collaboration with A. Lijoi and I. Prünster.

Abstract *I processi scambiabili sono usati di frequente per modellare dati scambiabili. Nella maggior parte dei casi la legge del processo richiede la specificazione di una misura aleatoria. Quando due sequenze scambiabili differiscono solamente nella distribuzione delle misure, la valutazione di una distanza tra misure aleatorie fornisce un modo naturale di metterle a confronto. In questo lavoro ne proponiamo una basata sulla distanza di Wasserstein. Inoltre, superiamo le difficoltà analitiche tramite la derivazione di limiti superiori e inferiori. La specializzazione dei limiti alle misure aleatorie Gamma fornisce il valore esatto della distanza di Wasserstein in termini di distanza di Kolmogorov tra misure base. I contenuti si basano su un lavoro di prossima pubblicazione in collaborazione con A. Lijoi e I. Prünster.*

Key words: Bayesian nonparametrics, Completely random measures, Increasing additive processes, Wasserstein distance.

Marta Catalano
Bocconi University, Milano, Italy. e-mail: marta.catalano@unibocconi.it

Antonio Lijoi
Bocconi University and Bocconi Institute of Data Science and Analytics (BIDSA), Milano, Italy. e-mail: antonio.lijoi@unibocconi.it

Igor Prünster
Bocconi University and BIDSA, Milano, Italy. e-mail: igor.pruenster@unibocconi.it

1 Introduction

Consider a generic parametric class of models $\mathscr{M} = \{M_\theta \,|\, \theta \in \Theta \subset \mathbb{R}^k\}$ assumed to describe or approximate the distribution of n observations $(x_1, \ldots, x_n)$. Many Bayesian inferential procedures rely on a notion of *discrepancy* between models, which is often translated into a distance between random variables. For example, sensitivity to the prior is assessed through a comparison between the posterior distributions, which often amounts to the evaluation of a (pseudo–)distance; see [2] for a review. Moreover, these are also used in model selection [10], variable selection [7], and in general in Bayesian testing when the hypotheses are nested. In this context many authors think that Bayes factors, corresponding to 0–1 losses, may be too restrictive and prefer considering distance–based losses instead [3, 16, 17].

A consistent portion of the Bayesian literature now focuses on the specification and properties of the so–called nonparametric priors, whose large support guarantees reliable predictions and estimations. The analytical tractability of the Dirichlet process [9] opened the way to the study of many nonparametric structures relying on random measures. For example, random measures are often normalized so to provide de Finetti measures [14, 15] or to define priors for density functions [1], which are both used to specify the law of an exchangeable process. In survival anaysis, moreover, they provide effective ways to specify the law of random hazard functions [8] or cumulative hazards [6, 11]. In all these cases, the specification of the law for a stochastic process of interest requires the distribution of a random measure. Typical inferential procedures analyse how this distribution is affected by the observed data. Nonparametric analogues of the previous procedures, such as sensitivity assessement and hypothesis testing, could then be based on distances between random measures. Interestingly, to the best of our knowledge there has not been any attempt to define such distances in the Bayesian nonparametric literature. We here propose a way to fill in this gap by exploiting the Wasserstein distance. While simulations of the Wasserstein distance are easily achieved [19], analytical evaluations are generally difficult. This raises the need for analytically tractable and sharp bounds. We achieve such bounds for a wide subclass of random measures, the so–called completely random measures.

The outline of the work is the following. After a brief recapitulation of basic notions about completely random measures and the Wasserstein distance, in Section 3 we provide general upper and lower bounds for the Wasserstein distance between completely random measures. These are expressed in terms of the underlying Lévy measures and are then specialized to Gamma completely random measures in Section 4.

2 Preliminaries

This section will be devoted to the definition of a distance between random measures. We first recall some useful properties of the Wasserstein distance and of completely random measures.

Let $\mathbb{X}$ be a Polish space with respect to a metric d, endowed with the Borel σ–algebra $\mathscr{B}(\mathbb{X})$, and let X_1 and X_2 be $\mathbb{X}$–valued random elements. The Wasserstein distance of order $p \in [1,+\infty)$ between X_1 and X_2 is defined as

$$W_{p,d}(X_1,X_2) = \inf_{(Z_1,Z_2)\in C(X_1,X_2)} \left\{\mathbb{E}(d(Z_1,Z_2)^p)^{\frac{1}{p}}\right\},$$

where $C(X_1,X_2)$ indicates the Fréchet class of X_1 and X_2, i.e. the set of distributions on the product space $\mathbb{X}^2$ whose marginal distributions on $\mathbb{X}$ coincide with the laws of X_1 and X_2. In the rest of the paper we will focus on the case $p = 1$ and $(\mathbb{X},d) = (\mathbb{R},|\cdot|)$, i.e. the real line with Euclidean distance, and we will denote such distance W. It can be shown that

$$|\mathbb{E}(X) - \mathbb{E}(Y)| \leq W(X,Y) \leq \mathbb{E}(|X|) + \mathbb{E}(|Y|). \tag{1}$$

In particular, the Wasserstein distance is finite when the random variables have finite mean.

Consider the space $\mathbb{M}(\mathbb{R})$ of boundedly finite measures on $\mathbb{R}$ endowed with the weak$^\sharp$ topology [5], and denote by $\mathscr{M}(\mathbb{R})$ the corresponding Borel σ-algebra. A random measure is a random element on the Borel space $(\mathbb{M}(\mathbb{R}),\mathscr{M}(\mathbb{R}))$. We identify each random measure $\tilde{\mu}$ with its corresponding cumulative process $\{\tilde{\mu}((-\infty,x])\}_{x\in\mathbb{R}}$ and propose the following distance between random measures.

Definition 1. Given two random measures $\tilde{\mu}_1$ and $\tilde{\mu}_2$ on $\mathbb{R}$ we define

$$d_W(\tilde{\mu}_1,\tilde{\mu}_2) = \sup_{x\in\mathbb{R}} W(\tilde{\mu}_1((-\infty,x]),\tilde{\mu}_2((-\infty,x])),$$

where W is the Wasserstein distance. It is easily shown that d_W is a distance on the laws of random measures and we refer to it as the Wasserstein distance between random measures. The distance d_W is finite whenever $\mathbb{E}(\tilde{\mu}_i(\mathbb{R})) < +\infty$, for $i = 1,2$.

The rest of the work concerns the evaluation of this distance on the so–called completely random measures. A random element $\tilde{\mu}$ taking values in $(\mathbb{M}(\mathbb{R}),\mathscr{M}(\mathbb{R}))$ is a completely random measure (CRM) if, given a finite collection of pairwise disjoint bounded sets $\{A_1,\cdots,A_n\}$ in $\mathscr{B}(\mathbb{R})$, the random variables $\{\tilde{\mu}(A_1),\cdots,\tilde{\mu}(A_n)\}$ are mutually independent [12]. Every CRM $\tilde{\mu}$ can be decomposed as the sum of three independent components, $\tilde{\mu} \overset{\mathrm{d}}{=} \mu + \tilde{\mu}_f + \tilde{\mu}_c$, where μ is a deterministic measure, $\tilde{\mu}_f$ is a random measure with fixed atoms and $\tilde{\mu}_c$ is a random measure without fixed atoms. Let $\mathbb{R}^+ = (0,+\infty)$. In particular, for every CRM without fixed atoms there exists a diffuse boundedly finite measure ν on $\mathbb{R}^+ \times \mathbb{R}$ such that

$$\tilde{\mu}_c(dy) \stackrel{\mathrm{d}}{=} \int_0^{+\infty} s\,\mathscr{N}(ds,dy),$$

where $\mathscr{N}$ is a Poisson random measure with intensity ν. This means that $\mathscr{N}$ is a CRM on $\mathbb{R}^+\times\mathbb{R}$ and, for any $B\in\mathscr{B}(\mathbb{R}^+)\otimes\mathscr{B}(\mathbb{R})$ such that $\nu(B)<\infty$, $\mathscr{N}(B)$ is a Poisson random variable with mean $\nu(B)$. The corresponding cumulative process $\{\tilde{\mu}_c((-\infty,x])\}_{x\in\mathbb{R}}$ is an increasing additive process on $\mathbb{R}$ with Lévy measures $\nu_x(ds)=\int_{(-\infty,x]}\nu(ds,dy)$ satisfying

$$\int_0^1 s\wedge 1\,\nu_x(ds)<+\infty \qquad \forall x\in\mathbb{R}.$$

In applications to Bayesian frameworks one is usually interested in CRMs that are infinitely active, i.e. such that $\nu_x((0,1])=+\infty\ \forall x\in\mathbb{R}$. Moreover, we point out that $\tilde{\mu}_c$ has finite mean if and only if $\sup_x\int_0^{+\infty}s\,\nu_x(ds)<+\infty$.

3 Wasserstein Bounds for Completely Random Measures

The Wasserstein distance can be easily simulated [19] but it is generally difficult to evaluate analytically. Nonetheless this is an important task since it can be used to quantify, for example, the sensitivity of the model to the prior specification of CRM. Our purpose for this section is to provide a general framework to derive upper and lower bounds for the Wasserstein distance between completely random measures in terms of their corresponding Lévy measure.

We focus on CRMs without fixed atoms, though our results can be extended in a natural way to CRMs with atoms. If $\tilde{\mu}$ is infinitely active, its Lévy measure ν is diffuse and not finite. Thus, for $i=1,2$ and $r>0$ there exists $\varepsilon_{i,r}>0$ such that

$$\nu_{i,x}([\varepsilon_{i,r},+\infty))=r.$$

We further define the probability measure $\rho_{i,r,x}$ to be proportional to the restriction of $\nu_{i,x}$ on the interval $[\varepsilon_{i,r},+\infty)$, i.e.

$$\rho_{i,r,x}(ds)=\frac{\nu_{i,x}(ds)}{r}\,\mathbb{1}_{[\varepsilon_{i,r},+\infty)}(s),$$

which can be shown to be the distribution of the jumps of the compound poisson approximation of $\mu_i((-\infty,x])$; see [18].

Theorem 1. *Let $\tilde{\mu}_1$ and $\tilde{\mu}_2$ be infinitely active* CRM*s with finite mean. Then for every $x\in\mathbb{R}$*

$$\left|\int_0^{+\infty} s\,(\nu_{1,x}-\nu_{2,x})(ds)\right| \le W(\tilde{\mu}_1((-\infty,x])),\tilde{\mu}_2((-\infty,x])) \le \lim_{r\to+\infty} r\,W(\rho_{1,r,x},\rho_{2,r,x}).$$

Moreover, the limit on the right hand side is finite.

The lower bound is an immediate consequence of (1). As for the right hand side, it can be seen as a generalization of [13], where the authors provide upper bounds on the Wasserstein distance between Lévy processes; see [4] for a proof.

4 Gamma Completely Random Measures

In this section we apply Theorem 1 to evaluate the exact expression of the Wasserstein distance between Gamma completely random measures. We recall that a Gamma CRM $\tilde{\mu}$ of parameter $b > 0$ and base measure α has Lévy intensity

$$\nu(ds,dy) = \frac{e^{-sb}}{s}\,\mathbb{1}_{(0,+\infty)}(s)\,ds\,\alpha(dy).$$

We use the notation $\tilde{\mu} \sim \mathrm{Ga}(b,\alpha)$. The random measure $\tilde{\mu}$ is easily shown to be infinitely active and, if α is a finite measure on $\mathbb{R}$, it has a finite mean. Moreover, we set $x \mapsto A(x) = \alpha((-\infty,x])$.

Theorem 2. *Let $\tilde{\mu}_i \sim Ga(b_i,\alpha_i)$, where $0 < b_1 < b_2$ and α_i is a finite measure on $\mathbb{R}$ for $i = 1,2$. Then,*

1. If $\alpha_1 = \alpha_2 = \alpha$,

$$W(\tilde{\mu}_1((-\infty,x]),\tilde{\mu}_2((-\infty,x])) = A(x)\left|\frac{1}{b_1} - \frac{1}{b_2}\right|;$$

2. If $b_1 = b_2 = b$,

$$W(\tilde{\mu}_1((-\infty,x]),\tilde{\mu}_2((-\infty,x])) = \frac{1}{b}\,|A_1(x) - A_2(x)|.$$

Remark 1. Theorem 2 clarifies the sharpness of the bounds derived in Theorem 1. Indeed, in this case the upper and lower bound coincide and can thus be used to derive the exact value of the Wasserstein distance.

Remark 2. Theorem 2 provides an immediate evaluation of the distance between completely random measures defined in Definition 1. By taking the supremum over $x \in \mathbb{R}$, one derives

1. If $\alpha_1 = \alpha_2 = \alpha$,
$$d_W(\tilde{\mu}_1,\tilde{\mu}_2) = \alpha(\mathbb{R})\left|\frac{1}{b_1} - \frac{1}{b_2}\right|;$$
2. If $b_1 = b_2 = b$,
$$d_W(\tilde{\mu}_1,\tilde{\mu}_2) = \frac{1}{b}K(\alpha_1,\alpha_2),$$

where $K(\alpha_1,\alpha_2)$ indicates the Kolmogorov distance between two finite measures, namely

$$K(\alpha_1, \alpha_2) = \sup_x |A_1(x) - A_2(x)|.$$

Intuitively, one expects that CRMs with similar parameters are close to each other. Our results confirm the intuition and allow for a precise quantification of the *closeness* in terms of Wasserstein distance.

References

1. Barrios, E. and Lijoi, A., Nieto-Barajas, L. E. and Prünster, I.: Modeling with normalized random measure mixture models. Stat. Sci. **28**, 313–334 (2013)
2. Berger, J.O.: An overview of robust Bayesian analysis (with discussion). Test. **3**, 5–124. Springer (1994)
3. Bernardo, J. M., Rueda, R.: Bayesian hypothesis testing: A reference approach. Int. Stat. Rev. **70**, 351–372 (2002)
4. Catalano, M., Lijoi, A., Prünster, I.: Wasserstein convergence rates of an approximate sampler for time-to-event data. In preparation.
5. Daley, D.J. and Vere-Jones, D.: An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods. Probability and Its Applications. Springer (2002)
6. Doksum, K.: Tailfree and Neutral Random Probabilities and Their Posterior Distributions. Ann. Probab. **2**, 183–201 (1974)
7. Dupuis, J. A., Robert, C. P.: Variable selection in qualitative models via an entropic explanatory power. J. Stat. Plan. Inference. **111**, 77–94 (2003)
8. Dykstra, R. L., Laud, P.: A Bayesian Nonparametric Approach to Reliability. Ann. Statist. **9**, 356–367 (1981)
9. Ferguson, T. S.: A Bayesian analysis of some nonparametric problems. Ann. Statist. **1**, 209–230 (1973)
10. Goutis, C., Robert, C.: Model choice in generalised linear models: A Bayesian approach via Kullback-Leibler projections. Biometrika. **85**, 29–37 (1998)
11. Hjort, N. L.: Nonparametric Bayes estimators based on beta processes in models for life history data. Ann. Statist. **18**, 1259–1294 (1990)
12. Kingman, J. F. C.: Completely random measures. Pacific J. Math. **21**, 59–78 (1967)
13. Mariucci, E. and Reiß, M.: Wasserstein and total variation distance between marginals of Lévy processes. Electron. J. Statist. **12**, 2482–2514 (2018)
14. Pitman, J. and Yor, M.: The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. Ann. Probab. **25**. 855–900 (1997)
15. Regazzini, E. and Lijoi, A. and Prünster, I.: Distributional results for means of normalized random measures with independent increments. Ann. Statist. **31**. 560–585 (2003)
16. Rousseau, J.: Approximating interval hypothesis: p-values and Bayes factors. Bayesian Statistics 8, Oxford Sci. Publ. **8**, 417–452 (2007)
17. Salomond, J.-B.: Testing Un-Separated Hypotheses by Estimating a Distance. Bayesian Anal. **13**. 461–484 (2018)
18. Sato, K.: Lévy Processes and Infinitely Divisible Distributions. Cambridge University Press (1999)
19. Sriperumbudur, B. K., Fukumizu, K., Gretton, A., Schölkopf, B., Lanckriet, G. R. G.: On the empirical estimation of integral probability metrics. Electron. J. Statist. **6**, 1550–1599 (2012)