Electronic Journal of Statistics
Vol. 14 (2020) 3366–3395
ISSN: 1935-7524
https://doi.org/10.1214/20-EJS1746

Approximation of Bayesian models for time-to-event data

Marta Catalano, Antonio Lijoi[*] **and Igor Prünster**[*]

Bocconi University, Italy

Abstract: Random measures are the key ingredient for effective nonparametric Bayesian modeling of time-to-event data. This paper focuses on priors for the hazard rate function, a popular choice being the kernel mixture with respect to a gamma random measure. Sampling schemes are usually based on approximations of the underlying random measure, both *a priori* and conditionally on the data. Our main goal is the quantification of approximation errors through the Wasserstein distance. Though easy to simulate, the Wasserstein distance is generally difficult to evaluate, making tractable and informative bounds essential. Here we accomplish this task on the wider class of completely random measures, yielding a measure of discrepancy between many noteworthy random measures, including the gamma, generalized gamma and beta families. By specializing these results to gamma kernel mixtures, we achieve upper and lower bounds for the Wasserstein distance between hazard rates, cumulative hazard rates and survival functions.

Keywords and phrases: Bayesian nonparametrics, completely random measures, gamma random measure, kernel mixtures, posterior sampling, survival analysis, Wasserstein distance.

Received February 2020.

1. Introduction

One of the most attractive features of the Bayesian nonparametric approach to statistical inference is the modeling flexibility implied by priors with large support. There are several classes of priors where this property is complemented by analytical tractability, thus contributing to making Bayesian nonparametrics very popular in several applied areas. See Hjort et al. [28] and Ghosal and van der Vaart [23] for broad overviews. In this framework, survival analysis stands out as one of the most lively fields of application. A prominent model for exchangeable time-to-event data is the extended gamma process for hazard rates [17], which allows for continuous observables and has been further generalized to kernel mixtures in Lo and Weng [37] and James [32]. These works paved the way for another active line of research that defines priors for the hazard rates by relaxing the dependence structure between the observables, going beyond the exchangeability assumption. For example, Pennell and Dunson [45], De Iorio et al. [13] and Hanson, Jara and Zhao [25] model subject specific hazards based

[*] Also affiliated to Bocconi Institute for Data Science and Analytics (BIDSA), Milano, Italy and Collegio Carlo Alberto, Torino, Italy.

on continuous covariates; Lijoi and Nipoti [35] and Zhou et al. [53] define priors for cluster specific hazards, while Nipoti, Jara and Guindani [43] account for both individual and cluster covariates simultaneously. In this work we rather focus on priors for the hazard rates of exchangeable time-to-event data.

An important feature shared by most classes of nonparametric priors is their definition in terms of random measures and transformations thereof. While there is a wealth of theoretical results that have eased their actual implementation in practice, sampling schemes are typically based on approximations of the underlying random measures. Nonetheless, with a very few exceptions [29, 2, 5], there is no extensive analysis on how to judge the quality of such approximations.

Consider the common situation where one is interested in making inference or sampling from a wide class of random measures $\mathcal{C}$, but can only treat a subclass $\mathcal{C}^\pi$ because of convenient analytical or computational properties. The restriction to $\mathcal{C}^\pi$ is usually argued through density statements, typically in terms of weak convergence of random measures. In many cases this reduces to the weak convergence of one-dimensional distributions, i.e. for every $\tilde{\mu} \in \mathcal{C}$ there exists an approximating sequence $\{\tilde{\mu}_n\}_{n\geq 1}$ in $\mathcal{C}^\pi$ such that $\tilde{\mu}_n(A)$ converges weakly to $\tilde{\mu}(A)$ for every Borel set A. This leaves out the possibility of establishing the rate of convergence and, more importantly, provides no guidance on the choice of the approximation $\tilde{\mu}_{\bar{n}} \in \{\tilde{\mu}_n\}_{n\geq 1}$ to use in practical implementations. Spurred by these considerations, the goal we pursue is to quantify the approximation errors by evaluating the Wasserstein distance between $\tilde{\mu}_n(A)$ and $\tilde{\mu}(A)$. Since convergence in Wasserstein distance implies weak convergence, this has the additional advantage of strengthening most results known in the literature.

The Wasserstein distance was first defined by Gini [24] as a *simple measure of discrepancy* between random variables. During the 20$^{\text{th}}$ century it has been redefined and studied in many other disciplines, such as transportation theory, partial differential equations, ergodic theory and optimization. Nowadays, depending on the field of study, it is known with different names, such as Gini distance, coupling distance, Monge-Kantorovich distance, Earth Moving distance and Mallows distance; see Villani [51], Rachev [46] and Cifarelli and Regazzini [8] for reviews. Indeed, one can find it scattered across the statistics literature [38, 16, 3, 6], though only in recent years it has achieved major success, especially in probability and machine learning. For a detailed review on the uses of the Wasserstein distance in statistics see Panaretos and Zemel [44]. As for the Bayesian literature, the Wasserstein distance was first used in Nguyen [42] and has been mainly used to evaluate approximations of the posterior distribution and to address consistency [42, 50, 7, 21, 15, 26]. These works deal with the convergence of the (random) Wasserstein distance between the attained values of random probability measures. In a similar vein, though without a specific statistical motivation, Mijoule, Peccati and Swan [40] examine the Wasserstein convergence rate of the empirical distribution to the prior, namely the de Finetti measure, for an exchangeable sequence of $\{0,1\}$-valued random variables. Our approach goes in a different direction: we are interested in a distance between the laws of random measures rather than a random distance between measures.

The Wasserstein distance is easy to simulate [49] but difficult to evaluate analytically and, hence, tractable bounds are needed for concrete applications. We achieve them in two steps. First, we determine bounds for the Wasserstein distance between so-called *completely random measures*, since they act as building blocks of most popular nonparametric priors. This is carried out by relying on results in Mariucci and Reiß [39] on Lévy processes. The techniques we develop in this first part measure the discrepancy between the laws of many noteworthy random measures, including the gamma, generalized gamma and beta families. Secondly, we move on to using these bounds in order to quantify the divergence between hazard rate mixture models that are used to analyze time-to-event data. These are then applied to evaluate the approximation error in a posterior sampling scheme for the hazards, in multiplicative intensity models, that relies on an algorithm for extended gamma processes [1].

The outline of the paper is as follows. After providing some basic notions and results on the Wasserstein distance and on completely random measures in Section 2, we determine upper and lower bounds for the Wasserstein distance between one-dimensional distributions associated to completely random measures in Section 3. This is, then, specialized to the case of gamma and beta completely random measures in Section 3.2. These results are the starting point for carrying out an in-depth analysis of hazard rate mixture models driven by completely random measures. In Section 4 we obtain a quantification of the discrepancy between two hazard rate mixtures and for the associated random survival functions. Examples related to its specification with mixing gamma random measures may be found in Section 4.3. Finally, in Section 5 we apply these results to evaluate the approximation error of a sampling scheme for the posterior hazards, conditional on the data. Proofs of the main results are deferred to Section 6.

2. Background and preliminaries

In this first section we recall some basic notions about completely random measures and their convergence in terms of the Wasserstein distance.

Let $\mathbb{X}$ be a Polish space with distance $d_{\mathbb{X}}$ and Borel σ-algebra $\mathcal{X}$. The space $M_{\mathbb{X}}$ of boundedly finite measures on $\mathbb{X}$ endowed with the weak$^\sharp$ topology is a Polish space as well; see Daley and Vere-Jones [9]. We denote by $\mathscr{M}_{\mathbb{X}}$ the corresponding Borel σ-algebra. A random measure is a measurable function from some probability space $(\Omega, \Sigma, \mathbb{P})$ to $(M_{\mathbb{X}}, \mathscr{M}_{\mathbb{X}})$.

Definition 1. If a random measure $\tilde{\mu}$ is such that, for any $n \geq 2$ and any collection of pairwise disjoint bounded sets $\{A_1, \cdots, A_n\}$ in $\mathcal{X}$, the random variables $\tilde{\mu}(A_1), \cdots, \tilde{\mu}(A_n)$ are mutually independent, then it is a *completely random measure* (CRM).

Every CRM can be uniquely represented as the sum of three independent components, $\mu + \tilde{\mu}_f + \tilde{\mu}$, where μ is a fixed measure on $\mathbb{X}$, $\tilde{\mu}_f$ is a random measure with fixed atoms and $\tilde{\mu}$ is a random measure without fixed atoms. See

Kingman [33]. Here we focus on CRMs without fixed atoms and rely on the fact that their distribution is uniquely determined by a Poisson random measure. Indeed,

$$\tilde{\mu}(dy) \overset{\mathrm{d}}{=} \int_{\mathbb{R}^+} s\,\mathcal{N}(ds, dy), \tag{1}$$

where $\mathcal{N}$ is a CRM on $\mathbb{R}^+ \times \mathbb{X}$ such that $\mathcal{N}(B)$ has a Poisson distribution of parameter $\nu(B) = \mathbb{E}\mathcal{N}(B)$ for every Borel set $B \in \mathscr{B}(\mathbb{R}^+) \otimes \mathcal{X}$ such that $\nu(B) < +\infty$. The mean measure ν on $\mathbb{R}^+ \times \mathbb{X}$ is referred to as *Lévy intensity* of $\tilde{\mu}$ and is such that for all $x \in \mathbb{X}$, $\nu(\mathbb{R}^+ \times \{x\}) = 0$, and for all bounded A in $\mathcal{X}$ and $\epsilon > 0$,

$$\int_A \int_{\mathbb{R}^+} (\epsilon \wedge s)\,\nu(ds, dy) < +\infty, \tag{2}$$

where $\wedge$ denotes the minimum. Motivated by Bayesian nonparametric modeling, we focus on Lévy intensities ν without atoms such that: (i) integrability condition (2) holds for every A in $\mathcal{X}$; (ii) for all A in $\mathcal{X}$ and $\epsilon > 0$, $\nu((0, \epsilon] \times A) = +\infty$. The latter corresponds to assuming $\tilde{\mu}$ *infinitely active*. Infinite activity is a necessary requirement for defining random probability measures by normalization of CRMs. See Lijoi and Prünster [36] for a review. Finally, in view of (1), the probability distribution of $\tilde{\mu}$ can be characterized through the Laplace functional transform

$$\mathbb{E}\Big(e^{-\int_{\mathbb{X}} f(y)\tilde{\mu}(dy)}\Big) = \exp\Big\{ - \int_{\mathbb{R}^+ \times \mathbb{X}} [1 - e^{-s\,f(y)}]\nu(ds, dy)\Big\}, \tag{3}$$

for all measurable functions $f : \mathbb{X} \to [0, +\infty)$.

When dealing with convergence of random measures we think of random measures in terms of probability distributions on $M_{\mathbb{X}}$. Results in strong convergence are often too hard to establish, so that one usually deals with weak convergence (of distributions) of random measures, $\mathcal{L}(\tilde{\mu}_n) \Rightarrow \mathcal{L}(\tilde{\mu})$, where $\mathcal{L}(X)$ denotes the probability distribution of a random element X, which can be either finite- or infinite-dimensional. A remarkable result establishes that this is equivalent to the weak convergence of all finite-dimensional distributions $\mathcal{L}(\tilde{\mu}_n(A_1), \cdots, \tilde{\mu}_n(A_d)) \Rightarrow \mathcal{L}(\tilde{\mu}(A_1), \cdots, \tilde{\mu}(A_d))$, for $A_1, \ldots, A_d \in \mathcal{X}$ stochastic continuity sets for $\tilde{\mu}$; see Theorem 11.1.VII in Daley and Vere-Jones [10]. Moreover, when dealing with CRMs, the weak convergence of finite-dimensional distributions is equivalent to the weak convergence of one-dimensional distributions. Thus, if d denotes a metric on the space of probability distributions on $\mathbb{R}$ whose convergence is stronger than the weak convergence, one has that if

$$d(\mathcal{L}(\tilde{\mu}_n(A)), \mathcal{L}(\tilde{\mu}(A))) \to 0 \tag{4}$$

for every $A \in \mathcal{X}$, then $\tilde{\mu}_n$ converges weakly to $\tilde{\mu}$. In the sequel, we will choose d as the Wasserstein distance with respect to the Euclidean norm on $\mathbb{R}$. See Villani [51]. In order to define this distance in its full generality, let $\mathbb{Y}$ be a Polish with respect to the metric $d_{\mathbb{Y}}$. For any pair of random elements Y_1 and Y_2 taking values in $\mathbb{Y}$, let $C(Y_1, Y_2)$ denote the Fréchet class of Y_1 and Y_2, i.e. the set of random elements (Z_1, Z_2) on the product space $\mathbb{Y}^2$ such that $\mathcal{L}(Z_i) = \mathcal{L}(Y_i)$ for $i = 1, 2$.

Definition 2. The *Wasserstein distance of order* $p \in [1, +\infty)$ between $\mathcal{L}(Y_1)$ and $\mathcal{L}(Y_2)$ is defined as

$$\mathcal{W}_{p,d_{\mathbb{Y}}}\left(\mathcal{L}(Y_1), \mathcal{L}(Y_2)\right) = \inf_{(Z_1,Z_2)\in C(Y_1,Y_2)} \left\{\mathbb{E}(d_{\mathbb{Y}}(Z_1, Z_2)^p)^{\frac{1}{p}}\right\}.$$

We focus on the case $p = 1$ and $(\mathbb{Y}, d_{\mathbb{Y}}) = (\mathbb{R}, |\cdot|)$. We denote such a distance as $\mathcal{W}$ and omit reference to the law $\mathcal{L}$ in the notation. In view of the previous discussion on weak convergence, a major goal that we pursue is evaluating or bounding $\mathcal{W}(\tilde{\mu}_1(A), \tilde{\mu}_2(A))$, for A in $\mathcal{X}$. The results and general techniques that will be detailed in the next sections make use of some known facts on Wasserstein distances and CRMs concisely recalled here. First, for any pair of random variables (X, Y),

$$|\mathbb{E}(X) - \mathbb{E}(Y)| \leq \mathcal{W}(X, Y) \leq \mathbb{E}(|X|) + \mathbb{E}(|Y|). \tag{5}$$

Thus the Wasserstein distance is finite when the random variables have finite mean. We will therefore focus our attention on CRMs whose total mass has finite mean and refer to them as CRMs with finite mean. By Campbell's Theorem this boils down to

$$\mathbb{E}(\tilde{\mu}(\mathbb{R})) = \mathbb{E}\left(\int_{\mathbb{R}^+\times\mathbb{X}} s\,\mathcal{N}(ds, dy)\right) = \int_{\mathbb{R}^+\times\mathbb{X}} s\,\nu(ds, dy) < +\infty. \tag{6}$$

Finally, we highlight two properties of the Wasserstein distance that will be used in many proofs. Let X, Y be random variables and let F_X, F_Y denote their distribution functions. Then by [11],

$$\mathcal{W}(X, Y) = \int_{-\infty}^{+\infty} |F_X(u) - F_Y(u)|\,du. \tag{7}$$

Moreover, if $X_1, \ldots, X_n$ are independent random variables and $Y_1, \cdots, Y_n$ are independent as well, then by [3],

$$\mathcal{W}(X_1 + \cdots + X_n, Y_1 + \cdots + Y_n) \leq \sum_{i=1}^{n} \mathcal{W}(X_i, Y_i). \tag{8}$$

3. Wasserstein bounds for CRMs

3.1. General result

There are situations where one is only interested in a numerical value for the Wasserstein distance: in such a case there are efficient ways to simulate it. See [49]. On the other hand, one may be interested in understanding how the distance is affected by the parameters of the distributions or by meaningful functionals, such as moments. This raises the need for an analytical evaluation of the Wasserstein distance, which in general is not an easy task. The most common practice is thus to develop informative bounds and to analyze how these are affected by the choices above. In this section we will express a bound for the Wasserstein distance between the one-dimensional distributions of CRMs in terms of their corresponding Lévy intensities. The proof is based on a compound Poisson approximation of CRMs.

Theorem 1. *Let $\tilde\mu_1$ and $\tilde\mu_2$ be infinitely active* CRM*s with finite mean. Then for every $A \in \mathcal{X}$*

$$g_\ell(A) \le W(\tilde\mu_1(A), \tilde\mu_2(A)) \le g_u(A),$$

where

$$g_\ell(A) = |\mathbb{E}(\tilde\mu_1(A)) - \mathbb{E}(\tilde\mu_2(A))| = \left| \int_{\mathbb{R}^+} s\,\nu_1(ds \times A) - \int_{\mathbb{R}^+} s\,\nu_2(ds \times A) \right|,$$
$$g_u(A) = \int_0^{+\infty} |\nu_1((u,+\infty) \times A) - \nu_2((u,+\infty) \times A)|\,du.$$

We observe that $g_u(A)$ has a compelling form with respect to the upper bound in (5), since it equals zero if $\tilde\mu_1 \overset{\mathrm{d}}{=} \tilde\mu_2$. We stress that this bound holds for all CRMs and may be evaluated through numerical integration. Nonetheless, when specializing to certain classes of CRMs, we manage to upper bound $g_u(A)$ with an expression that can be evaluated exactly, as we do in Section 3.2. In particular, easily computable upper bounds are available whenever the tails of the Lévy intensities are ordered, as we prove in the next corollary. In such a case not only we have a simple expression for $g_u(A)$, we may also prove that the upper and lower bounds coincide, providing the exact expression of the Wasserstein distance itself.

Corollary 2. *Consider the hypotheses of* Theorem 1 *and let $A \in \mathcal{X}$. If the tails of $\nu_1(ds \times A)$ and $\nu_2(ds \times A)$ are ordered, namely $\nu_i((u,+\infty) \times A) \le \nu_j((u,+\infty) \times A)$ for all $u \in \mathbb{R}^+$ and $i \ne j$ in $\{1,2\}$, then*

$$\mathcal{W}(\tilde\mu_1(A), \tilde\mu_2(A)) = \left| \int_{\mathbb{R}^+} s\,\nu_1(ds \times A) - \int_{\mathbb{R}^+} s\,\nu_2(ds \times A) \right|.$$

Remark 1. The condition of Corollary 2 holds whenever there exists a dominating measure η on $\mathbb{R}^+$ such that the Radon–Nikodym derivatives of $\nu_1(ds \times A)$ and $\nu_2(ds \times A)$ are ordered, i.e. $\nu_{i,A}(s) \le \nu_{j,A}(s)$ for all $s \in \mathbb{R}^+$ and $i \ne j$ in $\{1,2\}$. This more restrictive condition, which is however much easier to verify, holds true for many examples to be displayed in the sequel.

As underlined in Section 2, the convergence in the Wasserstein distance of $\tilde\mu_n(A)$ to $\tilde\mu(A)$, for every $A \in \mathcal{X}$, is sufficient to guarantee the weak convergence of the sequence $(\tilde\mu_n)_{n\ge1}$ and provide convergence rates. This motivates our main interest in set-wise results as those in Theorem 1. However, one can also define a uniform distance between laws of random measures with finite mean, by considering

$$d_{\mathcal{W}}(\tilde{\mu}_1, \tilde{\mu}_2) = \sup_{A \in \mathcal{X}} \mathcal{W}(\tilde{\mu}_1(A), \tilde{\mu}_2(A)).$$

Corollary 2 can be used to find the exact expression of such distance. We focus on *homogeneous* CRMs, i.e. such that their Lévy intensity is a product measure $\nu(ds, dx) = \rho(s)\, ds\, \alpha(dx)$. It will be next shown that $d_{\mathcal{W}}$ admits a very intuitive representation, being proportional to the total variation distance between the base measures

$$\mathrm{TV}(\alpha_1, \alpha_2) = \sup_{A \in \mathcal{X}} |\alpha_1(A) - \alpha_2(A)|.$$

Corollary 3. *Let $\tilde{\mu}_i$ be infinitely active homogeneous* CRM*s with finite mean such that the Lévy intensities $\nu_i(ds, dx) = \rho(s)\, ds\, \alpha_i(dx)$, for $i = 1, 2$. Then,*

$$d_{\mathcal{W}}(\tilde{\mu}_1, \tilde{\mu}_2) = \mathrm{TV}(\alpha_1, \alpha_2) \int_{\mathbb{R}^+} s\, \rho(s)\, ds.$$

3.2. Examples

When the conditions of Corollary 2 do not hold, one may often find upper bounds of $g_u(A)$ which may be evaluated exactly for specific examples of CRMs. In the next proposition we consider a gamma CRM with rate parameter $b > 0$ and base measure α whose Lévy intensity is

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

We use the notation $\tilde{\mu} \sim \mathrm{Ga}(b, \alpha)$. The random measure $\tilde{\mu}$ is infinitely active and, if α is a finite measure on $\mathbb{X}$, it has finite mean.

Proposition 4. *Let $\tilde{\mu}_i \sim \mathrm{Ga}(b_i, \alpha_i)$, where $0 < b_1 < b_2$ and α_i is a finite measure on $\mathbb{X}$ for $i = 1, 2$. Then,*

$$g_\ell(\boldsymbol{b}, \boldsymbol{\alpha}, A) \le \mathcal{W}(\tilde{\mu}_1(A), \tilde{\mu}_2(A)) \le g_u(\boldsymbol{b}, \boldsymbol{\alpha}, A),$$

where

$$g_\ell(\boldsymbol{b}, \boldsymbol{\alpha}, t) = \left| \frac{\alpha_1(A)}{b_1} - \frac{\alpha_2(A)}{b_2} \right|,$$
$$g_u(\boldsymbol{b}, \boldsymbol{\alpha}, t) = \frac{\alpha_1(A)}{b_1} - \frac{\alpha_2(A)}{b_2} + \mathbb{1}_{(0,+\infty)}(\alpha_2(A) - \alpha_1(A))\, 2\, \frac{\alpha_2(A) - \alpha_1(A)}{b_2 - b_1} \log \frac{b_2}{b_1},$$

and we have used the vector notations $\boldsymbol{b} = (b_1, b_2)$ and $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)$.

This result extends the ones in Gairing et al. [20], who develop upper bounds for similar integrals of gamma Lévy intensities in a more restrictive framework as they do not allow for both base measures and the scale parameter to differ between the two specifications. The bounds of Proposition 4 are informative

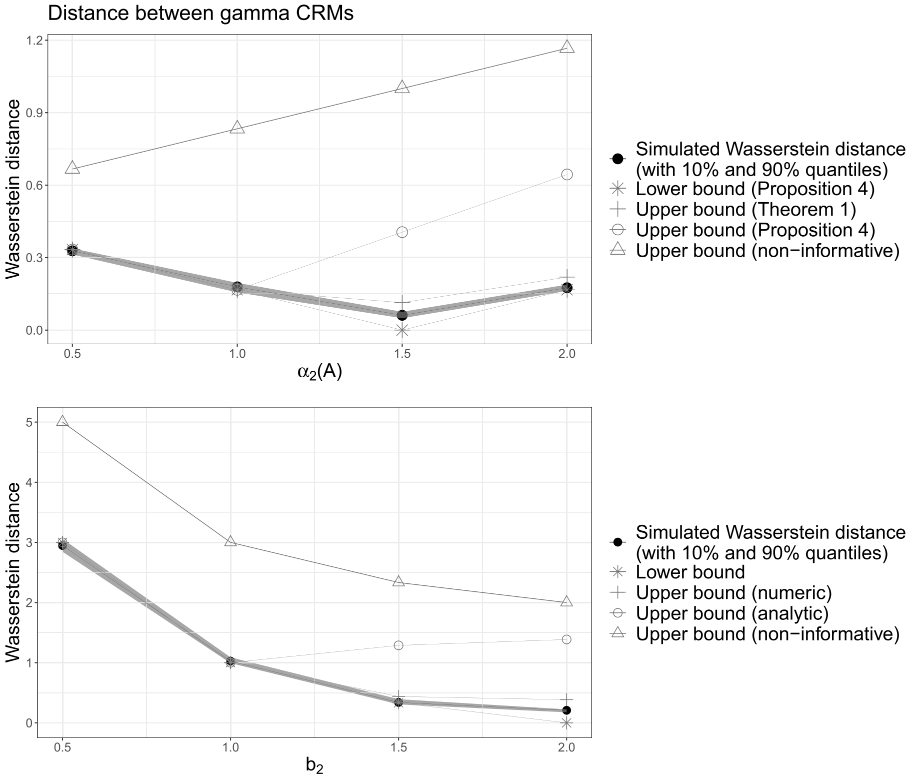

FIG 1. *Wasserstein distance* $\mathcal{W}(\tilde{\mu}_1(A), \tilde{\mu}_2(A))$ *between gamma* CRMs *and relative upper and lower bounds. In the upper panel* $\alpha_1(A) = 1, b_1 = 2, b_2 = 3$ *are fixed, whereas* $\alpha_2(A)$ *ranges from* 0.5 *to* 2. *In the lower one* $\alpha_1(A) = 1, b_1 = 1, \alpha_2(A) = 2$ *are fixed, whereas* b_2 *ranges from* 0.5 *to* 2. *In both plots the* Simulated Wasserstein distance *is based on 10 samples of 1000 observations using the Python Optimal Transport (POT) package [19].*

in the sense that, the closer the parameters of the two CRMs, the smaller the bound of the Wasserstein distance. Moreover, when the base measures are equal on A, the upper and lower bounds coincide, providing the exact expression for the Wasserstein distance, in accordance with Corollary 2. The same holds true if $b_1 = b_2$, since

$$\lim_{b_2 \to b_1^+} \frac{1}{b_2 - b_1} \log \frac{b_2}{b_1} = \frac{1}{b_1}.$$

In Figure 1 we compare the simulated Wasserstein distance between two gamma CRMs with the upper bound in Theorem 1, which can be evaluated numerically, the ones in Proposition 4, which can be evaluated exactly, and the upper bound in (5), which is non-informative. For a wide range of parameters the bounds of Theorem 1 and Proposition 4 coincide with the Wasserstein distance. In contrast, when the Lévy intensities are not ordered, the upper and lower bounds do not coincide. The upper bound of Proposition 4 is tight whenever at least one of the two parameters is close to the corresponding parameter of the other CRM (i.e. $\alpha_1(A) \approx \alpha_2(A)$ or $b_1 \approx b_2$), whereas the upper bound of Theorem 1 is tight on the whole range of parameters. Moreover, they are both more informative than the upper bound in (5). The lower bound, on the other hand, is always tight and becomes non-informative when the two CRMs have different parameters but equal ratios $(\alpha_i(A)/b_i)$, i.e. when they have equal mean.

A different situation occurs with beta CRMs, where the Lévy densities corresponding to different concentration parameters and same base measure are not ordered. We recall that $\tilde{\mu} \sim \mathrm{Be}(c, \alpha)$ is a beta CRM of concentration parameter c and base measure α if the Lévy intensity is

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

Proposition 5. *Let $\tilde{\mu}_i \sim \mathrm{Be}(c_i, \alpha_i)$, where $0 < c_1 \leq c_2$ and α_i is a finite measure on $\mathbb{X}$ for $i = 1, 2$.*

1) *If $c_1 = c_2 = c$, then $\mathcal{W}(\tilde{\mu}_1(A), \tilde{\mu}_2(A)) = c\,|\alpha_1(A) - \alpha_2(A)|$. Thus, $d_{\mathcal{W}}(\tilde{\mu}_1, \tilde{\mu}_2) = c\,\mathrm{TV}(\alpha_1, \alpha_2)$.*
2) *If $\alpha_1 = \alpha_2 = \alpha$, then $\mathcal{W}(\tilde{\mu}_1(A), \tilde{\mu}_2(A))) \leq 2\,\alpha(A)\,\log\left(\frac{c_2}{c_1}\right)$.*

We conclude this section with an immediate application of Corollary 3 on the distance $d_{\mathcal{W}}$ between the laws of generalized gamma CRMs.

Example 1. Consider *generalized gamma* CRMs with Lévy intensities

$$\nu_i(ds, dx) = \frac{1}{\Gamma(1-\sigma)} s^{-1-\sigma} e^{-bs} ds\,\alpha_i(dx),$$

where $\sigma \in [0, 1)$ and $b > 0$, for $i = 1, 2$. Then Corollary 3 ensures that $d_{\mathcal{W}}(\tilde{\mu}_1, \tilde{\mu}_2) = b^{\sigma-1}\,\mathrm{TV}(\alpha_1, \alpha_2)$. When $\sigma = 0$ we recover the distance between two gamma CRMs with same rate parameter.

4. Hazard rate mixtures

Applications in survival analysis and reliability involve *time-to-event* data and have spurred important developments in Bayesian nonparametric modeling. Stimulating and exhaustive overviews of popular models in the area can be found in Müller et al. [41] and in Ghosal and van der Vaart [23]. If $T_1, \ldots, T_n$ are from an exchangeable sequence of time-to-event data, i.e.

$$T_i \,|\, \tilde{P} \overset{\text{iid}}{\sim} \tilde{P} \quad (i = 1, \ldots, n), \qquad \tilde{P} \sim \Pi, \tag{9}$$

the choice of Π follows from specifying a prior on the survival function $t \mapsto \tilde{S}(t) = \tilde{P}((t, \infty))$. This may be done directly by resorting, e.g., to neutral to the right random probability measure [14], or by setting a prior on the corresponding cumulative hazard function by means of, e.g., the Beta process [27].

Alternatively, one may specify a prior on the hazard rate function if one can assume that $\tilde{S}$ is almost surely continuous: in this case a convenient option is a kernel mixture model [17]. For all these model specifications, one can also take into account the presence of censored observations. The most common mechanism is right-censoring, which associates to each T_i a censoring time C_i. In this case, the actual observations are the pairs (X_i, Δ_i), where $X_i = T_i \wedge C_i$ and $\Delta_i = \mathbb{1}_{(0,C_i]}(T_i)$ identifies exact observations whenever it equals 1. Here we focus on priors for the hazard rates, i.e. the instantaneous risk of failure, that are induced by kernel mixtures over a gamma CRM. The model, originally proposed as a prior for increasing hazard rates in Dykstra and Laud [17], is ideal for treating right censored observations and has led to several interesting generalizations. Henceforth, we consider a specification that has been investigated in its full generality by James [32].

Before focusing on our main results, let us first recall some basic definitions that will also allow us to set the notation to be used throughout. If F is a cumulative distribution function on $[0, +\infty)$ and $S = 1 - F$ the corresponding survival function, we assume it is absolutely continuous so that one can define the *hazard rate* $h = F'/(1 - F)$ and rewrite, for any $t \geq 0$,

$$S(t) = \exp\{-H(t)\}, \qquad H(t) = \int_0^t h(s)\, ds,$$

where H is the cumulative hazard function. Let $k : \mathbb{R}^+ \times \mathbb{X} \to [0, +\infty)$ be a measurable kernel function. If $\tilde{\mu}$ is a CRM, with corresponding Poisson random measure $\mathcal{N}$, and k is such that

$$\lim_{t\to\infty} \int_0^t \int_{\mathbb{R}^+\times\mathbb{X}} k(u \,|\, y)\, s\, du\, \mathcal{N}(ds, dy) = +\infty, \tag{10}$$

a prior for the hazard rates is the probability distribution of the process $\{\tilde{h}(t) \mid t \geq 0\}$ such that for any $t \geq 0$,

$$\tilde{h}(t) = \int_{\mathbb{X}} k(t \,|\, y)\, \tilde{\mu}(dy) \overset{\mathrm{d}}{=} \int_{\mathbb{R}^+\times\mathbb{X}} k(t \,|\, y)\, s\, \mathcal{N}(ds, dy). \tag{11}$$

Thus, condition (10) ensures that the mean cumulative hazards go to $+\infty$ as time increases. We use the techniques developed in the previous sections to obtain bounds on the Wasserstein distance between the marginal hazard rates coming from different kernels and different CRMs. Moreover, we successfully address the same issue when considering the Wasserstein distance between cumulative hazards and survival functions.

4.1. Bounds for hazard rates

Consider two random hazard rates $\tilde{h}_1 = \{\tilde{h}_1(t) \,|\, t \geq 0\}$ and $\tilde{h}_2 = \{\tilde{h}_2(t) \,|\, t \geq 0\}$ as in (11). From a statistical standpoint, these may be seen as different prior specifications corresponding, e.g., to different mixing CMRs or kernels. Alternatively, $\tilde{h}_2$ may be thought of as an approximation of the actual prior $\tilde{h}_1$ and one may be interested to ascertain the quality of such an approximation. The issue is of great interest when we need to sample from $\tilde{h}_1$, or its posterior distribution, while it is much easier to sample from $\tilde{h}_2$: in this case a bound on the error can provide an effective guidance as on how to fix the parameters of the approximating distribution. We first investigate how different CRMs and kernels impact the marginal hazards and use the Wasserstein distance as a measure. In other terms, we will be focusing on $\mathcal{W}(\tilde{h}_1(t), \tilde{h}_2(t))$ for every $t \geq 0$. The results in the previous sections will provide the necessary background for obtaining the desired bounds. Before displaying these, we state a technical result. To this end, we recall that if ν is a measure on $\mathbb{X}$ and $g : \mathbb{X} \to \mathbb{Y}$ is a measurable function, the pushforward measure $g \# \nu$ on $\mathbb{Y}$ is defined by $(g \# \nu)(A) = \nu(g^{-1}(A))$.

Lemma 6. *Let $\tilde{\mu}$ be a* CRM *with intensity measure ν and let $f : \mathbb{X} \to \mathbb{R}^+$ be a measurable function. Then the random measure $\tilde{\mu}_f(dy) = f(y)\, \tilde{\mu}(dy)$ is a* CRM *with Lévy intensity equal to the pushforward measure $\nu_f = p_f \# \nu$ where $p_f(s, y) = (s\, f(y), y)$. Thus for every $A \in \mathcal{X}$,*

$$\int_{\mathbb{R}^+ \times A} s\, \nu_f(ds, dy) = \int_{\mathbb{R}^+ \times A} s f(y)\, \nu(ds, dy). \tag{12}$$

When $\nu(ds, dy) = \nu(s, y)\, ds\, \alpha(dy)$, by Lemma 6 with a change of variable,

$$\nu_f(ds, dy) = \frac{1}{f(s)}\, \nu\Big(\frac{s}{f(y)}, y\Big) ds\, \alpha(dy).$$

Thus, we will use the notation $\nu_f(ds, dy) = \frac{1}{f(s)} \nu(d\frac{s}{f(y)}, dy)$. The relevance of this change of measure result is apparent since the prior specification in (11) involves a multiplicative structure with the kernel and the mixing CRM. The following example deals with the gamma case.

Example 2. Consider $\tilde{\mu} \sim \mathrm{Ga}(b, \alpha)$ and a generic kernel k. Then the random measures defined by $\tilde{\mu}_{k(t|\cdot)}(dy) = k(t \,|\, y)\tilde{\mu}(dy)$ are CRMs with Lévy intensity

$$\nu_{k(t|\cdot)}(ds, dy) = \frac{e^{-\frac{sb}{k(t\,|\,y)}}}{s}\, \mathbb{1}_{(0,+\infty)}(s)\, ds\, \alpha(dy).$$

Thus $\tilde{\mu}_{k(t|\cdot)}$ is an *extended gamma* CRM with scale function $\beta(y) = \frac{k(t\,|\,y)}{b}$ and base measure α. Extended gamma CRMs are easily shown to be infinitely active.

Lemma 6 ensures that marginally the hazard process in (11) satisfies $\tilde{h}(t) \overset{\mathrm{d}}{=} \tilde{\mu}_{k(t|\cdot)}(\mathbb{X})$, where $\tilde{\mu}_{k(t|\cdot)}$ is a CRM. In order to bound the Wasserstein distance between marginal hazards we may thus apply the results of Theorem 1 with $A = \mathbb{X}$. By (12), $\tilde{\mu}_{k(t|\cdot)}$ has finite mean and it is infinitely active if, respectively,

$$\int_{\mathbb{R}^+ \times \mathbb{X}} k(t \,|\, y)\, s\, \nu(ds, dy) < +\infty, \tag{13}$$

$$\int_{[0,\epsilon]\times A} \frac{1}{k(t\,|\,y)}\nu\Big(d\frac{s}{k(t\,|\,y)}, dy\Big) = +\infty, \tag{14}$$

for every $\epsilon \geq 0$, $A \in \mathcal{X}$ and $t \geq 0$. If ν is infinitely active, (14) holds.

Theorem 7. *Let $\tilde h_1 = \{\tilde h_1(t)\,|\,t \geq 0\}$ and $\tilde h_2 = \{\tilde h_2(t)\,|\,t \geq 0\}$ be random hazard rates as in (11) with associated infinitely active* CRMs *$\tilde\mu_i$, Lévy intensity ν_i, and kernel k_i that satisfy (10) and (13), for $i = 1, 2$. Then the Wasserstein distance between the marginal hazard rates is finite and for every $t \geq 0$,*

$$g_\ell(t) \leq \mathcal{W}(\tilde h_1(t), \tilde h_2(t)) \leq g_u(t),$$

where

$$\begin{aligned} g_\ell(t) &= \Big|\int_{\mathbb{R}^+\times\mathbb{X}} k_1(t\,|\,y)\, s\,\nu_1(ds, dy) - \int_{\mathbb{R}^+\times\mathbb{X}} k_2(t\,|\,y)\, s\,\nu_2(ds, dy)\Big|,\\ g_u(t) &= \int_0^{+\infty} \Big| \int_{(u,+\infty)\times\mathbb{X}} \frac{1}{k_1(t\,|\,y)}\nu_1\Big(d\frac{s}{k_1(t\,|\,y)}, dy\Big)\\ &\qquad\qquad - \frac{1}{k_2(t\,|\,y)}\nu_2\Big(d\frac{s}{k_2(t\,|\,y)}, dy\Big)\Big|\, du. \end{aligned}$$

In particular, if there exists a dominating measure η such that the Radon–Nikodym derivatives $\nu_i(s, y)$ satisfy, for $i \neq j$ in $\{1, 2\}$,

$$\frac{1}{k_i(t\,|\,y)}\nu_i\Big(\frac{s}{k_i(t\,|\,y)}, y\Big) \leq \frac{1}{k_j(t\,|\,y)}\nu_j\Big(\frac{s}{k_j(t\,|\,y)}, y\Big)$$

for all $(s, y) \in \mathbb{R}^+ \times \mathbb{X}$ and $t \geq 0$, then

$$\mathcal{W}(\tilde h_1(t), \tilde h_2(t)) = \Big|\int_{\mathbb{R}^+\times\mathbb{X}} k_1(t\,|\,y)\, s\,\nu_1(ds, dy) - \int_{\mathbb{R}^+\times\mathbb{X}} k_2(t\,|\,y)\, s\,\nu_2(ds, dy)\Big|.$$

4.2. Bounds for survival functions

The bounds we have derived for the hazard rates translate into bounds for the corresponding survival functions and these are of great interest since one typically targets estimation of functionals of the survival function. First, we consider the corresponding cumulative hazards processes $\tilde H = \{\tilde H(t)\,|\,t \geq 0\}$, defined by

$$\tilde H(t) = \int_0^t \tilde h(u)\, du = \int_{\mathbb{X}} K(t\,|\,y)\,\tilde\mu(dy), \tag{15}$$

where $K(t\,|\,y) = \int_0^t k(u\,|\,y)\, du$ is the *cumulative kernel*. Thus, the cumulative hazards can be treated as a kernel mixture as well, and an analogue of Theorem 7 is available.

Theorem 8. *Let $\tilde H_1 = \{\tilde H_1(t)\,|\,t \geq 0\}$ and $\tilde H_2 = \{\tilde H_2(t)\,|\,t \geq 0\}$ be two random cumulative hazards as in (15) with associated infinitely active* CRMs *$\tilde\mu_i$, with*

Lévy intensity ν_i, and kernel k_i that satisfy (10), *for $i = 1, 2$. If the cumulative kernels $K_i(t \,|\, y) = \int_0^t k_i(u \,|\, y) du$ satisfy* (13) *with $k = K_i$, the Wasserstein distance between the marginal cumulative hazards is finite and for every $t \geq 0$,*

$$g_\ell(t) \leq \mathcal{W}(\tilde{H}_1(t), \tilde{H}_2(t)) \leq g_u(t),$$

where

$$\begin{aligned} g_\ell(t) &= \bigg| \int_{\mathbb{R}^+ \times \mathbb{X}} K_1(t \,|\, y)\, s\, \nu_1(ds, dy) - K_2(t \,|\, y)\, s\, \nu_2(ds, dy) \bigg|, \\ g_u(t) &= \int_0^{+\infty} \bigg| \int_{(u, +\infty) \times \mathbb{X}} \frac{1}{K_1(t \,|\, y)} \nu_1\Big(d \frac{s}{K_1(t \,|\, y)}, dy\Big) \\ &\qquad\qquad - \frac{1}{K_2(t \,|\, y)} \nu_2\Big(d \frac{s}{K_2(t \,|\, y)}, dy\Big) \bigg| \, du. \end{aligned}$$

In particular, if there exists a dominating measure η such that the Radon–Nikodym derivatives $\nu_1(s, y), \nu_2(s, y)$ satisfy, for $i \neq j$ in $\{1, 2\}$,

$$\frac{1}{K_i(t \,|\, y)} \nu_i\Big(\frac{s}{K_i(t \,|\, y)}, y\Big) \leq \frac{1}{K_j(t \,|\, y)} \nu_j\Big(\frac{s}{K_j(t \,|\, y)}, y\Big)$$

for all $s, y \in \mathbb{R}^+ \times \mathcal{X}$, then

$$\mathcal{W}(\tilde{H}_1(t), \tilde{H}_2(t)) = \bigg| \int_{\mathbb{R}^+ \times \mathbb{X}} K_1(t \,|\, y)\, s\, \nu_1(ds, dy) - \int_{\mathbb{R}^+ \times \mathbb{X}} K_2(t \,|\, y)\, s\, \nu_2(ds, dy) \bigg|.$$

The bounds for the distance between cumulative hazards in Theorem 8 are also useful to identify a similar result for the survival function process $\tilde{S} = \{\tilde{S}(t) \,|\, t \geq 0\}$ defined by

$$t \mapsto \tilde{S}(t) = e^{-\tilde{H}(t)} = \exp\bigg\{ - \int_{\mathbb{X}} K(t \,|\, y)\, \tilde{\mu}(dy) \bigg\}. \tag{16}$$

Theorem 9. *Let $\tilde{H}_1$ and $\tilde{H}_2$ be as in* Theorem 8 *with survival process $\tilde{S}_i$ as in* (16)*, for $i = 1, 2$. Then for every $t \geq 0$,*

$$g_\ell(t) \leq \mathcal{W}(\tilde{S}_1(t), \tilde{S}_2(t)) \leq g_{u,1}(t) \wedge g_{u,2}(t),$$

where

$$\begin{gathered} g_\ell(t) = \big| \mathbb{E}\big(e^{-\tilde{H}_1(t)}\big) - \mathbb{E}\big(e^{-\tilde{H}_2(t)}\big) \big|, \qquad g_{u,1}(t) = 1 - e^{-\mathcal{W}(\tilde{H}_1(t), \tilde{H}_2(t))}, \\ g_{u,2}(t) = \mathbb{E}\big(e^{-\tilde{H}_1(t)}\big) + \mathbb{E}\big(e^{-\tilde{H}_2(t)}\big) - \big(e^{-\mathbb{E}(\tilde{H}_1(t))} + e^{-\mathbb{E}(\tilde{H}_2(t))}\big) e^{-\mathcal{W}(\tilde{H}_1(t), \tilde{H}_2(t))}. \end{gathered}$$

4.3. Examples

We now apply these results on kernels of the type of Dykstra and Laud [17], $k(t|y) = \beta(y) \mathbb{1}_{[0,t]}(y)$, which is a popular choice when one wants to model increasing hazards. In this setting $\mathbb{X} = [0, +\infty)$. For simplicity we will restrict our attention to constant functions $\beta(s) = \beta$, which is a common choice in applications [17, 34], and gamma CRMs with the same base measure α. In this scenario α may also be an infinite measure, though it must be boundedly finite. We will consider the Lebesgue measure on the positive real axis, $\mathrm{Leb}^+(ds) = \mathbb{1}_{[0,+\infty)}(s)\,ds$, which is the base measure proposed in the original paper of Dykstra and Laud [17] and meets the conditions of Theorem 7.

Example 3. Let $\tilde{\mu}_i \sim \mathrm{Ga}(b_i, \mathrm{Leb}^+)$ and let $k_i(t\,|\,y) = \beta_i \mathbb{1}_{[0,t]}(y)$, with $b_i, \beta_i > 0$, for $i = 1, 2$. If $\tilde{h}_1$ and $\tilde{h}_2$ are the corresponding hazard rate mixtures, then

$$\mathcal{W}(\tilde{h}_1(t), \tilde{h}_2(t)) = t\left|\frac{\beta_1}{b_1} - \frac{\beta_2}{b_2}\right|.$$

Proof. The general expression for $\nu_{k(t|\cdot)}$ was derived in Example 2. With our choices,

$$\frac{1}{k_i(t\,|\,y)}\nu_i\Big(d\frac{s}{k_i(t\,|\,y)}, dy\Big) = \frac{e^{-\frac{s\,b_i}{\beta_i}}}{s}\,\mathbb{1}_{(0,+\infty)}(s)\,\mathbb{1}_{[0,t]}(y)\,ds\,dy, \tag{17}$$

which corresponds to the Lévy intensity of a gamma CRM of parameter b_i/β_i and the restriction of Leb^+ to $[0,t]$ as base measure. Since the Lebesgue measure on a bounded set is finite, as observed in Proposition 4, (17) is infinitely active and has finite mean. Thus condition (13) holds. In order to check condition (10) on the expected cumulative hazards we first observe that for every $t > 0$,

$$\mathbb{E}(\tilde{h}_i(t)) = \int_{\mathbb{R}^+\times\mathbb{R}^+} \beta_i\, e^{-s\,b_i}\,\mathbb{1}_{[0,t]}(y)\,ds\,dy = t\,\frac{\beta_i}{b_i}.$$

Thus $\int_0^t \mathbb{E}(\tilde{h}_i(s))\,ds = t^2\,\frac{\beta_i}{2\,b_i}$ diverges as $t \to +\infty$, and condition (10) holds. The results in Theorem 7 apply and since the densities of (17) are ordered, we easily derive the expression for $\mathcal{W}(\tilde{h}_1(t), \tilde{h}_2(t))$ from the expression of $\mathbb{E}(\tilde{h}_i(t))$, in accordance with the results in Proposition 4 for gamma CRMs. □

The choice of the kernel allows for great flexibility and usually depends on the type of experiment one is considering. For example, if we are dealing with the failure of objects whose material wears out in time, the assumption of increasing hazard rates appears to be the most plausible. Besides the kernel of Dykstra and Laud [17], which leads to almost surely increasing hazard rates, one can resort to other options such as:

(a) Rectangular kernel with threshold τ: $k(t\,|\,x) = \mathbb{1}_{[t-\tau,t+\tau]}(x)$;
(b) Bathtub or U-shaped kernel with minimum $\eta > 0$: $k(t\,|\,x) = \mathbb{1}_{[0,|t-\eta|]}(x)$;
(c) Ornstein-Uhlenbeck kernel with $g > 0$: $k(t\,|\,x) = \sqrt{2g}\,e^{-g(t-x)}\mathbb{1}_{[0,t]}(x)$;
(d) Exponential kernel: $k(t\,|\,x) = e^{-tx}$.

More details can be found in Lo and Weng [37], James [31], De Blasi, Peccati and Prünster [12]. The choice of the kernel is typically driven by the type of data one is examining. As for the choice of the random measure, this may be dictated by specific inferential properties but it is usually motivated by analytical tractability and prior flexibility. In this regards, the gamma CRM is a popular alternative. We thus pick one of the kernels above and focus on gamma kernel mixtures. One is, then, left with the choice of the parameter b, which heuristically quantifies the prior belief on the steepness of the hazard. Given these specifications, one may be interested in quantifying the discrepancy induced by it on the corresponding hazards. Before proceeding, we underline how the same reasoning could be applied to the base measure, but for simplicity we consider gamma CRMs with a given shared base measure. We point out that in all cases we achieve the exact expression for the Wasserstein distance between the hazard rates.

Example 4. Let $\tilde{\mu}_i \sim \mathrm{Ga}(b_i, \mathrm{Leb}^+)$, with $b_i > 0$, for $i = 1, 2$. Let $k_1 = k_2 = k$ be one of the kernels (a)–(d) above. Then the Wasserstein distances between the corresponding hazard rates mixtures equal

(a′) $\mathcal{W}(\tilde{h}_1(t), \tilde{h}_2(t)) = (2\tau - (\tau - t)^+)\,|b_1^{-1} - b_2^{-1}|$;
(b′) $\mathcal{W}(\tilde{h}_1(t), \tilde{h}_2(t)) = |t - \eta|\,|b_1^{-1} - b_2^{-1}|$;
(c′) $\mathcal{W}(\tilde{h}_1(t), \tilde{h}_2(t)) = g\sqrt{2g}(1 - e^{-gt})\,|b_1^{-1} - b_2^{-1}|$;
(d′) $\mathcal{W}(\tilde{h}_1(t), \tilde{h}_2(t)) = t^{-1}(1 - e^{-t^2})\,|b_1^{-1} - b_2^{-1}|$,

where $f^+ = \max(f, 0)$ for any measurable function f with values in $\mathbb{R}$.

Proof. Kernel (a) is very similar to the one in Example 3. The Lévy intensity

$$\frac{1}{k_i(t\,|\,y)}\nu_i\Big(d\frac{s}{k_i(t\,|\,y)}, dy\Big) = \frac{e^{-s\,b_i}}{s}\,\mathbb{1}_{(0,+\infty)}(s)\,\mathbb{1}_{[0\wedge(t-\tau),t+\tau]}(y)\,ds\,dy$$

is the one of a gamma CRM with parameter b and Lebesgue base measure on $[0 \wedge (t - \tau), t + \tau]$. Since the corresponding densities are ordered, the exact Wasserstein distance is available and coincides with (a′). The same is true for kernel (b). With kernel (c) one has

$$\frac{1}{k_i(t\,|\,y)}\nu_i\Big(d\frac{s}{k_i(t\,|\,y)}, dy\Big) = \frac{1}{s}\exp\Big\{-\frac{sb_i}{\sqrt{2g}}e^{g(t-y)}\Big\}\mathbb{1}_{(0,+\infty)}(s)\,\mathbb{1}_{[0,t]}(y)\,dy\,ds.$$

The corresponding densities are ordered, thus if the conditions of Theorem 7 hold we only need to evaluate the expected value of the hazards to derive the exact Wasserstein distance:

$$\begin{aligned}\mathbb{E}(\tilde{h}_i(t)) &= \int_{\mathbb{R}^+\times\mathbb{R}} \sqrt{2g}\,e^{-g(t-y)}e^{-b_i s}\,\mathbb{1}_{[0,t]}(y)\,dy\,ds\\ &= \int_0^t -\frac{\sqrt{2g}}{b_i}e^{-g(t-y)}dy = \sqrt{\frac{2}{g}}(1 - e^{gt})\frac{1}{b_i}.\end{aligned}$$

This also proves the finite mean condition (13). Since $\int_0^t (1 - e^{gs})ds = \frac{1-e^{gt}}{g} + t$ diverges as $t \to +\infty$, also condition (10) holds.

Finally for kernel (d),

$$\frac{1}{k_i(t\,|\,y)}\nu_i\Big(d\frac{s}{k_i(t\,|\,y)}, dy\Big) = \frac{1}{s}\exp\big\{-sbe^{ty}\big\}\mathbb{1}_{(0,+\infty)}(s)\,\mathbb{1}_{[0,t]}(y)\,dy\,ds.$$

The mean hazard rates are

$$\mathbb{E}(\tilde{h}_i(t)) = \int_0^t \frac{e^{-ty}}{b_i} dy = \frac{1-e^{-t^2}}{b_i t},$$

and thus condition (13) holds. Moreover,

$$\int_0^t \frac{1-e^{-s^2}}{s} ds = \frac{\gamma}{2} + \frac{E_1(t^2)}{2} + \log(t),$$

where γ is the Euler gamma constant. This quantity diverges as $\log(t)$ for $t \to +\infty$ and thus condition (10) holds. We conclude as in the previous cases. $\square$

Next, we apply the bounds on cumulative hazards and survival functions of Theorem 8 and Theorem 9 to the case where mixtures of gamma CRMs are used, as in Example 3.

Example 5. Consider the prior specification in Example 3. Denote by $\tilde{H}_i$ the corresponding cumulative process (15) and by $\tilde{S}_i$ the corresponding survival process (16), for $i = 1, 2$. Then for every $t \geq 0$,

$$\mathcal{W}(\tilde{H}_1(t), \tilde{H}_2(t)) = \frac{t^2}{2}\left|\frac{\beta_1}{b_1} - \frac{\beta_2}{b_2}\right|, \tag{18}$$

$$g_\ell(\boldsymbol{b}, t) \leq \mathcal{W}(\tilde{S}_1(t), \tilde{S}_2(t)) \leq g_{u,1}(\boldsymbol{b}, t) \wedge g_{u,2}(\boldsymbol{b}, t), \tag{19}$$

where

$$g_\ell(\boldsymbol{b}, t) = e^t \left| \left(\frac{b_1}{b_1 + \beta_1 t}\right)^{\frac{b_1+\beta_1 t}{\beta_1}} - \left(\frac{b_2}{b_2 + \beta_2 t}\right)^{\frac{b_2+\beta_2 t}{\beta_2}} \right|,$$

$$g_{u,1}(\boldsymbol{b}, t) = 1 - e^{-\frac{t^2}{2}\left|\frac{\beta_1}{b_1} - \frac{\beta_2}{b_2}\right|},$$

$$g_{u,2}(\boldsymbol{b}, t) = e^t \left(\left(\frac{b_1}{b_1 + \beta_1 t}\right)^{\frac{b_1+\beta_1 t}{\beta_1}} + \left(\frac{b_2}{b_2 + \beta_2 t}\right)^{\frac{b_2+\beta_2 t}{\beta_2}} \right) \\ - \left(e^{-\frac{t^2\beta_1}{2b_1}} + e^{-\frac{t^2\beta_2}{2b_2}}\right) e^{-\frac{t^2}{2}\left|\frac{\beta_1}{b_1} - \frac{\beta_2}{b_2}\right|}.$$

Proof. Since the Lévy densities of

$$\frac{1}{K_i(t\,|\,y)} \nu_i\Big(d\frac{s}{K_i(t\,|\,y)}, dy\Big) = \frac{1}{s} \exp\Big\{-\frac{s b_i}{t-y}\Big\} \mathbb{1}_{(0,+\infty)}(s)\, \mathbb{1}_{[0,t]}(y)\, dy\, ds$$

are ordered, if the conditions of Theorem 8 hold the expression for the Wasserstein distance between the cumulative hazards easily derives from

$$\mathbb{E}(\tilde{H}_i(t)) = \int_{\mathbb{R}^+ \times [0,+\infty)} K_i(t\,|\,y)\, s\, \nu_i(ds, dy) = \int_{\mathbb{R}^+} \int_0^t \beta_i(t-y)\, e^{-s b_i} ds\, dy = \frac{t^2 \beta_i}{2 b_i}. \tag{20}$$

Now, condition (10) on the kernels has already been checked in Example 3. Moreover, (20) proves condition (13) on the finite mean.

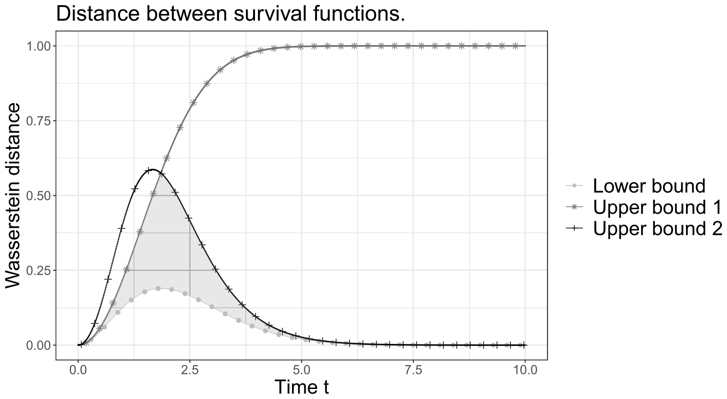

Fig 2. *Theoretical upper and lower bounds for the Wasserstein distance between marginals of the random survival functions in Example 5 with* $b_1 = 1$*,* $\beta_1 = 1$*,* $b_2 = 2$ *and* $\beta_2 = 1$*.*

As for the Wasserstein distance between the survival functions, in order to apply Theorem 9 it suffices to evaluate the mean of the survival functions. This is easily done thanks to the properties of the Laplace functional of a CRM (3). Specifically, $\mathbb{E}\big(e^{-\int_{\mathbb{R}} K(t\,|\,y)\,\tilde{\mu}_i(dy)}\big)$ is equal to

$$\exp\left\{-\int_{\mathbb{R}^+\times[0,+\infty)}\left(1-e^{s\,K_i(t\,|\,y)}\right)\nu_i(ds,dy)\right\}=\left(\frac{b_i}{b_i+\beta_i t}\right)^{\frac{b_i+\beta_i t}{\beta_i}}. \qquad \square$$

In Figure 2 a graphical representation of the upper and lower bounds for the Wasserstein distance between the corresponding survival functions is given. In particular, the distance between the survival functions lies in the gray area in the figure. The first upper bound $g_{u,1}$ appears to be tighter for small times, while the second $g_{u,2}$ is more informative as time increases. This depends on the fact that in the first case we are using the bound $e^{-\tilde{H}_1(t)\wedge\tilde{H}_2(t)} \leq 1$, which is effective for small values of the cumulative hazard function, i.e. for small times, while in the second one we are using $e^{-\tilde{H}_1(t)\wedge\tilde{H}_1(t)} \leq e^{-\tilde{H}_1(t)} + e^{-\tilde{H}_2(t)}$, which is effective for large values of the cumulative hazard function, i.e. for large t. Moreover, we point out that the Wasserstein distance between survival functions is considerably smaller than the one between the hazard rates, which is what we expect from a modeling perspective.

5. Posterior sampling scheme

The techniques we have developed in the previous sections may be fruitfully applied to evaluate approximation errors in posterior sampling schemes. In this section we focus on the gamma kernel mixture by Dykstra and Laud [17] and rely on the posterior analysis by James [32]. Even when the prior hazards are modeled as a gamma process (i.e. constant β), conditionally on the data and a set of latent variables, the non-atomic part of the posterior hazards is an extended gamma process. There are many available methods in the literature to sample from an extended gamma process, as the finite dimensional approximation by Ishwaran and James [30], the inverse Lévy methods of Ferguson and Klass [18] and Wolpert and Ickstadt [52], and the series representation of Bondesson [4], which serves as a basis for the algorithm in Laud, Smith and Damien [34]. Other available series representations can be found in Rosiński [47]. Recently, Al Masry, Mercier and Verdier [1] proposed a new algorithm based on a discretization of the scale function: in such case the extended gamma process can be approximated by a sum of gamma random increments. The construction of the discretization is not always simple but, when possible, it allows for a precise quantification of the approximation error. In Al Masry, Mercier and Verdier [1] the error is quantified through a bound on the L_2 distance. Here, we build the discrete approximation of the scale function of the posterior hazards corresponding to a gamma process prior and use the Wasserstein distance to quantify the approximation error between the induced hazard rates. Moreover, since one is usually interested in the cumulative hazards or in the survival function, we provide an estimate for their approximations as well, which yields a novel and meaningful guide for fixing the approximation error in the algorithm.

We first recall the posterior characterization of mixture hazard rates models, with censored data, as achieved by James [32]. This result is suited to our case, since it concerns CRM-driven mixtures under a multiplicative intensity model. In order to provide a summary description of the posterior distribution, henceforth $T_1, \dots, T_n$ are random elements from an exchangeable sequence as in (9) with Π being the law of a random probability measure with hazard rate $\tilde{h}$ as in (11). Furthermore, if $n_e = \sum_{i=1}^n \Delta_i$ is the number of exact observations in the sample, we may assume without loss of generality that $\Delta_1 = \cdots = \Delta_{n_e} = 1$ and, hence, the last $n - n_e$ observations are censored. The data are, then, given by $\{(x_j, \Delta_j)\}_{j=1}^n$. A representation of the likelihood function that is convenient for Bayesian computations is obtained by relying on a suitable augmentation that involves a collection of latent variables $Y_1, \dots, Y_{n_e}$ corresponding to the exact observations. Hence, the augmented likelihood is given by

$$\begin{aligned}\mathcal{L}(\tilde{\mu}; \boldsymbol{x}, \boldsymbol{y}) &= e^{-\int_{\mathbb{X}} K_n(y)\tilde{\mu}(dy)} \prod_{j=1}^{n_e} \tilde{\mu}(dy_j) k(x_j \,|\, y_j) \\ &= e^{-\int_{\mathbb{X}} K_n(y)\tilde{\mu}(dy)} \prod_{h=1}^{k} \tilde{\mu}(dy_h^*)^{n_h} \prod_{i \in C_h} k(x_i \,|\, y_h^*),\end{aligned}$$

where $\boldsymbol{x} = (x_1, \dots, x_n)$, $y_1^*, \cdots, y_k^*$ are the $k \leq n_e$ distinct values in $\boldsymbol{y} = (y_1, \dots, y_{n_e})$, $C_j = \{r : y_r = y_j^*\}$ and $n_j = \mathrm{card}(C_j)$. The function K_n is interpretable as a kernel for the cumlative hazards and, in general, accounts for different forms of censoring. For simplicity we henceforth focus on the case of right-censored observations and this yields

$$K_n(y) = \sum_{j=1}^{n} \int_0^{x_j} k(u \,|\, y)\, du. \tag{21}$$

The posterior characterization we rely on is as follows.

Theorem 10 (James [32]). *Let $T_1, \dots, T_n$ be random elements from an exchangeable sequence as in (9), with Π being the law of a random probability measure with hazard rate $\tilde{h}$ as in (11). Conditional on the observed data $\boldsymbol{x} = (x_1, \cdots, x_n)$ and latent variables $\boldsymbol{y} = (y_1, \cdots, y_{n_e})$, $\tilde{\mu}$ equals in distribution*

$$\tilde{\mu}^* \overset{d}{=} \tilde{\mu}_c^* + \sum_{h=1}^{k} J_h \delta_{y_h^*}, \tag{22}$$

where $\tilde{\mu}_c^*$ is a CRM *without fixed jump points and with intensity*

$$\nu^*(ds, dy) = e^{-sK_n(y)} \nu(ds, dy) = e^{-sK_n(y)} \rho_y(ds)\, \eta(dy),$$

while $J_1, \cdots, J_k$ are mutually independent and independent from $\tilde{\mu}_c^*$. For $h = 1, \dots, k$, the generic h-th jump J_h has distribution

$$G_h(ds) \propto s^{n_h} e^{-sK_n(y_h^*)} \rho_{y_h^*}(ds). \tag{23}$$

In the rest of the section we focus on the case $\mathbb{X} = \mathbb{R}$, $k(t|y) = \beta \mathbb{1}_{[0,t]}(y)$ and $\tilde{\mu}$ gamma CRM with rate parameter b and base measure α, which is a typical choice in applications. Thus the non-atomic posterior CRM $\tilde{\mu}_c^*$ has Lévy intensity

$$\nu^*(ds, dy) = \frac{e^{-s(b+\beta \sum_{i=1}^{n} (y-x_i)^+)}}{s} \mathbb{1}_{(0,+\infty)}(s)\, ds\, \alpha(dy).$$

It follows that $\tilde{\mu}_c^*$ is an extended gamma CRM with base measure α and scale function $1/(b + \beta \sum_{i=1}^{n} (y - x_i)^+)$. The non-atomic posterior hazards are an extended gamma process and can thus be written as

$$\tilde{h}^*(t) \overset{\mathrm{d}}{=} \int_0^t \beta^*(s)\, \tilde{\mu}(ds), \tag{24}$$

where $\beta^*(y) = \beta/(b + \beta \sum_{i=1}^{n} (y - x_i)^+)$ and $\tilde{\mu}$ is a gamma CRM with parameter 1 and base measure α.

Consider an interval of interest $[0, T]$, which can be thought of as the initial and final time of the study, so that $0 < x_1 \leq \cdots \leq x_n < T$. The algorithm proposed by Al Masry, Mercier and Verdier [1] to sample from $\{\tilde{h}^*(t) \,|\, t \in [0, T]\}$ is based on a piecewise constant approximation of β^* on the interval $[0, T]$. If $\beta^\epsilon(y) = \sum_{h=0}^{n(\epsilon)} \beta_h \mathbb{1}_{(t_h, t_{h+1}]}(y)$, then for every $t \geq 0$,

$$\tilde{h}^\epsilon(t) = \int_0^t \beta^\epsilon(s)\, \tilde{\mu}(ds) = \sum_{h=1}^{n_t} \beta_h\, \tilde{\mu}(t_h, t_{h+1}] + \beta_{n_t+1}\, \tilde{\mu}(t_{n_t}, t], \tag{25}$$

where n_t is such that $t_{n_t} \leq t \leq t_{n_t+1}$. The increments $\delta_h = \beta_h\, \tilde{\mu}(t_h, t_{h+1}]$ have a gamma distribution with scale β_h and shape $\alpha(t_h, t_{h+1})$. If the points $\{t_h \,|\, h = 1, \dots, n(\epsilon)\}$ are dense in the interval $[0, T]$ as $n(\epsilon) \to +\infty$, one samples directly from a sum of gamma random variables $\sum_{t_h \leq t} \delta_h$.

In order to apply this algorithm we need to build an approximating strictly positive piecewise constant function $\beta^\epsilon : [0, T] \to (0, +\infty)$ and find a reasonable

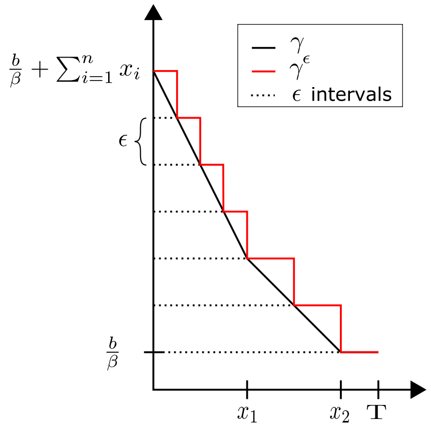

FIG 3. *Piecewise constant approximation γ^ϵ of the function γ on the interval $[0,T]$.*

criterion to fix the approximation error. We will build β^ϵ by discretizing the reciprocal of β^*, namely $\gamma(y) = b\,\beta^{-1} + \sum_{i=1}^n (y-x_i)^+$. Consider the points $t_0 \le t_1 \le \cdots \le t_{n(\epsilon)-1} = x_n \le t_{n(\epsilon)} = T$ defined by

$$t_{j+\sum_{i=0}^{j-1}[\epsilon^{-1}(n-i)(x_{i+1}-x_i)]+k} = x_j + \frac{k\,\epsilon}{n-j},$$

for every $j = 0,\dots,n-1$ and $k = 0,\dots,\big[(n-j)(x_{j+1}-x_j)\epsilon^{-1}\big]$, where $x_0 = 0$ and $[x]$ denotes the integer part of $x \ge 0$, so that $n(\epsilon) = n+1+\sum_{i=0}^{n-1}\big[(n-i)(x_{i+1}-x_i)\epsilon^{-1}\big]$. We observe that

$$\gamma\Big(t_{j+\sum_{i=0}^{j-1}[\epsilon^{-1}(n-i)(x_{i+1}-x_i)]+k}\Big) = \frac{b}{\beta} + \sum_{i=j+1}^{n} x_i - (n-j)x_j - k\,\epsilon.$$

Next we set $\gamma_h = \gamma(t_{h+1})$ and define

$$\gamma^\epsilon(y) = \sum_{h=0}^{n(\epsilon)} \gamma_i \mathbb{1}_{(t_h,t_{h+1}]}(y). \tag{26}$$

Theorem 11. *The function defined in* (26) *is piecewise constant and satisfies*

$$\sup_{y\in([0,T]} |\gamma(y)-\gamma^\epsilon(y)| \le \epsilon.$$

Moreover, $\gamma^\epsilon(y) \ge \gamma(y)$ for every $y \in [0,T]$.

From this theorem one deduces a very simple uniform bound for the discrepancy between β^ϵ and β^*.

Corollary 12. *If* $\beta^\epsilon(y) = \frac{1}{\gamma^\epsilon(y)} = \sum_{h=0}^{n(\epsilon)} \frac{1}{\gamma_h} \mathbb{1}_{(t_h, t_{h+1}]}(y)$*, then one has*

$$\sup_{y \in (0,T]} |\beta^*(y) - \beta^\epsilon(y)| \leq \frac{\beta^2}{b^2}\epsilon. \tag{27}$$

For a given ϵ, these results provide a constructive rule for determining an approximation of β^*, and hence an approximating hazard $\tilde{h}^\epsilon$. One may then wonder which value of ϵ should be specified to achieve a prescribed error of approximation for the posterior hazards and survivals. This is achieved in Corollary 13, where we propose three different rules based on the Wasserstein distance between the hazards, cumulative hazards and the survival functions respectively. The result is stated for the hypotheses of Example 3, $\alpha = Leb^+$, but can be easily adapted to any base measure.

Corollary 13. *Consider the hypotheses of* Theorem 10 *with* $k(y|t) = \beta \mathbb{1}_{(0,t]}(y)$ *and* $\tilde{\mu} \sim \mathrm{Ga}(b, \mathrm{Leb}^+)$*. Let* $\tilde{h}^* = \{\tilde{h}^*(t) \,|\, t \in [0,T]\}$ *be the non-atomic posterior hazard rates process* (24)*, and let* $\tilde{h}^\epsilon = \{\tilde{h}^\epsilon(t) \,|\, t \in [0,T]\}$ *be its approximation* (25)*. If* $\tilde{H}$*,* $\tilde{H}^\epsilon$*,* $\tilde{S}$*,* $\tilde{S}^\epsilon$ *denote their respective cumulative hazards and survival functions processes, then*

$$\sup_{t \in (0,T]} \mathcal{W}(\tilde{h}^*(t), \tilde{h}^\epsilon(t)) \leq \epsilon \frac{\beta^2}{b^2} T,$$

$$\sup_{t \in (0,T]} \mathcal{W}(\tilde{H}^*(t), \tilde{H}^\epsilon(t)) \leq \epsilon \frac{\beta^2}{2b^2} T^2,$$

$$\sup_{t \in (0,T]} \mathcal{W}(\tilde{S}^*(t), \tilde{S}^\epsilon(t)) \leq 1 - \exp\left\{ - \epsilon \frac{\beta^2}{2b^2} T^2 \right\}.$$

6. Proofs

6.1. Proof of Theorem 1

First of all we state a technical lemma.

Lemma 14. *Let* $\tilde{\mu}$ *be a* CRM *with Lévy intensity* ν *and finite mean* (6)*. Then for every* $A \in \mathcal{X}$*,*

$$\lim_{\epsilon \to 0^+} \epsilon\, \nu([\epsilon, +\infty) \times A) = 0.$$

Proof. For every $\delta > 0$ consider $\epsilon > 0$ such that $\epsilon < \delta$. Then

$$\epsilon\, \nu([\epsilon, +\infty) \times A) = \epsilon \int_\epsilon^\delta \int_A \nu(ds, dy) + \epsilon \int_\delta^{+\infty} \int_A \nu(ds, dy).$$

The second integral is finite by (2), thus $\epsilon \int_\delta^{+\infty} \int_A \nu(ds, dy) \to 0$ as $\epsilon \to 0$. As for the first one, this can be bounded by

$$\epsilon \int_\epsilon^\delta \int_A \nu(ds, dy) \leq \int_\epsilon^\delta \int_A s\, \nu(ds, dy).$$

Since the integrand is integrable in $[0,\delta]$ thanks to the finite mean condition (6), by the dominated convergence theorem,

$$\limsup_{\epsilon\to 0} \epsilon \int_\epsilon^\delta \int_A \nu(ds,dy) \leq \int_0^\delta \int_A s\,\nu(ds,dy).$$

Since this is true for every $\delta > 0$, by the absolute continuity of the integral $\epsilon\,\nu([\epsilon,+\infty)\times A) \to 0$ as $\epsilon \to 0$. □

We now prove the results in Theorem 1. The lower bound $g_\ell(A)$ is easily achieved by (5) and by Campbell's Theorem applied to the underlying Poisson random measures with respect to the measurable function $f(s,x) = s\,\mathbb{1}_A(x)$, similarly to (6). We thus concentrate on the upper bound.

Since the Lévy intensities are diffuse and infinitely active, for every $A \in \mathcal{X}$ and $r > 0$ there exists $\epsilon_{i,r,A} > 0$ such that

$$\nu_i([\epsilon_{i,r,A},+\infty)\times A) = r, \tag{28}$$

for $i = 1,2$. By denoting with $\mathcal{N}_i$ the Poisson random measure underlying $\tilde\mu_i$ as in (1),

$$\tilde\mu_i(A) \overset{\mathrm{d}}{=} \int_0^{\epsilon_{i,r,A}} \int_A s\,\mathcal{N}_i(ds,dy) + \int_{\epsilon_{i,r,A}}^{+\infty} \int_A s\,\mathcal{N}_i(ds,dy).$$

We use the notation $J^S_{i,r,A} = \int_0^{\epsilon_{i,r,A}} \int_A s\,\mathcal{N}_i(ds,dy)$ for the small jumps and $J^B_{i,r,A} = \int_{\epsilon_{i,r}}^\infty \int_A s\,\mathcal{N}_i(ds,dy)$ for the big jumps. The independence of the increments of a Poisson random measure ensures that $J^S_{i,r,A}$ and $J^B_{i,r,A}$ are independent, thus by (8)

$$\mathcal{W}(\tilde\mu_1(A),\tilde\mu_2(A)) \leq \mathcal{W}(J^S_{1,r,A},J^S_{2,r,A}) + \mathcal{W}(J^B_{1,r,A},J^B_{2,r,A}).$$

We first show that the small jumps do not play any role in the final bound. By (5),

$$\mathcal{W}(J^S_{1,r,A},J^S_{2,r,A}) \leq \mathbb{E}(J^S_{1,r,A}) + \mathbb{E}(J^S_{2,r,A}).$$

The means $\mathbb{E}(J^S_{i,r,A}) = \int_0^{\epsilon_{i,r,A}} s\,\nu_i(ds\times A)$ are finite by (2) and thus go to zero as $r \to +\infty$ by the absolute continuity of the integral. We now focus on the big jumps. By (2), these are integrals of Poisson random measures with finite mean measure, $\nu_i(ds,dy)\,\mathbb{1}_{[\epsilon_{i,r,A},+\infty)}(s)$. Proposition 19.5 in Sato [48] then ensures that $J^B_{i,r,A}$ has a compound Poisson distribution, so that

$$J^B_{i,r,A} \overset{\mathrm{d}}{=} \sum_{h=1}^{N_{i,r,A}} \xi^h,$$

where $N_{i,r,A}$ is a Poisson random variable with intensity $r = \nu_i([\epsilon_{i,r,A},+\infty)\times A)$ and $(\xi^h)_h$ are independent and identically distributed random variables, independent from $N_{i,r,A}$, with distribution

$$\rho_{i,r,A}(ds) = \frac{1}{r}\int_A \nu_i(ds,dy)\,\mathbb{1}_{[\epsilon_{i,r,A},+\infty)}(s). \tag{29}$$

Theorem 10 in Mariucci and Reiß [39] deals with the Wasserstein distance between compound Poisson distributions. Since $J^B_{1,r,A}$ and $J^B_{2,r,A}$ have the same total intensity measure r but different jump distribution $\rho_{i,r,A}$, an immediate adaptation of their result yields

$$\mathcal{W}(J^B_{1,r,A}, J^B_{2,r,A}) \leq r\,\mathcal{W}(\rho_{1,r,A}, \rho_{2,r,A}).$$

By (7), $\mathcal{W}(\rho_{1,r,A}, \rho_{2,r,A}) = \int_{-\infty}^{+\infty} |F_{1,r,A}(u) - F_{2,r,A}(u)|\,du$, where $F_{i,r,A}(u)$ is equal to

$$\frac{1}{r}\nu_i([\epsilon_{i,r,A}, u] \times A)\,\mathbb{1}_{[\epsilon_{i,r,A},+\infty)}(u) = \left(1 - \frac{1}{r}\nu_i((u,+\infty) \times A)\right)\mathbb{1}_{[\epsilon_{i,r,A},+\infty)}(u).$$

Define min $\in \{1,2\}$ such that $\epsilon_{\min} = \epsilon_{\min,r,A} = \epsilon_{1,r,A} \wedge \epsilon_{2,r,A}$ and similarly max $\in \{1,2\}$. Then

$$\mathcal{W}(\rho_{1,r,A}, \rho_{2,r,A}) = \int_{\epsilon_{\min}}^{\epsilon_{\max}} F_{\min,r,A}(u)\,du + \int_{\epsilon_{\max}}^{+\infty} |F_{1,r,A}(u) - F_{2,r,A}(u)|\,du.$$

Now, $r\int_{\epsilon_{\min}}^{\epsilon_{\max}} F_{\min,r,A}(u)\,du = \int_{\epsilon_{\min}}^{\epsilon_{\max}} \nu_{\min}([\epsilon_{\min}, u) \times A)\,dy \leq (\epsilon_{\max} - \epsilon_{\min})\,r$, which can be rewritten as $\epsilon_{\max}\nu_{\max,x}([\epsilon_{\max},+\infty)) - \epsilon_{\min}\nu_{\min,x}([\epsilon_{\min},+\infty))$. Thus by Lemma 14 it converges to zero as r goes to $+\infty$. On the other hand,

$$r\int_{\epsilon_{\max}}^{+\infty} |F_{1,r,A}(u) - F_{2,r,A}(u)|\,du = \int_{\epsilon_{\max}}^{+\infty} |\nu_1((u,+\infty)\times A) - \nu_2((u,+\infty)\times A)|\,du,$$

which attains the expression for $g_u(A)$ as r goes to $+\infty$.

6.2. *Proof of Corollary 2*

For every $u \in \mathbb{R}^+$, $\nu_i((u,+\infty) \times A) \leq \nu_j((u,+\infty) \times A)$ because the Radon–Nikodym derivatives are ordered. Thus $g_u(A)$ is equal to

$$\begin{aligned}&\left|\int_0^{+\infty} (\nu_1((u,+\infty) \times A) - \nu_2((u,+\infty) \times A))\,du\right|\\ &= \left|\int_0^{+\infty}\int_u^{+\infty} (\nu_1(ds \times A) - \nu_2(ds \times A))\,du\right|.\end{aligned}$$

By interchanging the integrals this is equal to the lower bound in Theorem 1.

6.3. *Proof of Corollary 3*

Without loss of generality we assume $\alpha_1(A) \leq \alpha_2(A)$. Then by taking $\eta(ds) = \mathbb{1}_{(0,+\infty)}(s)\,ds$,

$$\nu_1(s \times A) = \alpha_1(A)\,\rho(s) \leq \alpha_2(A)\,\rho(s) = \nu_2(s \times A),$$

for every $s \in \mathbb{R}^+$. Thus $\mathcal{W}(\tilde{\mu}_1(A), \tilde{\mu}_2(A)) = |\alpha_1(A) - \alpha_2(A)| \int_{\mathbb{R}^+} s\,\rho(s)\,ds$ by Corollary 2. We conclude by taking the supremum over $A \in \mathcal{X}$.

6.4. Proof of Proposition 4

Let $\tilde{\mu}_i \sim \mathrm{Ga}(b_i, \alpha_i)$, for $i = 1, 2$. Without loss of generality we assume $0 < b_1 \leq b_2$. Thus for every $A \in \mathcal{X}$,

$$\mathbb{E}(\tilde{\mu}_i(A)) = \int_{\mathbb{R}^+ \times A} s\, \nu_i(ds, dy) = \frac{\alpha_i(A)}{b_i}.$$

This implies that the CRM has finite mean. Since $\frac{e^{-sb_i}}{s}$ is not integrable near zero, the random measures are infinitely active. Thus Theorem 1 holds and from the expression of $\mathbb{E}(\tilde{\mu}_i(A))$ above we derive the lower bound in Proposition 4. We now focus on the upper bound. For every $u \in \mathbb{R}^+$,

$$\nu_i((u, +\infty) \times A) = \alpha_i(A) \int_u^{+\infty} \frac{e^{-sb_i}}{s} ds = \alpha_i(A)\, E_1(b_i u),$$

where $E_1(x) = \int_x^\infty \frac{e^{-y}}{y} dy$. Thus,

$$\mathcal{W}(\tilde{\mu}_1(A), \tilde{\mu}_2(A) \leq \int_0^{+\infty} |\alpha_1(A)\, E_1(b_1 u) - \alpha_2(A)\, E_1(b_2 u)|\, du.$$

The fundamental theorem of line integral ensures that

$$\int_0^{+\infty} |\alpha_1(A)\, E_1(b_1 u) - \alpha_2(A)\, E_1(b_2 u)|\, du = \int_0^{+\infty} \Big| \int_C \nabla \psi_u(a, b) \cdot ds \Big| du,$$

where ∇ denotes the gradient of a function, $\psi_u(a, b) = a\, E_1(by)$, and C is the segment in $\mathbb{R}^2$ connecting $(\alpha_1(A), b_1)$ to $(\alpha_2(A), b_2)$. We consider the parametrization $s(t) = (\alpha_1(A) + t(\alpha_2(A) - \alpha_1(A)), b_1 + t(b_2 - b_1))$. Since $\nabla \psi_u(a, b) = (E_1(by), -\frac{a}{b} e^{-by})$, this is equal to

$$\begin{aligned}
&\int_0^{+\infty} \Big| \int_0^1 \Big(E_1(s_2(t)u)\, s_1'(t) - \frac{s_1(t)}{s_2(t)} e^{-s_2(t)u}\, s_2'(t) \Big) dt \Big| du \\
&\leq \int_0^{+\infty} \int_0^1 \Big| E_1((b_1 + t(b_2 - b_1))u)\, (\alpha_2(A) - \alpha_1(A)) - \\
&\qquad \frac{\alpha_1(A) + t(\alpha_2(A) - \alpha_1(A))}{b_1 + t(b_2 - b_1)}\, e^{-(b_1 + t(b_2 - b_1))u}\, (b_2 - b_1) \Big| dt\, du.
\end{aligned}$$

Since we have assumed w.l.o.g. that $b_1 \leq b_2$,

$$\begin{aligned}
&\leq \int_0^{+\infty} \int_0^1 \Big(E_1((b_1 + t(b_2 - b_1))u)\, |\alpha_2(A) - \alpha_1(A)| + \\
&\qquad \frac{\alpha_1(A) + t(\alpha_2(A) - \alpha_1(A))}{b_1 + t(b_2 - b_1)} e^{-(b_1 + t(b_2 - b_1))u}\, (b_2 - b_1) \Big) dt\, du.
\end{aligned}$$

We invert the integrals thanks to Fubini's Theorem and use the fact that $\int_0^{+\infty} E_1(ax)dx = \frac{1}{a}$, which is a standard result on exponential integrals [22]. Thus,

$$\leq \int_0^1 \Big(\frac{1}{b_1 + t(b_2 - b_1)} \, |\alpha_2(A) - \alpha_1(A)| + \frac{\alpha_1(A) + t(\alpha_2(A) - \alpha_1(A))}{(b_1 + t(b_2 - b_1))^2} \, (b_2 - b_1)\Big) dt.$$

By standard integration techniques, this amounts to

$$= \frac{\alpha_1(A)}{b_1} - \frac{\alpha_2(A)}{b_2} + \mathbb{1}_{(0,+\infty)}(\alpha_2(A) - \alpha_1(A)) \, 2 \, \frac{\alpha_2(A) - \alpha_1(A)}{b_2 - b_1} \, \log \frac{b_2}{b_1}.$$

6.5. Proof of Proposition 5

When $c_1 = c_2$ the result is immediate once we observe that $\mathbb{E}(\tilde{\mu}(A)) = \alpha(A)$ for every $\tilde{\mu} \sim \mathrm{Be}(c, \alpha)$. We focus on the case $\alpha_1 = \alpha_2$, $0 < c_1 \leq c_2$. By reasoning as in the proof of Proposition 4, $\mathcal{W}(\tilde{\mu}_1(A), \tilde{\mu}_2(A)$ is upper bounded by

$$\alpha(A) \int_0^1 \int_0^s \int_{c_1}^{c_2} \left| \frac{d}{dc} \left(\frac{c\,(1-s)^{c-1}}{s} \right) \right| dc \, du \, ds$$

the derivative $\big((c\,(1-s)^{c-1})\, s^{-1}\big)' = (1-s)^{c-1}\,(1 + c\,\log(1-s)) s^{-1} \leq (1-s)^{c-1}\,(1 - c\,\log(1-s)) s^{-1}$ for $s \in (0,1)$. Thus by Fubini's Theorem the previous integral is upper bounded by

$$\alpha(A) \int_{c_1}^{c_2} \int_0^1 (1-s)^{c-1}\,(1 - c\,\log(1-s))\, ds \, dc = 2\alpha(A) \log\left(\frac{c_2}{c_1}\right),$$

by standard integration techniques.

6.6. Proof of Lemma 6

Let $\{A_1, \cdots A_n\}$ in $\mathcal{X}$ be disjoint sets. Then for $i = 1, \ldots, n$ the random variables $\tilde{\mu}_f(A_i) = \int_{A_i} f(x)\tilde{\mu}(dx)$ are independent since f is deterministic and $\tilde{\mu}(A_1), \cdots \tilde{\mu}(A_n)$ are independent. This proves that $\tilde{\mu}_f$ is a CRM. In order to find its Lévy intensity ν_f, we consider the Laplace functional transform (3):

$$\begin{aligned} \mathbb{E}\big(e^{-\int_{\mathbb{R}} g(y)\tilde{\mu}_f(dy)}\big) &= \exp\Big\{ - \int_{\mathbb{R}^+ \times \mathbb{R}} [1 - e^{-s\,g(y) f(y)}]\, \nu(ds, dy) \Big\} = \\ &= \exp\Big\{ - \int_{\mathbb{R}^+ \times \mathbb{R}} [1 - e^{-s\,g(y)}]\, (p_f \# \nu)(ds, dy) \Big\}, \end{aligned}$$

where $p_f(s, y) = (sf(y), y)$.

6.7. Proof of Theorem 9

The lower bound follows immediately from (5). As for the upper bounds, first of all we observe that for any $x, y \in \mathbb{R}^+$,

$$|e^{-x} - e^{-y}| = e^{-x \wedge y}(1 - e^{-|x-y|}). \tag{30}$$

In order to derive the function $g_{u,1}$ in the upper bound, we observe that since $e^{-x\wedge y} \leq 1$, the following upper bound holds for $\mathcal{W}(\tilde{S}_1(t), \tilde{S}_2(t))$:

$$\inf_{c(\tilde{H}_1(t), \tilde{H}_2(t))} \mathbb{E}\big(\big|e^{-\tilde{H}_1(t)} - e^{-\tilde{H}_2(t)}\big|\big) \leq \inf_{c(\tilde{H}_1(t), \tilde{H}_2(t))} \mathbb{E}\big(1 - e^{-|\tilde{H}_1(t) - \tilde{H}_2(t)|}\big).$$

Since $1-e^{-x}$ is a concave function, by using Jensen's Inequality, $\mathcal{W}(\tilde{S}_1(t), \tilde{S}_2(t))$ is upper bounded by

$$\inf_{c(\tilde{H}_1(t), \tilde{H}_2(t))} \Big\{1 - e^{-\mathbb{E}(|\tilde{H}_1(t) - \tilde{H}_2(t)|)}\Big\} = 1 - e^{-\inf_{c(\tilde{H}_1(t), \tilde{H}_2(t))} \mathbb{E}(|\tilde{H}_1(t) - \tilde{H}_2(t)|)}.$$

As for $g_{u,2}$, combining (30) with $e^{-x\wedge y} \leq e^{-x} + e^{-y}$ and by Jensen's Inequality,

$$\begin{aligned}
&\mathbb{E}\big(\big|e^{-\tilde{H}_1(t)} - e^{-\tilde{H}_2(t)}\big|\big) \\
&\leq \mathbb{E}\big(e^{-\tilde{H}_1(t)} + e^{-\tilde{H}_2(t)} - e^{-(\tilde{H}_1(t) + |\tilde{H}_1(t) - \tilde{H}_2(t)|)} - e^{-(\tilde{H}_2(t) + |\tilde{H}_1(t) - \tilde{H}_2(t)|)}\big) \\
&\leq \mathbb{E}\big(e^{-\tilde{H}_1(t)}\big) + \mathbb{E}\big(e^{-\tilde{H}_2(t)}\big) - e^{-\mathbb{E}(\tilde{H}_1(t) + |\tilde{H}_1(t) - \tilde{H}_2(t)|)} - e^{-\mathbb{E}(\tilde{H}_2(t) + |\tilde{H}_1(t) - \tilde{H}_2(t)|)} \\
&\leq \mathbb{E}\big(e^{-\tilde{H}_1(t)}\big) + \mathbb{E}\big(e^{-\tilde{H}_2(t)}\big) - \big(e^{-\mathbb{E}(\tilde{H}_1(t))} + e^{-\mathbb{E}(\tilde{H}_2(t))}\big)e^{-\mathbb{E}|\tilde{H}_1(t) - \tilde{H}_2(t)|}.
\end{aligned}$$

By taking the infimum over all couplings in $c(\tilde{H}_1(t), \tilde{H}_2(t))$ we derive $g_{u,2}$.

6.8. Proof of Theorem 11

The proof relies on $\gamma(y)$ being a decreasing continuous piecewise linear function. We first include $x_1, \dots, x_n$ in the set $\{t_h \,|\, h = 1, \dots, n(\epsilon)\}$. Then, for every $i = 1, \dots n$ we iteratively include all points $t \in (x_i, x_{i+1})$ such that the counterimage $\gamma^{-1}(t)$ is at distance ϵ from the previous point. We easily conclude by observing that on $(x_i, x_{i+1}]$ the function γ is linear with coefficient equal to $-(n-i)$. See Figure 3.

6.9. Proof of Corollary 13

The proof is based on observing that $\tilde{h}_1 = \tilde{h}^*$ and $\tilde{h}_2 = \tilde{h}^\epsilon$ are two kernel mixture hazards with $k_1(y|t) = k_2(y|t) = \mathbb{1}_{[0,t]}(y)$ and $\tilde{\mu}_i$ extended gamma CRMs with scale function $\beta_1(y) = \beta^*(y)$ and $\beta_2(y) = \beta^\epsilon(y)$ and Lebesgue base measure on the positive axis, i.e.

$$\nu_i(ds, dy) = \frac{\exp\big\{-\frac{s}{\beta_i(y)}\big\}}{s} \mathbb{1}_{[0,+\infty)}(y)\, ds\, dy.$$

These kernel and Lévy intensities satisfy both the conditions of Theorem 7 and of Theorem 8. Since by construction $\beta^\epsilon(y) \leq \beta^*(y)$ for every $y \in [0, +\infty)$, the Lévy densities of the hazards and of the cumulative hazards are ordered. Thus the Wasserstein distance reduces to the absolute difference of their means:

$$\mathcal{W}(\tilde{h}^*(t), \tilde{h}^\epsilon(t)) = \left| \int_0^t \beta^*(y) - \beta^\epsilon(y)\, dy \right| \leq \int_0^t |\beta^*(y) - \beta^\epsilon(y)|\, dy \leq \epsilon\, \frac{\beta^2}{b^2}\, t,$$

by (27). Similarly,

$$\mathcal{W}(\tilde{H}^*(t), \tilde{H}^\epsilon(t)) = \left| \int_0^t (t-y)(\beta^*(y) - \beta^\epsilon(y))\, dy \right| \leq \epsilon\, \frac{\beta^2}{b^2} \int_0^t (t-y) dy = \epsilon\, \frac{\beta^2}{2b^2}\, t^2.$$

Finally, the bound for the survival function derives directly from the one on the cumulative hazards, as in Theorem 9.

Acknowledgments

The authors are grateful to an Associate Editor and a Referee for insightful comments and suggestions, which led to a substantial improvement of the paper. Antonio Lijoi and Igor Prünster are partially supported by MIUR, PRIN Project 2015SNS29B.

References

[1] Al Masry, Z., Mercier, S. and Verdier, G. (2017). Approximate simulation techniques and distribution of an extended gamma process. *Methodology and Computing in Applied Probability* **19** 213–235. MR3611541

[2] Arbel, J., De Blasi, P. and Prünster, I. (2019). Stochastic approximations to the Pitman-Yor process. *Bayesian Analysis* **15** 1303–1356. MR4136558

[3] Bickel, P. J. and Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **9** 1196–1217. MR0630103

[4] Bondesson, L. (1982). On simulation from infinitely divisible distributions. *Advances in Applied Probability* **14** 855–869. MR0677560

[5] Campbell, T., Huggins, J. H., How, J. P. and Broderick, T. (2019). Truncated random measures. *Bernoulli* **25** 1256–1288. MR3920372

[6] Chen, J. (1995). Optimal rate of convergence for finite mixture models. *Ann. Statist.* **23** 221–233. MR1331665

[7] Cifarelli, D. M., Dolera, E. and Regazzini, E. (2016). Frequentistic approximations to Bayesian prevision of exchangeable random elements. *International Journal of Approximate Reasoning* **78**. MR3543878

[8] Cifarelli, D. M. and Regazzini, E. (2017). On the centennial anniversary of Gini's theory of statistical relations. *Metron* **75** 227-242. MR3695007

[9] Daley, D. J. and Vere-Jones, D. (2002). *An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods. Probability and Its Applications*. Springer. MR1950431

[10] Daley, D. J. and Vere-Jones, D. (2007). *An Introduction to the Theory of Point Processes: Volume II: General Theory and Structure. Probability and Its Applications*. Springer New York. MR2371524

[11] Dall'Aglio, G. (1956). Sugli estremi dei momenti delle funzioni di ripartizione doppia. *Annali della Scuola Normale Superiore di Pisa - Classe di Scienze* **10** 35-74. MR0081577

[12] De Blasi, P., Peccati, G. and Prünster, I. (2009). Asymptotics for posterior hazards. *Ann. Statist.* **37** 1906–1945. MR2533475

[13] De Iorio, M., Johnson, W. O., Müller, P. and Rosner, G. L. (2009). Bayesian nonparametric nonproportional hazards survival modeling. *Biometrics* **65** 762-771. MR2649849

[14] Doksum, K. (1974). Tailfree and neutral random probabilities and their posterior distributions. *Ann. Probab.* **2** 183–201. MR0373081

[15] Donnet, S., Rivoirard, V., Rousseau, J. and Scricciolo, C. (2018). Posterior concentration rates for empirical Bayes procedures with applications to Dirichlet process mixtures. *Bernoulli* **24** 231–256. MR3706755

[16] Dudley, R. M. (1976). *Probabilities and metrics: convergence of laws on metric spaces, with a view to statistical testing. Lecture notes series*. Aarhus Universitet, Matematisk Institut. MR0488202

[17] Dykstra, R. L. and Laud, P. (1981). A Bayesian nonparametric approach to reliability. *Ann. Statist.* **9** 356–367. MR0606619

[18] Ferguson, T. S. and Klass, M. J. (1972). A representation of independent increment processes without Gaussian components. *Ann. Math. Statist.* **43** 1634–1643. MR0373022

[19] Flamary, R. and Courty, N. (2017). POT Python Optimal Transport library.

[20] Gairing, J., Högele, M., Kosenkova, T. and Kulik, A. (2015). Coupling distances between Lévy measures and applications to noise sensitivity of SDE. *Stochastics and Dynamics* **15** 1550009. MR3332269

[21] Gao, F. and van der Vaart, A. (2016). Posterior contraction rates for deconvolution of Dirichlet-Laplace mixtures. *Electron. J. Statist.* **10** 608–627. MR3471990

[22] Geller, M. and W. Ng, E. (1969). A table of integrals of exponential integral. *Journal of Research of the National Bureau of Standards, Section B: Mathematical Sciences* **73B**. MR0249669

[23] Ghosal, S. and van der Vaart, A. (2017). *Fundamentals of Nonparametric Bayesian Inference. Cambridge Series in Statistical and Probabilistic Mathematics* **44**. Cambridge University Press, Cambridge. MR3587782

[24] Gini, C. (1914). Di una misura delle relazioni tra le graduatorie di due caratteri. *Saggi monografici del Comune di Roma, Tip. Cecchini.*

[25] Hanson, T. E., Jara, A. and Zhao, L. (2012). A Bayesian semiparametric temporally-stratified proportional hazards Model with Spatial Frailties. *Bayesian Anal.* **7** 147–188. MR2896715

[26] Heinrich, P. and Kahn, J. (2018). Strong identifiability and optimal minimax rates for finite mixture estimation. *Ann. Statist.* **46** 2844–2870. MR3851757

[27] Hjort, N. L. (1990). Nonparametric Bayes estimators based on beta processes in models for life history data. *Ann. Statist.* **18** 1259–1294. MR1062708

[28] Hjort, N. L., Holmes, C., Müller, P. and Walker, S. G. (2010). *Bayesian Nonparametrics*. *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press. MR2722988

[29] Ishwaran, H. and James, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *J. Amer. Statist. Assoc.* **96** 161–173. MR1952729

[30] Ishwaran, H. and James, L. F. (2004). Computational methods for multiplicative intensity models using weighted gamma processes. *Journal of the American Statistical Association* **99** 175-190. MR2054297

[31] James, L. F. (2003). Bayesian calculus for gamma processes with applications to semiparametric intensity models. *Sankhyā: The Indian Journal of Statistics (2003-2007)* **65** 179–206. MR2016784

[32] James, L. F. (2005). Bayesian Poisson process partition calculus with an application to Bayesian Lévy moving averages. *Ann. Statist.* **33** 1771–1799. MR2166562

[33] Kingman, J. F. C. (1967). Completely random measures. *Pacific J. Math.* **21** 59–78. MR0210185

[34] Laud, P. W., Smith, A. F. M. and Damien, P. (1996). Monte Carlo methods for approximating a posterior hazard rate process. *Statistics and Computing* **6** 77–83.

[35] Lijoi, A. and Nipoti, B. (2014). A class of hazard rate mixtures for combining survival data from different experiments. *Journal of the American Statistical Association* **109** 802-814. MR3223751

[36] Lijoi, A. and Prünster, I. (2010). *Models beyond the Dirichlet process*. In *Bayesian Nonparametrics*. *Cambridge Series in Statistical and Probabilistic Mathematics* 80–136. Cambridge University Press. MR2730661

[37] Lo, A. and Weng, C.-S. (1989). On a class of Bayesian nonparametric estimates: II. Hazard rate estimates. *Annals of the Institute of Statistical Mathematics* **41** 227-245. MR1006487

[38] Mallows, C. L. (1972). A note on asymptotic joint normality. *Ann. Math. Statist.* **43** 508–515.

[39] Mariucci, E. and Reiss, M. (2018). Wasserstein and total variation distance between marginals of Lévy processes. *Electron. J. Statist.* **12** 2482–2514. MR3833470

[40] Mijoule, G., Peccati, G. and Swan, Y. (2016). On the rate of convergence in de Finetti's representation theorem. *ALEA Lat. Am. J. Probab. Math. Stat.* **13** 1165–1187. MR3582913

[41] Müller, P., Quintana, F. A., Jara, A. and Hanson, T. (2015). *Bayesian nonparametric data analysis*. *Springer Series in Statistics*. Springer, Cham. MR3309338

[42] Nguyen, X. (2013). Convergence of latent mixing measures in finite and

infinite mixture models. *Ann. Statist.* **41** 370–400. MR3059422

[43] Nipoti, B., Jara, A. and Guindani, M. (2018). A Bayesian semiparametric partially PH model for clustered time-to-event data. *Scandinavian Journal of Statistics* **45** 1016-1035. MR3884898

[44] Panaretos, V. M. and Zemel, Y. (2018). Statistical aspects of Wasserstein distances. *To appear in Annual Review of Statistics and Its Applications. arXiv:1806.05500*. MR3939527

[45] Pennell, M. L. and Dunson, D. B. (2006). Bayesian semiparametric dynamic frailty models for multiple event time Data. *Biometrics* **62** 1044-1052. MR2297675

[46] Rachev, S. (1985). The Monge-Kantorovich mass transference problem and its stochastic applications. *Theory of Probability & Its Applications* **29** 647-676.

[47] Rosiński, J. (2001). *Series representations of Lévy processes from the perspective of point processes* In *Lévy Processes: Theory and Applications* 401–415. Birkhäuser Boston, Boston, MA. MR1833707

[48] Sato, K. (1999). *Lévy Processes and Infinitely Divisible Distributions. Cambridge Studies in Advanced Mathematics*. Cambridge University Press. MR1739520

[49] Sriperumbudur, B. K., Fukumizu, K., Gretton, A., Schölkopf, B. and Lanckriet, G. R. G. (2012). On the empirical estimation of integral probability metrics. *Electron. J. Statist.* **6** 1550–1599. MR2988458

[50] Srivastava, S., Li, C. and Dunson, D. (2015). Scalable Bayes via barycenter in Wasserstein space. *Journal of Machine Learning Research* **19**. MR3862415

[51] Villani, C. (2008). *Optimal Transport: Old and New. Grundlehren der mathematischen Wissenschaften*. Springer Berlin Heidelberg. MR2459454

[52] Wolpert, R. L. and Ickstadt, K. (1998). Poisson/gamma random field models for spatial statistics. *Biometrika* **85** 251–267. MR1649114

[53] Zhou, H., Hanson, T., Jara, A. and Zhang, J. (2015). Modeling county level breast cancer survival data using a covariate-adjusted frailty proportional hazards model. *The Annals of Applied Statistics* **9** 43–68. MR3341107