



Department of Economics and Management

**DEM Working Paper Series**

**A class of hazard rate mixtures for combining  
survival data from different experiments**

Antonio Lijoi  
(Università di Pavia & Collegio Carlo Alberto)

Bernardo Nipoti  
(Università di Torino & Collegio Carlo Alberto)

**# 59 (11-13)**

Via San Felice, 5  
I-27100 Pavia  
<http://epmq.unipv.eu/site/home.html>

**November 2013**

# A class of hazard rate mixtures for combining survival data from different experiments

A. Lijoi

University of Pavia & Collegio Carlo Alberto, Italy

B. Nipoti

University of Torino & Collegio Carlo Alberto, Italy

*November 2013*

## **Abstract**

Mixture models for hazard rate functions are widely used tools for addressing the statistical analysis of survival data subject to a censoring mechanism. The present paper introduces a new class of vectors of random hazard rate functions that are expressed as kernel mixtures of dependent completely random measures. This leads to define dependent nonparametric prior processes that are suitably tailored to draw inferences in the presence of heterogeneous observations. Besides its flexibility, an important appealing feature of our proposal is analytical tractability: we are, indeed, able to determine some relevant distributional properties and a posterior characterization that is also the key for devising an efficient MCMC sampler. For illustrative purposes, we specialize our general results to a class of dependent extended gamma processes. We finally display a few numerical examples, including both simulated and real two-sample datasets: these allow us to identify the effect of a borrowing strength phenomenon and provide evidence of the effectiveness of the prior to deal with datasets for which the proportional hazards assumption does not hold true.

KEYWORDS: Bayesian nonparametrics; Completely random measures; Dependent processes; Extended gamma processes; Partial exchangeability.

# 1 Introduction

The analysis of survival data has been one of the first areas of applications of Bayesian nonparametric techniques. Seminal contributions as those, e.g., in Doksum (1974), Ferguson (1974), Susarla and Van Ryzin (1976), Ferguson and Phadia (1979) and Dykstra and Laud (1981) contain deep theoretical results that paved the way for later advances in the field. Furthermore, the rapid development of Markov Chain Monte Carlo techniques in the last two decades has made the use of nonparametric priors for survival analysis accessible to practitioners as well. A large portion of this literature in Bayesian nonparametrics has been developed under the assumption that survival data are exchangeable. Such an assumption does not hold true when some source of heterogeneity affects the data. A typical example corresponds to observations that originate from different related studies, such as in randomized clinical trials where subjects are enrolled at different study centers or undertake different treatments. Indeed, factors specific to each centre or to each treatment might exert significant influence. Therefore, while it is reasonable to assume that exchangeability holds true for subjects in the same group or sample (i.e. recruited at the same centre or undertaking the same treatment), exchangeability between samples is hard to motivate. A possible treatment of these data in a Bayesian nonparametric framework relies on the definition of a collection of dependent process priors indexed by a covariate identifying each sample or group. We shall pursue this goal and define a class of dependent hazard rate mixture models that are suitably tailored to statistical inference with possibly censored survival observations that originate from different studies or experiments. In this respect, the present contribution inserts itself in a large body of research that, inspired by pioneering contributions due to S.N. MacEachern (1999; 2000), has been recently devoted to the proposal of dependent random probability measures with potential applications to nonparametric regression, time series analysis, meta-analysis, spatial statistics and so on. Examples include dependent Dirichlet processes obtained by means of a stick-breaking procedure (see, e.g. De Iorio et al., 2004; Griffin and Steel, 2008; Rodríguez et al., 2008; Teh et al., 2006), Pólya trees (Wong and Ma, 2010; Trippa et al., 2011), neutral to the right priors (Epifani and Lijoi, 2010), more general tail-free processes (Jara and Hanson, 2011) and normalized random measures with independent increments (Lijoi et al., 2013; Griffin et al., 2013).

The key to our construction is a vector of dependent random measures  $\tilde{\boldsymbol{\mu}} = (\tilde{\mu}_1, \dots, \tilde{\mu}_\gamma)$  recently proposed and investigated in Lijoi et al. (2013) and whose components are *completely random*. This means that, for any  $\ell = 1, \dots, \gamma$ , when  $\tilde{\mu}_\ell$  is evaluated at pairwise disjoint sets it gives rise to

mutually independent non-negative random variables. A vector  $(\tilde{S}_1, \dots, \tilde{S}_\gamma)$  of survival functions, or of the corresponding dependent random probability measures  $(\tilde{P}_1, \dots, \tilde{P}_\gamma)$ , is thus defined by

$$\tilde{S}_\ell(t) := \tilde{P}_\ell((t, \infty)) = \exp\left(-\int_0^t \int_{\mathbb{Y}} k(s, y) \tilde{\mu}_\ell(dy) ds\right) \quad (1)$$

for any  $\ell = 1, \dots, \gamma$ , for any  $t > 0$  and for some kernel function  $k : \mathbb{R}^+ \times \mathbb{Y} \rightarrow \mathbb{R}^+$ . Equation (1) amounts to specifying a prior for the vector of hazard rate functions (RHRs)  $\tilde{\mathbf{h}} = (\tilde{h}_1, \dots, \tilde{h}_\gamma)$  as kernel mixtures of dependent CRMs, namely  $\tilde{h}_\ell(t) = \int_{\mathbb{Y}} k(t, y) \tilde{\mu}_\ell(dy)$ . Our approach leads to a much more flexible specification than models based, e.g., on the proportional hazards assumption which may fail in some situations of practical interest. It further yields smooth estimates of the marginal survival functions and incorporates a borrowing strength effect due to the presence, in the definition of each  $\tilde{\mu}_\ell$ , of a component which is shared among samples. On the other hand, the nice probabilistic structure featured by the vector of CRMs  $\tilde{\boldsymbol{\mu}}$  is a key property that allows us to prove important theoretical results on the distributional properties of  $\tilde{\mathbf{h}}$ , both *a priori* and *a posteriori*. For example, we obtain a simple expression for the correlation between  $\tilde{S}_i(t_1)$  and  $\tilde{S}_j(t_2)$ , for  $i \neq j$  and any  $t_1$  and  $t_2$ , and provide an explicit characterization of the posterior distribution of  $\tilde{\mathbf{h}}$ . As a by-product, the representation of the posterior distribution of  $\tilde{\mathbf{h}}$  that we achieve straightforwardly yields a Gibbs sampling algorithm that allows an approximate evaluations of Bayesian estimators for several quantities of interest.

As first displayed in Lijoi and Prünster (2010), CRMs constitute the basic building-blocks of a wide variety of popular priors for Bayesian nonparametric inference; these then enjoy some common structural properties which can be characterized in terms of the corresponding CRMs. It should be also recalled that proposals of dependent priors induced by CRMs in survival analysis can be found in James (2003), where vectors of dependent gamma CRMs are defined in terms of mixtures similar to those set forth in Antoniak (1974), and in Epifani and Lijoi (2010) who introduce a new class of dependent neutral-to-the-right processes by resorting to Lévy copulae.

The outline of the paper is as follows. In Section 2 we provide a quick overview of a standard prior specification for hazard rate functions with exchangeable censored data. In Section 3 we develop our proposal of vectors of dependent RHRs and investigate their dependence structure. In Section 4 we state a characterization of the posterior distribution of the vector of dependent RHRs we have introduced. This is the main result of the paper and leads us to establish an estimate of the survival functions corresponding to each component of the vector  $\tilde{\mathbf{h}} = (\tilde{h}_1, \dots, \tilde{h}_\gamma)$ , conditional on  $\gamma \geq 2$

samples of survival times and on suitable latent variables. In Section 5 we specialize our general construction to the case of dependent extended gamma processes, with  $\gamma = 2$ , and provide an expression of a Bayesian estimator of the survival function for each group of data that is examined. The posterior characterization of  $\tilde{\mathbf{h}}$  is also relevant for devising a Gibbs sampling algorithm that allows for an approximate evaluations of the estimates of both the hazard rates and the survival functions. Finally, in Section 6 we provide two illustrative examples with simulated and real data. In order to ease the exposition in the paper, the proofs of propositions and the derivation of the full conditional distributions of random variables entering the model specification are all postponed to the Appendix.

## 2 Priors for hazard rate functions

Let  $F$  be the cumulative distribution function (cdf) associated to a probability distribution on  $\mathbb{R}^+$ . The corresponding survival and cumulative hazard functions are denoted as  $S(t) = 1 - F(t)$  and  $H(t) = -\log(S(t))$ , for any  $t > 0$ . If  $F$  is absolutely continuous, the hazard rate function associated to  $F$  is, thus, defined as  $h(t) = F'(t)/[1 - F(t-)]$ , where  $F(t-) := \lim_{\varepsilon \downarrow 0} F(t - \varepsilon)$ . Nonparametric priors for hazard rates  $h$  have been originally proposed in Dykstra and Laud (1981) and, then, generalized in later work by Lo and Weng (1989) and James (2005). Their uses in practical applications for the analysis of censored survival data have been eased by the recent developments of powerful computational techniques that allow for an approximate evaluation of posterior inferences on quantities of statistical interest. The key probabilistic tool shared by these contributions is represented by *completely random measures* (CRMs). Recall that a CRM  $\mu$  on  $\mathbb{Y}$  is a random measure that, when evaluated at any collection of pairwise disjoint sets  $A_1, \dots, A_d$ , gives rise to mutually independent random variables  $\mu(A_1), \dots, \mu(A_d)$ , for any  $d \geq 1$ . A more detailed description of properties of CRMs is postponed to the Appendix.

Henceforth, we shall assume that  $\mu$  does not have jumps at fixed locations, which implies that there exists a measure  $\nu$  on  $\mathbb{R}^+ \times \mathbb{Y}$  such that  $\int_{\mathbb{R}^+ \times \mathbb{Y}} \min\{s, 1\} \nu(ds, dy) < \infty$  and

$$\mathbb{E} \left[ \exp \left( - \int_{\mathbb{Y}} f(y) \mu(dy) \right) \right] = \exp \left( - \int_{\mathbb{R}^+ \times \mathbb{Y}} [1 - \exp(-s f(y))] \nu(ds, dy) \right), \quad (2)$$

for any measurable function  $f : \mathbb{Y} \rightarrow \mathbb{R}$  such that  $\int_{\mathbb{Y}} |f| d\mu < \infty$ , with probability 1. We refer to the

measure  $\nu$  as the *Lévy intensity* of  $\mu$ . For our purposes, it will be useful to rewrite  $\nu$  as

$$\nu(ds, dy) = \rho_y(s) ds c P_0(dy), \quad (3)$$

where  $P_0$  is a probability measure on  $\mathbb{Y}$  and  $\rho_y(s)$  is some transition kernel on  $\mathbb{Y} \times \mathbb{R}^+$ . If  $\rho_y = \rho$ , for any  $y$  in  $\mathbb{Y}$ , the CRM  $\mu$  is said *homogeneous*. In the sequel we further suppose that  $P_0$  is non-atomic. A well-known example corresponds to  $\rho_y(s) = \rho(s) = e^{-s}/s$ , which identifies a so-called *gamma CRM*  $\mu$ . With such a choice of the Lévy intensity, from (A.1) it can be seen that for any  $A$  such that  $P_0(A) > 0$ , the random variable  $\mu(A)$  has a  $\text{Ga}(1, cP_0(A))$ , where by  $\text{Ga}(a, b)$  one denotes a probability distribution admitting density function  $g_{a,b}(x) = \mathbb{1}_{\mathbb{R}^+}(x) a^b x^{b-1} \exp(-ax)/\Gamma(b)$  and  $\mathbb{1}_C$  denotes the indicator function of set  $C$ .

If  $k(\cdot; \cdot)$  is a transition kernel on  $\mathbb{R}^+ \times \mathbb{Y}$ , a prior for  $h$  is the distribution of the random hazard rate (RHR)

$$\tilde{h}(t) = \int_{\mathbb{Y}} k(t; y) \mu(dy), \quad (4)$$

where  $\mu$  is a CRM on  $\mathbb{Y}$ . We observe that, if  $\lim_{t \rightarrow \infty} \int_0^t \tilde{h}(s) ds = \infty$  with probability 1, then  $\tilde{h}$  defines a random survival function  $\tilde{S}(t) = \exp(-\int_0^t \tilde{h}(s) ds)$ . Such a specification of  $\tilde{h}$  is used to define the prior for an exchangeable sequence of (possibly censored) survival data, namely  $\mathbf{X} = (X_1, \dots, X_n)$  is such that  $X_i | \tilde{P} \stackrel{\text{iid}}{\sim} \tilde{P}$  and  $\tilde{P}((\cdot, \infty)) = \exp(-\int_0^\cdot \tilde{h}(s) ds)$ . In this setting, Dykstra and Laud (1981) characterize the posterior distribution of the so-called *extended gamma process*: this is obtained when  $\mu$  is a gamma CRM and  $k(t; y) = \mathbb{1}_{(0,t]}(y) \beta(y)$  for some positive right-continuous function  $\beta : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ . The same kind of result is proved in Lo and Weng (1989) for the *weighted gamma processes* corresponding to RHRs obtained when  $\mu$  is still a gamma CRM and  $k(\cdot; \cdot)$  is an arbitrary kernel. Finally, a posterior characterization has been derived in James (2005) for any CRM  $\mu$  and kernel  $k(\cdot; \cdot)$ . A clever augmentation argument that introduces a vector of auxiliary variables  $\mathbf{Y} = (Y_1, \dots, Y_n)$ , combined with results concerning disintegrations of Poisson random measures, leads to a description of the posterior distribution of  $\tilde{h}$ , given  $\mathbf{X}$  and  $\mathbf{Y}$ , that is also very convenient for simulation purposes. See, e.g., Ishwaran and James (2004).

REMARK 1. It should be noted that a specification of the hazard rate similar to (4) is present also in the non-Bayesian literature. For example, in Gjessing et al. (2003) and in Aalen et al. (2008) one can find a detailed description on the use of Lévy processes to model time-dependent individual frailties. Moreover, there is an apparent analogy with the well known class of frailty models where, unlike in

(4), the random effect is considered fixed in time and affects the hazard function multiplicatively (see, e.g., Hougaard, 2000).  $\square$

As mentioned in the Introduction, exchangeability corresponds to assuming that the observations are analogous or homogeneous. In situations where there are  $\gamma \geq 2$  different experimental conditions under which data are generated, one cannot rely on such a dependence scheme. If  $\mathbf{X}_\ell = (X_{1,\ell}, \dots, X_{n_\ell,\ell})$  are, for  $\ell = 1, \dots, \gamma$ , the data originating from the  $\ell$ -th sample, a possible dependence structure that takes into account the nature of the data is

$$\begin{aligned} (X_{j_1,1}, \dots, X_{j_\gamma,\gamma}) | (\tilde{P}_1, \dots, \tilde{P}_\gamma) &\stackrel{\text{iid}}{\sim} \tilde{P}_1 \times \dots \times \tilde{P}_\gamma \\ (\tilde{P}_1, \dots, \tilde{P}_\gamma) &\sim Q \end{aligned} \tag{5}$$

for any  $j_\ell = 1, \dots, n_\ell$ . A collection of sequences  $\{(X_{n,\ell})_{n \geq 1} : \ell = 1, \dots, \gamma\}$  for which (5) holds true is said *partially exchangeable*. Here  $Q$  plays the role of a prior distribution on the space of all vectors  $(P_1, \dots, P_\gamma)$  whose components are probability measures on  $\mathbb{R}^+$ . When  $Q$  is degenerate on a single coordinate, i.e.  $Q(\{\tilde{P}_1 = \dots = \tilde{P}_\gamma\}) = 1$ , condition (5) reduces to the more familiar conditional independence and identity in distribution that characterizes exchangeability. In the next section a new class of priors  $Q$  in (5) will be specified in terms of the hazard rate functions  $\tilde{h}_\ell$  associated to the coordinates  $\tilde{P}_\ell$  as highlighted in (1).

### 3 A model for partially exchangeable survival data

We aim at defining a flexible and tractable prior  $Q$  for a vector of dependent random probability measures  $(\tilde{P}_1, \dots, \tilde{P}_\gamma)$ . To achieve this goal, we express each  $\tilde{P}_\ell$  as a suitable transformation of a CRM  $\tilde{\mu}_\ell$  so that dependence among  $\tilde{P}_1, \dots, \tilde{P}_\gamma$  is induced by dependence at the level of the CRMs  $\tilde{\boldsymbol{\mu}} = (\tilde{\mu}_1, \dots, \tilde{\mu}_\gamma)$ . More precisely, we shall resort to the specification (1). For example, if  $\gamma = 2$  one can define each hazard function  $\tilde{h}_1$  and  $\tilde{h}_2$  as a mixture of a gamma CRM  $\tilde{\mu}_1$  and  $\tilde{\mu}_2$ , i.e.  $\tilde{h}_\ell(t) = \int_{(0,t]} \beta(y) \tilde{\mu}_\ell(dy)$ , for some positive and right-continuous function  $\beta$ : this gives rise to a vector of extended gamma processes. Dependence between  $\tilde{\mu}_1$  and  $\tilde{\mu}_2$  is induced via the construction suggested in Lijoi et al. (2013), which makes use of a specific vector of dependent Poisson random measures (PRMs) discussed by Griffiths and Milne (1978). For easing the exposition, we shall henceforth confine to the case where  $\gamma = 2$ , even though the analysis carries over to any dimension  $\gamma$ , as apparent

from Proposition 3 below.

### 3.1 GM–dependent completely random measures

Let  $N$  be a Poisson random measure (PRM) on  $\mathbb{R}^+ \times \mathbb{Y}$  with intensity measure  $\nu$ , i.e. there exists a sequence  $(J_i, Y_i)_{i \geq 1}$  of independent random elements with  $J_i > 0$  and  $Y_i \in \mathbb{Y}$  such that if  $A$  is a measurable subset of  $\mathbb{R}^+ \times \mathbb{Y}$  such that  $\nu(A) < \infty$ , the random variable  $N(A) = \text{card}(A \cap \{(J_i, Y_i) : i \geq 1\})$  has Poisson distribution with parameter  $\nu(A)$ . It is, then, well-known that

$$\mu(B) = \int_{\mathbb{R}^+ \times B} s N(ds, dy) = \sum_{i: Y_i \in B} J_i \quad (6)$$

is a CRM on  $\mathbb{Y}$  and the Laplace functional transform of  $\mu$  equals the right-hand-side of (A.1). It is apparent from (6) that a pair of dependent CRMs  $(\tilde{\mu}_1, \tilde{\mu}_2)$  can be defined as linear functionals of dependent PRMs  $(\tilde{N}_1, \tilde{N}_2)$ . This is the starting point of Lijoi et al. (2013): for their construction they rely on a class of dependent PRMs investigated by Griffiths and Milne (1978) that feature a nice probabilistic structure. Indeed, Griffiths and Milne (1978) provide an interesting and useful characterization of a vector of (possibly dependent) PRMs  $(\tilde{N}_1, \tilde{N}_2)$  on  $\mathbb{R}^+ \times \mathbb{Y}$  with the same marginal intensity measure  $\bar{\nu}$  that admit an additive representation

$$\tilde{N}_\ell = N_\ell + N_0 \quad \ell = 1, 2, \quad (7)$$

where  $N_1$ ,  $N_2$  and  $N_0$  are independent Cox processes with respective random intensities  $\tilde{\nu}$ ,  $\tilde{\nu}$  and  $\tilde{\nu}_0$  such that  $\tilde{\nu}_0 \leq \nu$  (a.s.) and  $\tilde{\nu} = \nu - \tilde{\nu}_0$ . In particular, the appeal of their result can be traced back to a nice representation of the joint Laplace functional of  $(\tilde{N}_1, \tilde{N}_2)$  which is an important operational tool for deriving analytical results useful for Bayesian inference. Moreover, (7) allows us to gain some insight on the dependence between  $\tilde{N}_1$  and  $\tilde{N}_2$  by stating that each may be decomposed in a common ( $N_0$ ) and an idiosyncratic ( $N_\ell$ ) component, for  $\ell = 1, 2$ . We can now provide the following

**DEFINITION 1.** Let  $\tilde{\mu}_\ell(dy) = \int_{\mathbb{R}^+} s \tilde{N}_\ell(ds, dy)$ , for  $\ell = 1, 2$ , where  $\tilde{N}_1$  and  $\tilde{N}_2$  are defined as in (7) with  $\nu(ds, dy) = cP_0(dy)\rho(s)ds$ . Then  $(\tilde{\mu}_1, \tilde{\mu}_2)$  is said to be vector of *GM–dependent* CRMs on  $\mathbb{Y}$  and the marginal intensity of  $\tilde{\mu}_\ell$  is equal to  $\nu$ .

Henceforth we shall specifically deal with the case where  $N_1$ ,  $N_2$  and  $N_0$  are three independent Cox



processes with respective intensities

$$\tilde{\nu}(ds, dy) = cZP_0(dy) \rho(s) ds, \quad \tilde{\nu}_0(ds, dx) = c(1 - Z)P_0(dy) \rho(s) ds \quad (8)$$

for some  $[0, 1]$ -valued random variable  $Z$  independent of  $N_i$ , for  $i = 0, 1, 2$ . Moreover, in order to ease the exposition we work conditioning on a fixed value  $Z = z$ . According to the definition above, the marginals of a vector of GM-dependent CRMs are equally distributed and  $\tilde{\mu}_\ell(dx) = \mu_\ell(dx) + \mu_0(dx)$ . Moreover  $\mu_\ell$ , with  $\ell = 1, 2$ , and  $\mu_0$  are independent CRMs with  $\mathbb{E}[\exp(-\mu_\ell(f))] = \exp(-cz\psi(f))$  and  $\mathbb{E}[\exp(-\mu_0(f))] = \exp(-c(1 - z)\psi(f))$ , where  $\psi(f) := \int_{\mathbb{R}^+ \times \mathbb{Y}} [1 - \exp(-sf(y))] \nu(ds, dy)$ . One can then easily recover the joint Laplace transform of  $(\tilde{\mu}_1, \tilde{\mu}_2)$ . Indeed, for any pair of measurable functions  $f_\ell : \mathbb{Y} \rightarrow \mathbb{R}$ , for  $\ell = 1, 2$ , such that  $\mathbb{P}[\tilde{\mu}_\ell(|f_\ell|) < \infty] = 1$ , one has

$$\mathbb{E}[\exp(-\tilde{\mu}_1(f_1) - \tilde{\mu}_2(f_2))] = \exp(-c\psi_z(f_1, f_2)), \quad (9)$$

where  $\psi_z(f_1, f_2) = z[\psi(f_1) + \psi(f_2)] + (1 - z)\psi(f_1 + f_2)$ . As a simple example, consider a vector of GM-dependent gamma CRMs, namely  $\rho(s) = e^{-s} s^{-1}$ . In this case  $\psi(f_\ell) = \int \log(1 + f_\ell) dP_0$  for any measurable function  $f_\ell$  such that  $\int \log(1 + |f_\ell|) dP_0 < \infty$ , with  $\ell = 1, 2$ . The bivariate Laplace functional transform  $\exp(-c\psi_z(f_1, f_2))$  of the vector of dependent gamma CRMs, then, is

$$\exp\left(-cz \int \log((1 + f_1)(1 + f_2)) dP_0 - c(1 - z) \int \log(1 + f_1 + f_2) dP_0\right). \quad (10)$$

### 3.2 GM-dependent random hazard rates

Let  $(\tilde{\mu}_1, \tilde{\mu}_2)$  be a vector of GM-dependent CRMs and  $(\tilde{h}_1, \tilde{h}_2)$  be such that  $\tilde{h}_\ell(t) = \int_{\mathbb{Y}} k(t; y) \tilde{\mu}_\ell(dy)$ , for  $\ell = 1, 2$ . If  $\lim_{t \rightarrow \infty} \int_0^t \tilde{h}_\ell(s) ds = \infty$ , almost surely, then we say that  $(\tilde{h}_1, \tilde{h}_2)$  is a vector of *GM-dependent RHRs* and define survival functions

$$\tilde{S}_1(t) = \exp\left(-\int_0^t \tilde{h}_1(s) ds\right), \quad \tilde{S}_2(t) = \exp\left(-\int_0^t \tilde{h}_2(s) ds\right). \quad (11)$$

If  $y \mapsto \bar{K}_t(y) := \int_0^t k(s; y) ds$ , it can be easily seen that

$$\mathbb{E}\left[\tilde{S}_1(t_1) \tilde{S}_2(t_2)\right] = \exp(-c\psi_z(\bar{K}_{t_1}, \bar{K}_{t_2})), \quad (12)$$

where  $\psi_z$  is the joint Laplace exponent defined in (9). Given the simple structure of  $\tilde{h}_\ell$ , for  $\ell = 1, 2$ , it is possible to analyze the dependence between the survival functions  $\tilde{S}_1(t_1)$  and  $\tilde{S}_2(t_2)$ , for some fixed  $t_1, t_2 > 0$ . This will be expressed through the correlation function as described in the following.

PROPOSITION 1. *Let  $(\tilde{S}_1, \tilde{S}_2)$  be as in (11). Then the correlation between  $\tilde{S}_1(t_1)$  and  $\tilde{S}_2(t_2)$  is given by*

$$\text{Corr}(\tilde{S}_1(t_1), \tilde{S}_2(t_2)) = \frac{\exp(c(1-z)[\psi(\bar{K}_{t_1}) + \psi(\bar{K}_{t_2}) - \psi(\bar{K}_{t_1} + \bar{K}_{t_2})]) - 1}{\sqrt{\prod_{i=1}^2 [\exp(c(2\psi(\bar{K}_{t_i}) - \psi(2\bar{K}_{t_i}))) - 1]}}.$$

The proof is omitted since the result easily follows from (9). As expected, the correlation vanishes when  $z = 1$ . A simplification occurs if  $t_1 = t_2 = t$ , when

$$\text{Corr}(\tilde{S}_1(t), \tilde{S}_2(t)) = \frac{\exp(c(1-z)[2\psi(\bar{K}_t) - \psi(2\bar{K}_t)]) - 1}{\exp(c[2\psi(\bar{K}_t) - \psi(2\bar{K}_t)]) - 1}. \quad (13)$$

Under the additional conditions

- (i)  $\int_{\mathbb{Y}} k(t_1; y) k(t_2; y) P_0(dy) < \infty$  for any positive  $t_1$  and  $t_2$ ,
- (ii)  $\int_0^\infty s^2 \rho(s) ds < \infty$ ,

it is possible to find an explicit expression for the correlation of  $\tilde{h}_1(t_1)$  and  $\tilde{h}_2(t_2)$ . Interestingly, when  $t_1 = t_2 = t$ , we obtain  $\text{Corr}(\tilde{h}_1(t), \tilde{h}_2(t)) = (1 - z)$ , that does not depend on  $t$ . Details are provided in Proposition B.1 in Appendix B.

The probability distribution of the vector of survival functions in (11) defines a nonparametric prior for Bayesian analysis of partially exchangeable survival data. More precisely, if  $\mathbf{X}_1 = (X_{1,1}, \dots, X_{n_1,1})$ , and  $\mathbf{X}_2 = (X_{1,2}, \dots, X_{n_2,2})$  are a two-sample survival dataset, a model we shall use is

$$\begin{aligned} X_{i,1}, X_{j,2} | \tilde{h}_1, \tilde{h}_2 &\sim \tilde{S}_1 \times \tilde{S}_2 & i, j = 1, 2, \dots \\ \tilde{h}_\ell(t) &\stackrel{d}{=} \int_{\mathbb{Y}} k(t; y) \tilde{\mu}_\ell(dy) & \ell = 1, 2. \end{aligned} \quad (14)$$

Given this specification, one can address the issue of determining the posterior distribution of the pair of hazard rate functions  $(\tilde{h}_1, \tilde{h}_2)$  or, equivalently, of  $(\tilde{S}_1, \tilde{S}_2)$ , with the aim of evaluating Bayesian estimates of the marginal hazard rates or of the marginal survival functions. This problem is tackled in the next section.

Even if not directly discussed here, note that one can also consider distinct kernels  $k_1(\cdot; \cdot)$  and  $k_2(\cdot; \cdot)$  for defining RHRs: the results that will be presented in the next section still hold true with

some minor suitable adaptation.

REMARK 2. The definition of dependent nonparametric priors for hazard rate functions makes our proposal suited for applications even beyond the setting described through (14). Indeed, a natural problem where a prior for  $\tilde{\mathbf{h}} = (\tilde{h}_1, \dots, \tilde{h}_\gamma)$  is of interest is related to the analysis of competing risks data with  $\gamma$  distinct sources of failure. In this case, each  $\tilde{h}_i$  would be the  $i$ -th cause specific random hazard rate function. Techniques similar to those that are used to deduce a posterior characterization of  $\tilde{\mathbf{h}}$  here can be used to develop a full Bayesian analysis also in a competing risks framework. Another example of possible applications of the proposal concerns inverse problems when observations are available on the hazard rate  $h$ , or on functionals of  $h$ , and one wishes to estimate the mixing random measure  $\mu$  in (4). See, e.g., Wolpert et al. (2003).  $\square$

## 4 Characterization of the posterior distribution

For a full Bayesian analysis of survival data modeled as in (14), one needs to determine a posterior characterization of the vector  $(\tilde{h}_1, \tilde{h}_2)$ . For the sake of simplicity we suppose that the conditioning  $n_1 + n_2$  data  $\mathbf{X}_1$  and  $\mathbf{X}_2$  are all exact: the extension to the case where censored data are also present is straightforward and will be considered in the examples of Section 6. As in the univariate case, we find a characterization of the posterior distribution of  $(\tilde{h}_1, \tilde{h}_2)$  by conditioning on some latent (non observable) random variables. Even though one can formally establish a characterization of the posterior distribution of  $(\tilde{h}_1, \tilde{h}_2)$  without resorting to the introduction of auxiliary random variables, such an augmentation is very convenient since it avoids dealing with unfeasible computations that involve sums over spaces of partitions.

Throughout we work with the multiplicative intensity likelihood

$$\mathcal{L}(\mu_1, \mu_2; \mathbf{X}_1, \mathbf{X}_2) = \prod_{\ell=1}^2 \exp\left(-\int_{\mathbb{Y}} K_\ell(y) \mu_\ell(dy)\right) \prod_{i=1}^{n_\ell} \int_{\mathbb{Y}} k(X_{i,\ell}; y) \mu_\ell(dy), \quad (15)$$

where  $K_\ell(y) = \sum_{i=1}^{n_\ell} \int_0^{X_{i,\ell}} k(s; y) ds$  for  $\ell = 1, 2$  and we have used the fact that all data are exact. We introduce here some latent random variables  $\mathbf{Y}_\ell = (Y_{1,\ell}, \dots, Y_{n_\ell,\ell})$ , for  $\ell = 1, 2$ , such that the joint

law of  $(\tilde{\mu}_1, \tilde{\mu}_2, \mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}_1, \mathbf{Y}_2)$  is given by

$$\prod_{\ell=1}^2 \exp\left(-\int_{\mathbb{Y}} K_{\ell}(y) \mu_{\ell}(dy)\right) \prod_{i=1}^{n_{\ell}} k(X_{i,\ell}; Y_{i,\ell}) \mu_{\ell}(dY_{i,\ell}) Q^*(d\mu_1, d\mu_2), \quad (16)$$

where  $Q^*$  is the joint distribution of the GM-dependent CRMs  $(\tilde{\mu}_1, \tilde{\mu}_2)$ . We observe that, with positive probability, the latent variables  $\mathbf{Y}_1$  and  $\mathbf{Y}_2$  can display ties. Indeed, for  $\ell = 1, 2$  and for every  $j = 1, \dots, n_{\ell}$ , we have that

$$\mathbb{P}[Y_{j,\ell} \in dy, X_{j,\ell} \in dx \mid \tilde{\mu}_{\ell}] \propto \exp\left(-\int_{\mathbb{Y}} \int_0^x k(s; y') ds \tilde{\mu}_{\ell}(dy')\right) k(x; y) dx \tilde{\mu}_{\ell}(dy)$$

and  $\mathbb{P}[Y_{j,\ell} \in dy \mid \tilde{\mu}_{\ell}] \propto \tilde{\mu}_{\ell}(dy) \int_0^{\infty} \exp(-\int_{\mathbb{Y}} \int_0^x k(s; y') ds \tilde{\mu}_{\ell}(dy')) k(x; y) dx$ . Thus, the almost sure discreteness of  $\tilde{\mu}_{\ell}$  implies that  $\mathbb{P}[Y_{j,\ell} = Y_{i,\ell}] > 0$  for  $j \neq i$ . Moreover, by definition of GM-dependence,  $\tilde{\mu}_1$  and  $\tilde{\mu}_2$  share a common component and, therefore, we have that  $\mathbb{P}[Y_{i,1} = Y_{j,2}] > 0$  for any  $i$  and  $j$ . On the basis of these considerations, the partition structure within  $\mathbf{Y}_1$  and  $\mathbf{Y}_2$  can be described as follows

$$\begin{aligned} \{Y_{1,1}, \dots, Y_{n_{1,1}}\} &= \{Y_{1,1}^*, \dots, Y_{k_{1,1}}^*\} \cup \{Y_1^*, \dots, Y_{k_0}^*\}, \\ \{Y_{1,2}, \dots, Y_{n_{2,2}}\} &= \{Y_{1,2}^*, \dots, Y_{k_{2,2}}^*\} \cup \{Y_1^*, \dots, Y_{k_0}^*\}, \end{aligned}$$

where  $\{Y_{1,\ell}^*, \dots, Y_{k_{\ell,\ell}}^*, Y_1^*, \dots, Y_{k_0}^*\}$  are the distinct values of  $\mathbf{Y}_{\ell}$ , for  $\ell = 1, 2$ , and  $\{Y_{1,1}^*, \dots, Y_{k_{1,1}}^*\} \cap \{Y_{1,2}^*, \dots, Y_{k_{2,2}}^*\} = \emptyset$ . Moreover,  $1 \leq k_{\ell} + k_0 \leq n_{\ell}$  for  $\ell = 1, 2$ . We introduce also the corresponding frequencies  $n_{j,\ell} = \sum_{i=1}^{n_{\ell}} \mathbb{1}_{\{Y_{i,\ell} = Y_{j,\ell}^*\}}$ ,  $q_{m,\ell} = \sum_{i=1}^{n_{\ell}} \mathbb{1}_{\{Y_{i,\ell} = Y_m^*\}}$ , for any  $\ell = 1, 2$ ,  $j = 1, \dots, k_{\ell}$  and  $m = 1, \dots, k_0$ . One accordingly has the obvious constraint  $n_{\ell} = \sum_{j=1}^{k_{\ell}} n_{j,\ell} + \sum_{m=1}^{k_0} q_{m,\ell}$ , for  $\ell = 1, 2$ , if we agree that  $\sum_{i=1}^0 \equiv 0$ . Taking into account this specific partition of the latent variables, (16) can be rewritten as follows

$$\begin{aligned} \prod_{\ell=1}^2 \exp\left(-\int_{\mathbb{Y}} K_{\ell}(y) \mu_{\ell}(dy)\right) &\prod_{i=1}^{k_{\ell}} (\mu_{\ell}(dY_{i,\ell}^*))^{n_{i,\ell}} \prod_{j \in C_{i,\ell}} k(X_{j,\ell}; Y_{i,\ell}^*) \\ &\times \prod_{m=1}^{k_0} (\mu_{\ell}(dY_m^*))^{q_{m,\ell}} \prod_{j \in C'_{m,\ell}} k(X_{j,\ell}; Y_m^*) Q^*(d\mu_1, d\mu_2), \quad (17) \end{aligned}$$

where we have set  $C_{i,\ell} := \{j : Y_{j,\ell} = Y_{i,\ell}^*\}$  and  $C'_{m,\ell} = \{j : Y_{j,\ell} = Y_m^*\}$  for  $\ell = 1, 2$ . The representation in (17) is the starting point for establishing the main result. Before stating it, we set some notation.

First, let  $\mathbf{V}_1 = (V_{i,1})_{i \geq 1}$  and  $\mathbf{V}_2 = (V_{j,2})_{j \geq 1}$  be independent sequences of i.i.d. random variables with  $\mathbb{P}[V_{i,\ell} = 0] = 1 - \mathbb{P}[V_{i,\ell} = 1] = z$  for  $i = 1, 2, \dots$  and  $\ell = 1, 2$ . Moreover, introduce exponentially tilted Lévy intensities

$$\begin{aligned} \nu_0^*(ds, dy) &= c(1-z) \exp(-(K_1(y) + K_2(y))t) \rho(s) ds P_0(dy), \\ \nu_\ell^*(ds, dx) &= cz \exp(-K_\ell(y)s) \rho(s) ds P_0(dy) \end{aligned} \quad (18)$$

for  $\ell = 1, 2$ , and, for any  $\kappa \in \mathbb{N}$  and  $\xi > 0$ , the density function

$$f(t | \kappa, \xi) \propto t^\kappa \exp(-\xi t) \rho(t) \mathbb{1}_{\mathbb{R}^+}(t). \quad (19)$$

PROPOSITION 2. *Let  $(\tilde{h}_1, \tilde{h}_2)$  be a vector of GM-dependent RHRs. Then, given  $\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}_1, \mathbf{Y}_2, \mathbf{V}_1$  and  $\mathbf{V}_2$ , the posterior distribution of  $(\tilde{h}_1, \tilde{h}_2)$  equals the distribution of*

$$(\tilde{h}_1^*, \tilde{h}_2^*) + \sum_{i=1}^{k_1} (1, V_{i,1}) k(\cdot, Y_{i,1}^*) J_{i,1} + \sum_{j=1}^{k_2} (V_{j,2}, 1) k(\cdot, Y_{j,2}^*) J_{j,2} + \sum_{m=1}^{k_0} (1, 1) k(\cdot, Y_m^*) J_m, \quad (20)$$

where  $\tilde{h}_\ell^*(t) = \int_{\mathbb{Y}} k(t; y) \tilde{\mu}_\ell^*(dy)$ , for each  $\ell = 1, 2$ , and

- (i)  $\tilde{\mu}_\ell^*$  is CRM such that  $\tilde{\mu}_\ell^* \stackrel{d}{=} \mu_\ell^* + \mu_0^*$ , for  $\ell = 1, 2$  and  $\mu_0^*, \mu_1^*$  and  $\mu_2^*$  are independent CRMs with Lévy intensities are equal to  $\nu_0^*, \nu_1^*$  and  $\nu_2^*$  in (18), respectively;
- (ii) the jumps  $J_{1,1}, \dots, J_{k_1,1}, J_{1,2}, \dots, J_{k_2,2}$  and  $J_1, \dots, J_{k_0}$  are mutually independent and independent from  $(\tilde{h}_1, \tilde{h}_2)$ . Moreover,  $J_{i,1}$  has density  $f(\cdot | n_{i,1}, K_1(Y_{i,1}^*) + K_2(Y_{i,1}^*)V_{i,1})$ ,  $J_{j,2}$  has density  $f(\cdot | n_{j,2}, K_1(Y_{j,2}^*)V_{j,2} + K_2(Y_{j,2}^*))$  and  $J_m$  has density  $f(\cdot | q_{m,1} + q_{m,2}, K_1(Y_m^*) + K_2(Y_m^*))$ , where  $f(\cdot | \cdot, \cdot)$  is as in (19).

It is worth noting that item (i) in Proposition 2 implies that the dependence in  $(\tilde{h}_1^*, \tilde{h}_2^*)$  is not of GM-type since  $\tilde{\mu}_1^*$  and  $\tilde{\mu}_2^*$  are not generally equal in distribution. Furthermore, the structure of the posterior distribution of  $(\tilde{h}_1, \tilde{h}_2)$  is somehow reminiscent of the one in James (2005). However, in contrast to the univariate case, one now has three groups of fixed points of discontinuity and this feature reflects the specific construction of  $(\tilde{\mu}_1, \tilde{\mu}_2)$ . Indeed, in the first two groups the locations of the jumps coincide, respectively, with the distinct values of the latent variables  $\mathbf{Y}_1$  and  $\mathbf{Y}_2$  that appear only in the first and in the second sample. The value of the auxiliary random variables  $V_{i,\ell}$  determines whether the jumps of these two groups affect the posterior distribution of both  $\tilde{h}_1$  and  $\tilde{h}_2$  or of only

one of them. For instance, if  $V_{j,2} = 0$  then the  $j$ -th distinct value  $Y_{j,2}^*$  specific to the second group does not affect posterior inference on  $\tilde{h}_1$  or on functionals of  $\tilde{h}_1$ . On the other hand, all the distinct values  $Y_{i,1}^*$  specific to the first group impact the posterior distribution of  $\tilde{h}_1$  through the jump components. The third group includes jumps that occur at the distinct values of  $\mathbf{Y}_1$  and  $\mathbf{Y}_2$  shared by both samples: such jumps affect the posterior distribution of both components in  $(\tilde{h}_1, \tilde{h}_2)$ . Similar remarks apply also to the distributions of the random jumps: as a matter of fact only the density of  $J_m$  depends on both samples, whereas the density of  $J_{i,\ell}$  may or may not depend on the data in  $\mathbf{X}_\kappa$ , with  $\ell \neq \kappa$ , according to the value of the latent variable  $V_{i,\ell}$ . In summary, each  $V_{i,\ell}$  identifies which component of  $\tilde{\mu}_\ell$ , i.e. either  $\mu_\ell$  or  $\mu_0$ , the corresponding distinct value  $Y_{i,\ell}^*$  is associated to: if  $Y_{i,\ell}^*$  is generated by the idiosyncratic component  $\mu_\ell$ , then it is not shared by the other sample, whereas, if it is generated by the common component  $\mu_0$ , it affects the other sample as well. There is no need to specify the  $V$ 's for the distinct  $Y_m^*$  since these can only be generated by  $\mu_0$ .

REMARK 3. The posterior representation in (20) can be extended to the case where  $(\tilde{\mu}_1, \tilde{\mu}_2)$  are non-homogeneous GM-dependent CRMs, namely the marginal Lévy intensity of  $\tilde{\mu}_\ell$ , for  $\ell = 1, 2$ , is equal to  $\nu(ds, dy) = \rho_y(s) ds cP_0(dy)$  and the dependence of  $\rho_y$  on  $y$  cannot be dropped. One can proceed along the lines of the proof of Proposition 2 and the main decomposition displayed in the main result (20) is unchanged. The only differences concern the tilted Lévy intensities in (18), where  $\rho$  is replaced by  $\rho_x$ , and the densities of the jumps  $J_{i,\ell}$ , with  $i = 1, \dots, k_\ell$  and  $\ell = 1, 2$ , and  $J_r$ , for  $r = 1, \dots, k_0$ . As for the latter, the density function of  $J_{i,1}$  becomes

$$f_{i,1}(t | n_{i,1}, K_1(Y_{i,1}^*) + K_2(Y_{i,1}^*)V_{i,1}) \propto t^{n_{i,1}} \exp(-K_1(Y_{i,1}^*) + K_2(Y_{i,1}^*)V_{i,1}) \rho_{Y_{i,1}^*}(t)$$

for any  $t > 0$ . Similarly, one deduces the densities of  $J_{i,2}$  and  $J_r$  and the required modifications of the MCMC sampler described in Appendix F.  $\square$

If interest lies in estimating the survival functions of both populations  $\tilde{S}_1$  and  $\tilde{S}_2$ , Proposition 2 provides all the necessary ingredients for evaluating an estimator of  $\tilde{S}_\ell(t)$ , for any  $t > 0$ , with a squared loss function. The estimator provided in the next proposition depends on observations  $\mathbf{X}_1$  and  $\mathbf{X}_2$  and latent variables  $\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{V}_1$  and  $\mathbf{V}_2$ .

PROPOSITION 3. *Conditionally on latent variables  $\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{V}_1, \mathbf{V}_2$  and exact observations  $\mathbf{X}_1, \mathbf{X}_2$ , the*

random survival function  $\tilde{S}_1(t)$  has expected value equal to

$$\begin{aligned} & \exp \left( -c z \int_{\mathbb{R}^+ \times \mathbb{Y}} [1 - \exp(-s \bar{K}_t(y))] \exp(-s K_1(y)) \rho(s) ds P_0(dy) \right) \\ & \times \exp \left( -c(1-z) \int_{\mathbb{R}^+ \times \mathbb{Y}} [1 - \exp(-s \bar{K}_t(y))] \exp(-s (K_1(y) + K_2(y))) \rho(s) ds P_0(dy) \right) \\ & \times \prod_{i=1}^{k_1} \prod_{j=1}^{k_2} \prod_{m=1}^{k_0} C_{i,1}(t) C_{j,2}(t) C_m(t), \quad (21) \end{aligned}$$

where we recall that  $\bar{K}_t(y) = \int_0^t k(s, y) ds$ . Moreover,

$$\begin{aligned} C_{i,1}(t) &= \frac{\int_0^\infty s^{n_{i,1}} \exp(- (K_1(Y_{i,1}^*) + K_2(Y_{i,1}^*) V_{i,1} + \bar{K}_t(Y_{i,1}^*)) s) \rho(s) ds}{\int_0^\infty s^{n_{i,1}} \exp(- (K_1(Y_{i,1}^*) + K_2(Y_{i,1}^*) V_{i,1}) s) \rho(s) ds}, \\ C_{j,2}(t) &= \frac{\int_0^\infty s^{n_{j,2}} \exp(- (K_1(Y_{j,2}^*) V_{j,2} + K_2(Y_{j,2}^*) + \bar{K}_t(Y_{j,2}^*) V_{j,2}) s) \rho(s) ds}{\int_0^\infty s^{n_{j,2}} \exp(- (K_1(Y_{j,2}^*) V_{j,2} + K_2(Y_{j,2}^*)) s) \rho(s) ds}, \quad (22) \\ C_m(t) &= \frac{\int_0^\infty s^{q_{m,1} + q_{m,2}} \exp(- (K_1(Y_m^*) + K_2(Y_m^*) + \bar{K}_t(Y_m^*)) s) \rho(s) ds}{\int_0^\infty s^{q_{m,1} + q_{m,2}} \exp(- (K_1(Y_m^*) + K_2(Y_m^*)) s) \rho(s) ds}. \end{aligned}$$

Obvious changes lead to a closed form expression for an estimator of  $\tilde{S}_2$  as well. Note that once a Lévy intensity  $\nu$  and a kernel  $k(\cdot; \cdot)$  are specified, the expression in (21) can be evaluated. At this point, in order to obtain an estimator that does not depend on non-observable variables, one needs to marginalize with respect to the conditional distribution of  $(\mathbf{V}_1, \mathbf{V}_2, \mathbf{Y}_1, \mathbf{Y}_2)$ , given the data  $\mathbf{X}_1$  and  $\mathbf{X}_2$ . This task that cannot be completed analytically and we, then, resort to a Gibbs sampler that is described in Appendix F. Before moving on to the specific case of the dependent extended gamma process, it is worth noting that this construction can be generalized to include  $\gamma > 2$  GM-dependent CRMs. Suppose  $(\tilde{\mu}_1, \dots, \tilde{\mu}_\gamma)$  is such that  $\tilde{\mu}_\ell = \mu_\ell + \mu_0$ , for  $\ell = 1, \dots, \gamma$ , and  $\mu_0, \mu_1, \dots, \mu_\gamma$  are independent homogeneous CRMs. Accordingly, one can define dependent random hazard rates  $\tilde{h}_\ell(t) = \int_{\mathbb{Y}} k(t, y) \tilde{\mu}_\ell(dy)$  and a corresponding vector of  $\gamma$  dependent survival functions related to  $\gamma$  partially exchangeable populations. If  $\mathbf{X}_\ell = (X_{1,\ell}, \dots, X_{n_\ell,\ell})$ , the model can be described as

$$\begin{aligned} X_{j_1,1}, \dots, X_{j_\gamma,\gamma} \mid (\tilde{h}_1, \dots, \tilde{h}_\gamma) &\stackrel{\text{iid}}{\sim} \tilde{S}_1 \times \dots \times \tilde{S}_\gamma & j_1, \dots, j_\gamma \geq 1 \\ \tilde{h}_\ell &\stackrel{d}{=} \int_{\mathbb{Y}} k(\cdot; y) \tilde{\mu}_\ell(dy) & \ell = 1, \dots, \gamma, \end{aligned} \quad (23)$$

where  $-\log(\tilde{S}_\ell(\cdot)) \stackrel{d}{=} \int_0^\cdot \tilde{h}_\ell(s) ds$ . Analogously to the bivariate case, the latent variables  $\mathbf{Y}_\ell =$

$(Y_{1,\ell}, \dots, Y_{n_\ell, \ell})$ , for  $\ell = 1, \dots, \gamma$ , may exhibit ties both within and between samples. To describe the resulting partition, we let  $\mathcal{C}_i$  be the set of all  $i$ -tuples of distinct indices  $(j_1, \dots, j_i)$  identifying the samples that share the same latent value  $Y$ , for  $i = 2, \dots, \gamma$ . Moreover  $\mathcal{C}_i^{(\ell)} := \{(j_1, \dots, j_i) \in \mathcal{C}_i : \ell \in \{j_1, \dots, j_i\}\}$  is, for any  $\ell = 1, \dots, \gamma$ , the subset of  $\mathcal{C}_i$  which yields all possible  $i$ -tuples that involve also the  $\ell$ -th group. According to this, the latent variables  $\mathbf{Y}_\ell$  can be partitioned as  $\{Y_{1,\ell}, \dots, Y_{n_\ell, \ell}\} = \{Y_{1,\ell}^*, \dots, Y_{k_\ell, \ell}^*\} \cup \mathcal{Y}_0^{(\ell)}$ , where

$$\mathcal{Y}_0^{(\ell)} = \bigcup_{i=2}^{\gamma} \bigcup_{(j_1, \dots, j_i) \in \mathcal{C}_i^{(\ell)}} \left\{ Y_{j_1, \dots, j_i}^{(1)}, \dots, Y_{j_1, \dots, j_i}^{(k_{j_1, \dots, j_i})} \right\}$$

is the set of latent variable values shared by the  $\ell$ -th group with one or more other groups among the remaining  $\gamma - 1$ . Note that  $Y_{j_1, \dots, j_i}^{(r)}$  is the  $r$ th value among the  $k_{j_1, \dots, j_i}$  shared by the  $i$  groups indexed by  $j_1, \dots, j_i$ . Clearly, one may have  $\mathcal{C}_i^{(\ell)} = \emptyset$  for some or all  $i$ . In the latter case,  $\mathcal{Y}_0^{(\ell)} = \emptyset$  and the  $\ell$ -th sample does not share any latent variable value with other samples. Given this notation, for each  $\ell = 1, \dots, \gamma$ , the obvious constraints  $1 \leq k_\ell + \sum_{i=2}^{\gamma} \sum_{\mathcal{C}_i^{(\ell)}} k_{j_1, \dots, j_i} \leq n_\ell$  hold true and the frequencies  $n_{j, \ell} = \sum_{i=1}^{n_\ell} \mathbb{1}_{\{Y_{i, \ell} = Y_{j, \ell}^*\}}$ ,  $q_{j_1, \dots, j_i, r}^\ell = \sum_{m=1}^{n_\ell} \mathbb{1}_{\{Y_{m, \ell} = Y_{j_1, \dots, j_i}^{(r)}\}}$ , for any  $j = 1, \dots, k_\ell$ ,  $(j_1, \dots, j_i) \in \mathcal{C}_i^{(\ell)}$  and  $r \in \{1, \dots, k_{j_1, \dots, j_i}\}$ , are such that

$$n_\ell = \sum_{j=1}^{k_\ell} n_{j, \ell} + \sum_{(j_1, \dots, j_i) \in \mathcal{C}_i^{(\ell)}} \sum_{r=1}^{k_{j_1, \dots, j_i}} q_{j_1, \dots, j_i, r}^\ell.$$

In order to provide an explicit characterization of the posterior for the more general model displayed in (23), let us introduce independent CRMs  $\mu_0^*, \mu_1^*, \dots, \mu_\gamma^*$  with respective intensity measures  $\nu_0^*(ds, dy) = c(1-z) \exp(-s \sum_{i=1}^{\gamma} K_i(y)) \rho(s) ds P_0(dy)$ , and  $\nu_\ell^*(ds, dy) = cz \exp(-s K_\ell(y)) \rho(s) ds P_0(dy)$ , for  $\ell = 1, \dots, \gamma$ . Moreover, sequences of independent and identically distributed latent variables  $\mathbf{V}_\ell = (V_{i, \ell})_{i \geq 1}$  are defined, for each  $\ell \in \{1, \dots, \gamma\}$ , by  $\mathbb{P}[V_{i, \ell} = 0] = 1 - \mathbb{P}[V_{i, \ell} = 1] = z$ . One can, then, replicate the strategy used in the proof of Proposition 2 and show the validity of the following.

**PROPOSITION 4.** *Let  $\mathbf{X}_1, \dots, \mathbf{X}_\gamma$  be vectors of (exact) survival times for which model assumptions summarized in (23) hold true. Then, given  $(\mathbf{Y}_1, \dots, \mathbf{Y}_\gamma)$  and  $(\mathbf{V}_1, \dots, \mathbf{V}_\gamma)$ , the posterior distribution of  $(\tilde{h}_1, \dots, \tilde{h}_\gamma)$  equals the distribution of*

$$(\tilde{h}_1^*, \dots, \tilde{h}_\gamma^*) + \sum_{\ell=1}^{\gamma} \sum_{i=1}^{k_\ell} \mathbf{W}_{i, \ell} k(\cdot; Y_{i, \ell}^*) J_{i, \ell} + \sum_{i=2}^{\gamma} \sum_{(j_1, \dots, j_i) \in \mathcal{C}_i} \sum_{r=1}^{k_{j_1, \dots, j_i}} \mathbf{1}_\gamma k(\cdot; Y_{j_1, \dots, j_i}^{(r)}) J_{j_1, \dots, j_i}^{(r)},$$



where both  $\mathbf{W}_{i,\ell}$  and  $\mathbf{1}_\gamma$  are in  $\{0,1\}^\gamma$ : the former has all components equal to  $V_{i,\ell}$  but the  $\ell$ -th which is 1 and the latter has entries all equal to 1. Moreover

- (i)  $\tilde{h}_\ell^*(t) = \int_{\mathbb{Y}} k(t; y) \tilde{\mu}_\ell^*(dy)$  and  $\tilde{\mu}_\ell^* = \mu_\ell^* + \mu_0^*$ , for each  $\ell = 1, \dots, \gamma$ ;
- (ii) the jumps  $J_{1,\ell}, \dots, J_{k_\ell,\ell}$ , for  $\ell = 1, \dots, \gamma$ , and  $J_{j_1, \dots, j_i}^{(r)}$  for  $(j_1, \dots, j_i) \in \mathcal{C}_i$ ,  $i \in \{2, \dots, \gamma\}$  and  $r \in \{1, \dots, k_{j_1, \dots, j_i}\}$  are mutually independent and independent from  $(\tilde{h}_1^*, \dots, \tilde{h}_\gamma^*)$
- (iii) The jumps  $J_{i,\ell}$  and  $J_{j_1, \dots, j_i}^{(r)}$  have densities equal to  $f(\cdot | n_{i,\ell}, K_\ell(Y_{i,\ell}^*) + V_{i,\ell} \sum_{j \neq \ell} K_j(Y_{i,\ell}^*))$  and  $f(\cdot | \sum_{\ell \in \{j_1, \dots, j_i\}} q_{j_1, \dots, j_i, r}^\ell, \sum_{\ell=1}^\gamma K_\ell(Y_{j_1, \dots, j_i}^{(r)}))$ , respectively, with  $f(\cdot | \cdot, \cdot)$  as in (19).

The proof is omitted since it reproduces the one of Proposition 2, with obvious changes.

## 5 Dependent extended gamma processes

Even if the result stated in Proposition 2 is general, for illustration purposes we focus on a special case where each  $\tilde{\mu}_\ell$ , for  $\ell = 1, 2$ , is a gamma CRM. In this case,  $(\tilde{h}_1^*, \tilde{h}_2^*)$  in (20) is a vector of dependent RHRs obtained from dependent gamma CRMs, though not of GM-type. The conditional densities of the jumps can also be easily determined and it turns out that  $J_{i,1} \sim \text{Ga}(1 + K_1(Y_{i,1}^*) + V_{i,1} K_2(Y_{i,1}^*); n_{i,1})$  for any  $i = 1, \dots, k_1$ ,  $J_{j,2} \sim \text{Ga}(1 + V_{j,2} K_1(Y_{j,2}^*) + K_2(Y_{j,2}^*); n_{j,2})$  for any  $j = 1, \dots, k_2$  and  $J_m \sim \text{Ga}(1 + K_1(Y_m^*) + K_2(Y_m^*); q_{m,1} + q_{m,2})$  for  $m = 1, \dots, k_0$ .

We further suppose that  $\mathbb{Y} = \mathbb{R}^+$  and set  $k(t; y) := \mathbb{1}_{(0,t]}(y) \beta(y)$  for some positive right-continuous real-valued function  $\beta$ . With such a choice, each  $\tilde{h}_i$  marginally is an *extended gamma process* with parameters  $(cP_0, \beta)$ . In the case of exchangeable survival data, such a prior has been proposed in Dykstra and Laud (1981) for modelling increasing hazard rates. In the following we set  $\beta(t) = \beta$  for a any  $t > 0$  and we specify a prior for  $\beta$ . One correspondingly has, for  $\ell = 1, 2$  and  $t > 0$ ,

$$K_\ell(y) = \beta \sum_{i=1}^{n_\ell} (X_{i,\ell} - y) \mathbb{1}_{[y, +\infty)}(X_{i,\ell}), \quad \bar{K}_t(y) = \beta (t - y) \mathbb{1}_{[y, +\infty)}(t). \quad (24)$$

Estimators of  $\tilde{S}_1(t)$  and  $\tilde{S}_2(t)$ , under square loss functions, can be determined for any  $t > 0$  by specializing the result stated in Proposition 3. To this end, and with no loss of generality, we suppose that, for  $\ell = 1, 2$ , the observations  $\mathbf{X}_\ell$  are ordered so that  $X_{i+1,\ell} \leq X_{i,\ell}$  for each  $i = 1, \dots, n_\ell$ . Moreover we denote with  $\tilde{\mathbf{X}} = (\tilde{X}_1, \dots, \tilde{X}_{n_1+n_2})$  the vector including the components of both  $\mathbf{X}_1$  and  $\mathbf{X}_2$  in such a way that  $\tilde{X}_{i+1} \leq \tilde{X}_i$ , for  $i = 1, \dots, n_1 + n_2$ . In order to simplify notation,

we set  $\mathbf{L} = (\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{V}_1, \mathbf{V}_2)$  and  $\Theta = (c, z, \beta)$ . As for the general case displayed in Proposition 3, we first determine an explicit expression for  $\mathbb{E}[\tilde{S}_\ell | \mathbf{X}_1, \mathbf{X}_2, \mathbf{L}, \Theta]$  and, then, establish an approximate evaluation of a Bayesian estimator of  $\tilde{S}_\ell$  by means of a suitably devised Gibbs sampler. In order to ease the illustration, we introduce some notation. For every  $\ell = 1, 2, j = 1, \dots, n_\ell$  and  $\kappa = 1, \dots, n_1 + n_2$ , we set  $\xi_{j,\ell} = \sum_{i=1}^j X_{i,\ell}$  and  $\tilde{\xi}_\kappa = \sum_{i=1}^\kappa \tilde{X}_i$ . We suppose that  $\mathbf{Y}_1$  and  $\mathbf{Y}_2$  are partitioned as in Proposition 3 and we define

$$\zeta_{i,1} = \sum_{j=1}^{n_1} (X_{j,1} - Y_{i,1}^*) \mathbb{1}_{[Y_{i,1}^*, \infty)}(X_{j,1}) + V_{i,1} \sum_{j=1}^{n_2} (X_{j,2} - Y_{i,1}^*) \mathbb{1}_{[Y_{i,1}^*, \infty)}(X_{j,2})$$

for  $i = 1, \dots, k_1$ ,

$$\zeta_{i,2} = V_{i,2} \sum_{j=1}^{n_1} (X_{j,1} - Y_{i,2}^*) \mathbb{1}_{[Y_{i,2}^*, \infty)}(X_{j,1}) + \sum_{j=1}^{n_2} (X_{j,2} - Y_{i,2}^*) \mathbb{1}_{[Y_{i,2}^*, \infty)}(X_{j,2})$$

for  $i = 1, \dots, k_2$ , and finally, for any  $m = 1, \dots, k_0$ ,

$$\tilde{\zeta}_m = \sum_{j=1}^{n_1+n_2} (\tilde{X}_j - Y_m^*) \mathbb{1}_{[Y_m^*, \infty)}(\tilde{X}_j).$$

Moreover, we define the functions

$$G_{i,\ell}(t) = \int_{X_{i,\ell} \wedge t}^{X_{i-1,\ell} \wedge t} \log \left( 1 + \frac{\beta(t-y)}{1 + \beta \xi_{i-1,\ell} - \beta(i-1)y} \right) P_0(dy) \quad (25)$$

for  $\ell = 1, 2$  and  $i = 1, \dots, n_\ell + 1$ , provided  $X_{0,\ell} \equiv +\infty$  and  $X_{n_\ell+1,\ell} \equiv 0$ , and

$$\tilde{G}_j(t) = \int_{\tilde{X}_j \wedge t}^{\tilde{X}_{j-1} \wedge t} \log \left( 1 + \frac{\beta(t-y)}{1 + \beta \tilde{\xi}_{i-1} - \beta(i-1)y} \right) P_0(dy) \quad (26)$$

for  $j = 1, \dots, n_1 + n_2 + 1$ , where  $\tilde{X}_0 \equiv +\infty$  and  $\tilde{X}_{n_1+n_2+1} \equiv 0$ .

**COROLLARY 1.** *Suppose that  $(\tilde{h}_1, \tilde{h}_2)$  is a vector of GM-dependent RHRs. Moreover,  $k(t, y) = \beta \mathbb{1}_{(0,t]}(y)$ . Assuming a square loss function, the estimator of  $\tilde{S}_1(t)$ , for any  $t > 0$ , conditional on the latent variables  $\mathbf{L}$  is*

$$\exp \left( -c z \sum_{i=1}^{n_1} G_{i,1}(t) - c(1-z) \sum_{j=1}^{n_1+n_2} \tilde{G}_j(t) \right) \prod_{i=1}^{k_1} \left( 1 + \frac{\beta(t - Y_{i,1}^*) \mathbb{1}_{[Y_{i,1}^*, +\infty)}(t)}{1 + \beta \zeta_{i,1}} \right)^{-n_{i,1}}$$

$$\times \prod_{j=1}^{k_2} \left( 1 + V_{j,2} \frac{\beta(t - Y_{j,2}^*) \mathbb{1}_{[Y_{j,2}^*, +\infty)}(t)}{1 + \beta \zeta_{j,2}} \right)^{-n_{j,2}} \prod_{m=1}^{k_0} \left( 1 + \frac{\beta(t - Y_m^*) \mathbb{1}_{[Y_m^*, +\infty)}(t)}{1 + \beta \tilde{\zeta}_m} \right)^{-(q_{m,1} + q_{m,2})}. \quad (27)$$

Note that (27) turns out to be very convenient from an operational point of view. Indeed, one can integrate out the vector of latent variables  $\mathbf{L}$  and obtain an estimate of  $\tilde{S}_1(t)$  conditionally on the observations  $(\mathbf{X}_1, \mathbf{X}_2)$ . For computational purposes, in the sequel we will set  $P_0(dx) = \mathbb{1}_{[0, T]}(x)/T$ , for some  $T > 0$  large enough so that  $\tilde{X}_1 < T$ . One can think of  $T$  as, for example, a maximum follow-up time in a survival analysis experiment. This is a specific choice we stick to since it simplifies some of the computations we shall perform for the MCMC sampling scheme. Clearly, any  $P_0$  on  $\mathbb{R}^+$  would work.

The results in this section allow us to devise an MCMC algorithm that we will use to estimate  $\tilde{S}_\ell$ , for  $\ell = 1, 2$ , when  $(\tilde{h}_1, \tilde{h}_2)$  are dependent extended gamma processes. In order to marginalize  $\mathbb{E}[\tilde{S}_\ell | \mathbf{X}_1, \mathbf{X}_2, \mathbf{L}, \Theta]$ , for  $\ell = 1, 2$ , with respect to hyperparameters and the latent variables, we determine the full conditional distributions within a Gibbs sampler that allows to generate samples from the posterior distribution of  $(\mathbf{L}, \Theta)$  given the observations  $(\mathbf{X}_1, \mathbf{X}_2)$ . The details are provided in Appendix F. Furthermore, though we confine ourselves to the specific extended gamma case, the general distributional representations stated in Section 4 make it feasible to implement any GM-dependent RHR model, for any choice of GM-dependent CRMs  $(\tilde{\mu}_1, \tilde{\mu}_2)$  and of the mixing kernel  $k(\cdot, \cdot, \cdot)$ .

## 6 Illustrations

We provide two illustrative examples where we model two hazard rates using dependent extended gamma processes and we apply the devised algorithm to estimate the corresponding dependent survival functions, given two samples  $\mathbf{X}_1$  and  $\mathbf{X}_2$  of possibly right-censored survival times. As a byproduct, we provide estimates of functionals such as mean and median lifetimes for the two groups that we obtain by numerically evaluating such functionals for the estimated survival functions.

The goal of our analysis is two-fold: on one side we aim at showing that, by dropping the proportionality of hazards assumption, we obtain a model that allows for crossing survival functions; on the other side we want to highlight the borrowing strength mechanism that under certain circumstances may lead the survival curves to shrink. In order to develop both examples we need to introduce censored data and include them in the posterior representation that has been determined

in Proposition 2. To this end, we set  $C_{i,\ell}$  to be the censoring time corresponding to  $X_{i,\ell}$  and define  $\Delta_{i,\ell} = \mathbb{1}_{(0, C_{i,\ell}]}(X_{i,\ell})$  so that  $\Delta_{i,\ell}$  is either 0 or 1 according as to whether  $X_{i,\ell}$  is either censored or not, respectively. Hence, the actual observation is  $T_{i,\ell} = \min\{X_{i,\ell}, C_{i,\ell}\}$  and the data consist of the collections of pairs  $\mathbf{D}_1 = \{(T_{j,1}, \Delta_{j,1})\}_{j=1}^{n_1}$  and  $\mathbf{D}_2 = \{(T_{j,2}, \Delta_{j,2})\}_{j=1}^{n_2}$ . These information yield the number of exact and censored observations in each group, namely  $n_\ell^{(e)} = \sum_{j=1}^{n_\ell} \Delta_{j,\ell}$  and  $n_\ell^{(c)} = n_\ell - n_\ell^{(e)}$ . We shall, then, consider the data as ordered, namely  $T_{i+1,\ell} \leq T_{i,\ell}$  for any  $i = 1, \dots, n_\ell - 1$  and for each  $\ell = 1, 2$  and identify the whole set of ordered data as  $\tilde{T}_1 > \tilde{T}_2 > \dots > \tilde{T}_{n_1+n_2}$ . The extension of the distributional results in Section 4 to the case that includes right-censored data is quite straightforward once the likelihood in (15) is rewritten as

$$\mathcal{L}(\mu_1, \mu_2; \mathbf{D}_1, \mathbf{D}_2) = \prod_{\ell=1}^2 \exp\left(-\int_{\mathbb{Y}} K_\ell^*(y) \mu_\ell(dy)\right) \prod_{\{i: \Delta_{i,\ell}=1\}} \int_{\mathbb{Y}} k(T_{i,\ell}; y) \mu_\ell(dy),$$

where

$$K_\ell^*(y) = \sum_{i=1}^{n_\ell} \int_0^{T_{i,\ell}} k(s; y) ds, \quad \ell = 1, 2 \quad (28)$$

and the censored survival times, i.e. those with  $\Delta_{i,\ell} = 0$ , are involved in the likelihood function only through  $K_1^*$  and  $K_2^*$ . The results derived in the previous sections under the assumption that only exact observations are detected, e.g. Propositions 2 and Corollary 1, carry over to the case with right-censored observations: the only changes concern  $K_\ell$  and the jump components which occur only at the distinct values of the latent variables that correspond to exact observations. For example, in (20) and (27) one has  $n_{j,\ell} = \sum_{\{i: \Delta_{i,\ell}=1\}} \mathbb{1}_{\{Y_{i,\ell}=Y_{j,\ell}^*\}}$ , and  $q_{m,\ell} = \sum_{\{i: \Delta_{i,\ell}=1\}} \mathbb{1}_{\{Y_{i,\ell}=Y_m^*\}}$ , for any  $\ell = 1, 2$ ,  $j = 1, \dots, k_\ell$  and  $m = 1, \dots, k_0$ . The non-negative integers  $k_1$ ,  $k_2$  and  $k_0$  are subject to the constraints  $0 \leq k_\ell + k_0 \leq n_\ell^{(e)}$ , for  $\ell = 1, 2$ . The tilted Lévy intensities in (18) of the component without fixed points of discontinuity in the posterior are

$$\begin{aligned} \nu_0^*(dt, dy) &= c(1-z) \exp(-(K_1^*(y) + K_2^*(y))t) \rho(t) dt P_0(dy), \\ \nu_\ell^*(dt, dy) &= cz \exp(-K_\ell^*(y)t) \rho(t) dt P_0(dy) \end{aligned}$$

for  $\ell = 1, 2$ . Moreover, the parameters of the densities of the jumps in Proposition 2(iii) are such that  $K_\ell$  is replaced by  $K_\ell^*$  and the frequencies  $n_{j,\ell}$  and  $q_{m,\ell}$  are defined as above. The adaptation of the full conditionals distributions obtained in the previous section and in Appendix F is, then, straightforward.

We can now illustrate an application of our model to both simulated and real datasets. The former experiment is devised in such a way that the two samples are generated from distributions with intersecting hazard rates, a situation that cannot be fitted through a proportional hazards model specification. The latter refers to remission times of two separate groups of leukemia patients receiving different treatments. In both cases, in order to obtain a method that is stable under rescaling of the data, we divide all lifetimes by  $\tilde{T}_1$  thus implying that all components in  $\mathbf{T}_\ell = (T_{1,\ell}, \dots, T_{n_\ell,\ell})$  are, for any  $\ell = 1, 2$ , in  $[0, 1]$ . Finally, we set  $T = 1.5$  as the maximum follow-up time. This leads to a vector of scaled hazard functions  $(\tilde{r}_1, \tilde{r}_2)$  that are related to the unscaled hazards  $(\tilde{h}_1, \tilde{h}_2)$  by means of  $(T \tilde{r}_1(t_1), T \tilde{r}_2(t_2)) = (\tilde{h}_1(t_1/T), \tilde{h}_2(t_2/T))$ . After computing the posterior of  $(\tilde{r}_1, \tilde{r}_2)$ , we recover the posterior estimate of the vector  $(\tilde{h}_1, \tilde{h}_2)$  by transforming back. The described procedure is similar to the one proposed in Ishwaran and James (2004). The framework is completed by specifying the prior distribution for the hyperparameters that, in our examples, are set as  $z \sim \mathcal{U}[0, 1]$ ,  $c \sim \Gamma(1, 10)$  and  $\beta \sim \Gamma(1, 10)$ . The choice of a uniform distribution for the parameter  $z$  formalizes the idea that, a priori, we do not have any reason to favor total exchangeability ( $z = 0$ ), independence of the two samples ( $z = 1$ ) or any situation in between ( $z \in (0, 1)$ ). The parameters of the prior distributions of  $c$  and  $\beta$  were chosen in order to obtain a large variance, namely  $\sigma^2 = 100$ . Finally, the estimates in the examples that follow were obtained by running the MCMC algorithm for 1,000,000 iterations, after 50,000 of burn-in. After thinning, we have stored a sample of 2000 iterations, for which the considered diagnostic tools suggest convergence of the chain and indicate excellent mixing. In particular the chains corresponding to  $\mathbb{E}[\tilde{S}_1(t)|\mathbf{X}_1, \mathbf{X}_2, \mathbf{L}, \Theta]$  and  $\mathbb{E}[\tilde{S}_2(t)|\mathbf{X}_1, \mathbf{X}_2, \mathbf{L}, \Theta]$ , evaluated on a grid of values  $t \in [0, 2 \max\{\mathbf{X}_1, \mathbf{X}_2\}]$ , have effective sample size, on average, equal to about 1900.

REMARK 4. Since the MCMC algorithm we propose is based on the marginalization with respect to the underlying vector of CRMs  $\tilde{\boldsymbol{\mu}} = (\tilde{\mu}_1, \tilde{\mu}_2)$ , the output of the Gibbs sampler does not take into account the uncertainty associated to  $\tilde{\boldsymbol{\mu}}$ , given the data. Hence, Bayesian credible intervals based on the MCMC output tend to be narrower than the actual ones: this is an important issue that typically arises in Bayesian nonparametrics when so-called marginal methods are used. In order to clarify this point, we consider, as a measure of dispersion of the posterior distribution of  $\tilde{S}_1(t)$ , its variance  $\text{Var}_{\boldsymbol{\mu}}(\tilde{S}_1(t)|\mathbf{X}_1, \mathbf{X}_2)$ , where the subscript  $\boldsymbol{\mu}$  indicates the random variable that is integrated out. A classic result about conditional variance allows us to decompose  $\text{Var}_{\boldsymbol{\mu}}(\tilde{S}_1(t)|\mathbf{X}_1, \mathbf{X}_2)$  in the sum of two components, namely  $\text{Var}_{\mathbf{L}, \Theta}(\mathbb{E}_{\boldsymbol{\mu}}[\tilde{S}_1(t)|\mathbf{X}_1, \mathbf{X}_2, \mathbf{L}, \Theta])$ , that takes into account the volatility (w.r.t.  $\mathbf{L}$  and  $\Theta$ ) of the posterior mean (w.r.t.  $\boldsymbol{\mu}$ ), and  $\mathbb{E}_{\mathbf{L}, \Theta}[\text{Var}_{\boldsymbol{\mu}}(\tilde{S}_1(t)|\mathbf{X}_1, \mathbf{X}_2, \mathbf{L}, \Theta)]$  that is related to the

expected (w.r.t.  $\mathbf{L}$  and  $\Theta$ ) posterior volatility (w.r.t  $\mu$ ). Analogously, the intervals obtained with the MCMC output allow to quantify only the volatility that is related to the first component, that is the uncertainty around the posterior mean. Nonetheless, we are able to control the loss in posterior uncertainty quantification by studying the ratio

$$\frac{\text{Var}_{\mathbf{L}, \Theta}(\mathbb{E}_{\mu}[\tilde{S}_1(t)|\mathbf{X}_1, \mathbf{X}_2, \mathbf{L}, \Theta])}{\text{Var}_{\mu}(\tilde{S}_1(t)|\mathbf{X}_1, \mathbf{X}_2)}. \quad (29)$$

A numerical evaluation of (29), over a grid of values of  $t$ , suggests that, for the example considered in this section, credible intervals obtained by means of the MCMC output, although smaller, provide, in average, a reasonable approximation to the real ones. As an alternative measure of uncertainty one can consider  $\sigma^2(t) = \text{Var}_{\mu}(\tilde{S}_1(t)|\mathbf{X}_1, \mathbf{X}_2)$  and use

$$\left( \max \left( \hat{S}_1(t) - 3\hat{\sigma}(t), 0 \right), \min \left( \hat{S}_1(t) + 3\hat{\sigma}(t), 1 \right) \right), \quad (30)$$

which, by Chebychev inequality, contain approximately 90% credible intervals. Details on the evaluation of (29) and  $\sigma^2(t) = \text{Var}_{\mu}(\tilde{S}_1(t)|\mathbf{X}_1, \mathbf{X}_2)$  are provided in Appendix E. A similar line of reasoning can be obviously replicated for  $\tilde{S}_2(t)$ .

## 6.1 Simulation study

In survival analysis problems, it is often reasonable to drop the proportionality of hazards assumption that is behind the widely used Cox proportional model. Several ways to extend Cox model have appeared in literature. See, among others, Lee et al. (1993), Kalbfleish and Prentice (2002), Klein and Moeschberger (1997), De Iorio et al. (2009), Hanson et al. (2012), Choi and Huang (2012) and a nice recent overview based on the theory of counting processes in Zeng and Lin (2007). It is worth mentioning that models such as the accelerated failure time, while dropping the proportionality assumption, do not accommodate for crossing survival functions. One of the simplest extensions that circumvents this problem, consists in stratifying on treatments. In the context of nonparametric hazard functions defined through CRMs, this could be done by assuming as independent the CRMs underlying the hazard functions of groups of patients under different treatments. It is apparent that this approach has the drawback of not allowing borrowing information across groups. Through this example we aim at showing that the model we propose, while inducing a borrowing of information effect, well accommodates for non proportional hazards and can also lead to estimate crossing survival

functions. In a similar fashion as it was done in De Iorio et al. (2009), we simulate two-sample datasets of several sizes, from distributions that correspond to crossing survival functions. More specifically, we consider two Weibull distributions  $f_1 \sim \text{Wbl}(2, 3/2)$  and  $f_2 \sim \text{Wbl}(2, 3)$ , where  $f \sim \text{Wbl}(a, b)$  means that  $f$  is a density function on the positive real line such that  $f(x) = (b/a)(x/a)^{b-1} \exp(-x/a)^b$ . Hence, two samples of exact observations of size  $n_1 = n_2 = n$ , with  $n$  that ranges in  $\{25, 50, 100\}$  are generated from  $f_1$  and  $f_2$ . In Figure 1 we compare estimated and true survival functions. It is not

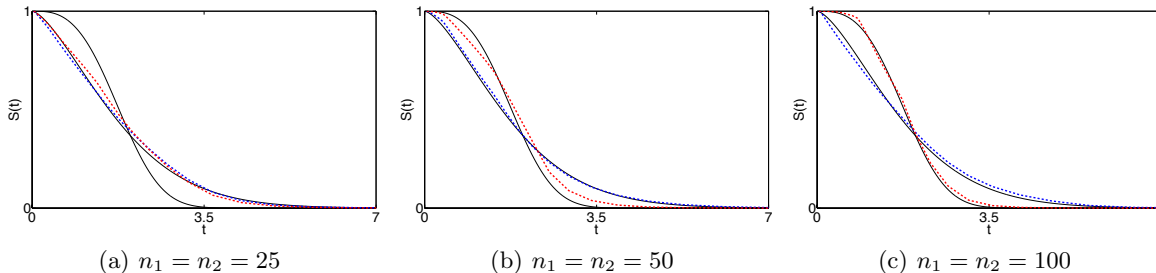


Figure 1: Comparison of true (solid lines) and estimated (dashed lines) survival functions, for different sample sizes.

surprising that the estimates improve as the sample size  $n$  increases: while the survival curves shrink significantly for  $n = 25$ , for larger sample size the fact that the two survival curves cross is better detected. Such trend is confirmed by the numerical evaluation of the Kolmogorov distance  $d_K(\hat{S}_1, \hat{S}_2)$  proposed in the second row of Table 1, where  $d_K(S_1, S_2) = \sup_{t \geq 0} |S_1(t) - S_2(t)|$ . This seems like an interesting feature of the algorithm since sharing information across studies is particularly convenient when the samples have small size. A comparison of rows 3 and 4 with rows 5 and 6 of Table 1 shows that, for every considered value of  $n$ , the estimates our model provides are closer to the true survival curves than the corresponding Kaplan-Meier estimators. Finally, the crossing time (around  $t = 2$ ) is well estimated for  $n = 100$ .

	$n = 25$	$n = 50$	$n = 100$
$d_K(S_1, S_2)$	0.181	0.181	0.181
$d_K(\hat{S}_1, \hat{S}_2)$	0.051	0.118	0.232
$d_K(\hat{S}_1, S_1)$	0.029	0.025	0.042
$d_K(\hat{S}_2, S_2)$	0.149	0.071	0.032
$d_K(\text{KM}_1, S_1)$	0.106	0.041	0.071
$d_K(\text{KM}_2, S_2)$	0.166	0.132	0.051

Table 1: Kolmogorov distances between true survival functions (row 1), survival functions estimated via GM-dependent model (row 2), true and estimated survivals (row 3 and 4 for estimates obtained via GM-dependent model, rows 5 and 6 for Kaplan-Meier estimates).

We generate also two-sample datasets, of sizes  $n_1 = n_2 = 100$ , that involve right-censored observations. To this end, we first select uniformly a portion of the observations from each sample and censor them by means of censoring times that, conditional on the exact observations  $X_{i,\ell}$  (i.e., with  $\Delta_{i,\ell} = 1$ ), are distributed uniformly on  $[0, T_{i,\ell}]$ , for each  $\ell = 1, 2$  and  $i = 1, \dots, n_\ell^{(c)}$ . More specifically, we have considered three datasets with same portion of censored observations in both samples, namely 10%, 20% and 30%. In Table 2 we report estimated mean and median lifetimes for all the

	true	$n = 25$	$n = 50$	$n = 100$	10%	20%	30%
mean $\mathbf{X}_1$	1.81	1.81	1.84	1.83	1.90	1.96	2.04
95% c.i.		(1.49,2.22)	(1.56,2.17)	(1.64,2.04)	(1.70,2.13)	(1.76,2.20)	(1.81,2.31)
mean $\mathbf{X}_2$	1.79	1.85	1.86	1.83	1.87	1.87	1.88
95% c.i.		(1.55,2.26)	(1.62,2.13)	(1.71,1.96)	(1.74, 2.01)	(1.74,2.00)	(1.73,2.05)
median $\mathbf{X}_1$	1.57	1.60	1.61	1.59	1.65	1.67	1.77
95% c.i.		(1.27,1.97)	(1.36,1.83)	(1.39,1.78)	(1.44,1.85)	(1.47,1.86)	(1.55,2.00)
median $\mathbf{X}_2$	1.77	1.69	1.84	1.81	1.85	1.82	1.80
95% c.i.		(1.42,1.98)	(1.64,2.06)	(1.68,1.90)	(1.72,1.93)	(1.71,1.91)	(1.66,1.91)

Table 2: Estimated mean and median expected lifetimes for the two samples and corresponding 95%-credible intervals. First column provides the true values for mean and median lifetime of the two samples. Columns 2–4 refer to datasets of exact observations of size  $n$  equal to 25, 50 and 100, respectively. Last three columns refer to datasets of size  $n = 100$  with, respectively, 10%, 20% and 30% of right-censored observations.

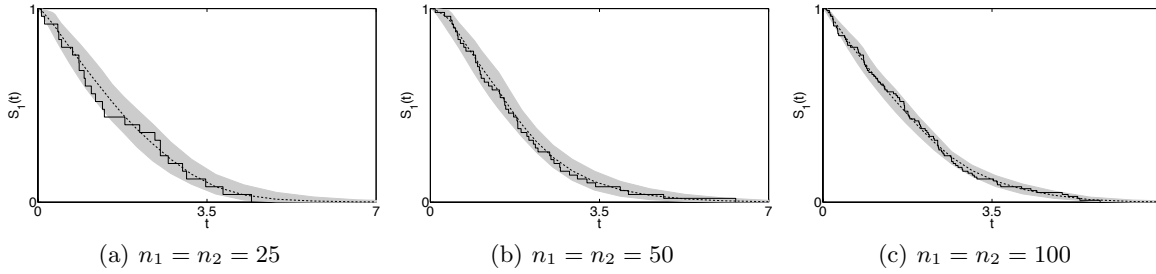


Figure 2: Kaplan-Meier estimate (continuous line), estimated survival (dashed line) with 95%-credible intervals for the mean (gray area), for first group only and for different sample sizes.

synthetic datasets we have considered, together with approximate 95%-credible intervals. As for the two-sample datasets with only exact observations (columns 2–4), it is apparent that intervals for both estimated mean and median lifetimes of the two samples tend to narrow around the true values as the sample size increases. Similarly, Figure 2 shows that the estimated credible intervals, for the vector of observations  $\mathbf{X}$ , shrink around the true curve for large values of  $n$ . On the contrary, the more right-censored observations are present the larger are the intervals for mean and median lifetimes (see



columns 5–7 of Table 2) and the less precise are the corresponding estimates.

## 6.2 Leukemia datasets

We consider two well known two-sample datasets involving leukemia remission times, in weeks, under different treatments. The first was studied, e.g., in Cox (1972) and consists of treatment versus placebo; the second was analyzed in Lawless (1982) and compares the effects of two different therapies. Both datasets were studied more recently in Damien and Walker (2002) where the focus is on discerning situations of full exchangeability from situations of partial exchangeability.

*Dataset 1.* The remission times, expressed in weeks, of patients under an active drug are

$$\{6, 6, 6, 6^*, 7, 9^*, 10, 10^*, 11, 13, 16, 17^*, 19^*, 20^*, 22, 23, 25^*, 32^*, 32^*, 34^*, 35^*\},$$

where a star denotes a censored observation. The remission times of patients under placebo are all exact and coincide with

$$\{1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 11, 11, 12, 12, 15, 17, 22, 23\}.$$

*Dataset 2.* Two groups of patients subject to two different treatments, A and B, are considered. The remission times of patients under treatment A are

$$\{1, 3, 3, 6, 7, 7, 10, 12, 14, 15, 18, 19, 22, 26, 28^*, 29, 34, 40, 48^*, 49^*\},$$

while those of patients under treatment B are

$$\{1, 1, 2, 2, 3, 4, 5, 8, 8, 9, 11, 12, 14, 16, 18, 21, 27^*, 31, 38^*, 44\}.$$

For both datasets, our analysis proceeds in two directions. First, in Figures 3(a) and 3(c), we report the estimated survival functions, with estimated 95% credible intervals, and compare them with the Kaplan–Meier estimates. Second, we compare the estimates provided by our GM-dependent model with those that are provided by the equivalent stratified model, that is a model where  $z = 1$  and, therefore,  $\tilde{h}_1$  and  $\tilde{h}_2$  are independent (for a complete description of the stratified model we need to mention that  $\tilde{h}_1$  and  $\tilde{h}_2$  do not share the same hyperparameters  $c$  and  $\beta$ ). By inspecting Figures

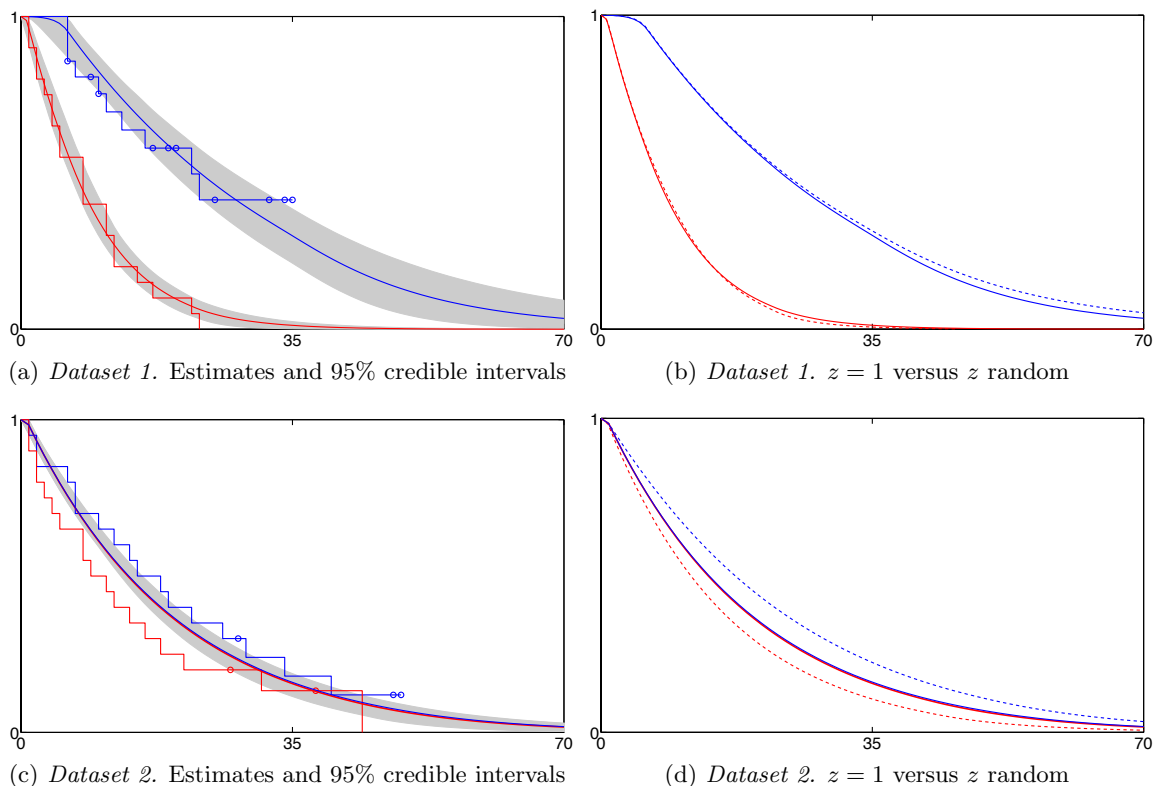


Figure 3: [*Leukemia datasets.*] In (a) and (c) survival functions estimated through a GM-dependent extended gamma processes model, with 95% credible intervals, and Kaplan–Meier estimates. Circles denote the presence of a censor. In (b) and (d) comparison of the survival functions estimated through a GM-dependent extended gamma processes model with  $z$  random (solid lines) and the estimates obtained with a stratified models, i.e.  $z = 1$  (dashed lines).

3(b) and 3(d), it is evident that while, for the first dataset, the estimates obtained by the two models do not differ significantly, for the second dataset, the model based on GM-dependent CRMs gives rise to substantial shrinkage phenomenon that makes the two survival curves hardly distinguishable. This indicates a heavier borrowing of information across the two studies. Such an intuition is straightened by looking at the distances between estimated survivals under the two settings, reported in Table 3: for each dataset, the ratio between the distance between survivals estimated via GM-dependent and independent models is equal to 0.99 for the first dataset and 0.04 for the second. Our observation is

Dataset 1		Dataset 2	
$z$ random	$z = 1$	$z$ random	$z = 1$
0.498	0.503	0.005	0.113

Table 3: Kolmogorov distances  $d_K(\hat{S}_1, \hat{S}_2)$  between estimated survival functions under GM-dependent model (columns 1,3) and independent models (columns 2,4).

supported by Damien and Walker (2002) where similar conclusions on the two cases were derived and, as for the second dataset, by Lawless (1982) that, by means of frequentist tests, concludes that there is no evidence of difference in the two distributions.

## 7 Concluding remarks

The present paper has proposed a new dependent nonparametric prior process for hazard rate functions associated to heterogeneous survival data. The nice structure featured by the underlying CRMs leads us to obtain a posterior characterization, which is the key for a Bayesian analysis as the one we have performed on both simulated and real datasets. Most of the technicalities related to our model are postponed to the Appendix. Nonetheless we wish to point out that (20) has a simple structure that naturally extends the known results for the posterior in the univariate exchangeable case. Given its simplicity, one can actually use (20) or the representation displayed in Proposition 4 in order to simulate the trajectories of the posterior process instead of performing a marginalization procedure as the one discussed in the present paper. This will be the object of future research.

Two other relevant issues that deserve further investigation in future work concern the analysis of asymptotic properties of the proposed model and an extension to more general covariate indexing. As for the former, one can analyze frequentist asymptotics of GM-dependent RHRs by relying on a suitable adaptation of the approach set forth in Peccati and Prünster (2008) and in De Blasi et al. (2009) for the exchangeable case. As for the latter, a possible variant of our proposal would be the specification of a covariate-dependent proportional hazards model for the  $\ell$ -th group and baseline hazard coinciding with GM-dependent RHRs  $\tilde{h}_\ell$  as the ones we have analyzed in this paper. This would allow to take into account both covariate effect and center heterogeneity in multicenter studies. Possible elaborations in this direction will benefit from the theoretical findings displayed in the previous sections.

## Acknowledgements

The authors are grateful to an Associate Editor and to two Referees for their valuable comments and suggestions that have led to a considerable improvement of the manuscript. Antonio Lijoi's research is supported by the European Research Council (ERC) through StG "N-BNP" 306406.

# Appendix

## A Some basics on completely random measures

Let  $M_{\mathbb{Y}}$  be the set of boundedly finite measures on a complete and separable metric space  $\mathbb{Y}$ . This means that for any bounded set  $A$  in the Borel  $\sigma$ -algebra  $\mathcal{Y}$  on  $\mathbb{Y}$  and for any  $m \in M_{\mathbb{Y}}$ , one has  $m(A) < \infty$ . Moreover, set  $\mathcal{M}_{\mathbb{Y}}$  as the Borel  $\sigma$ -algebras on  $M_{\mathbb{Y}}$ , respectively. See Daley and Vere-Jones (2003) for technical details on the construction of the space  $(M_{\mathbb{Y}}, \mathcal{M}_{\mathbb{Y}})$ .

DEFINITION 2. A *completely random measure* (CRM)  $\mu$  on  $(\mathbb{Y}, \mathcal{Y})$  is a measurable function defined on some probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  and taking values in  $(M_{\mathbb{Y}}, \mathcal{M}_{\mathbb{Y}})$  such that, for any integer  $j \geq 1$  and for any collection  $A_1, \dots, A_j$  of pairwise disjoint sets in  $\mathcal{Y}$ , the random variables  $\mu(A_1), \dots, \mu(A_j)$  are mutually independent.

Any CRM  $\mu$  may be represented as

$$\mu = \mu_c + \mu_0 = \sum_{i=1}^{\infty} J_i \delta_{Y_i} + \sum_{i=1}^L V_i \delta_{y_i}$$

where  $(J_i)_{i \geq 1}$ ,  $(V_i)_{i \geq 1}$  and  $(Y_i)_{i \geq 1}$  are independent sequences of random variables and the jump points  $\{y_1, \dots, y_L\}$  are fixed, with  $L \in \{0, 1, \dots\} \cup \{\infty\}$ . See Kingman (1967). The Lévy–Khintchine representation states that, for every CRM  $\mu_c = \sum_i J_i \delta_{Y_i}$  without fixed jumps, there exists a measure  $\nu$  on  $\mathbb{R}^+ \times \mathbb{Y}$  such that  $\int_{\mathbb{R}^+ \times \mathbb{Y}} \min\{s, 1\} \nu(ds, dy) < \infty$  and

$$\mathbb{E} \left[ \exp \left( - \int_{\mathbb{Y}} f(y) \mu_c(dy) \right) \right] = \exp \left( - \int_{\mathbb{R}^+ \times \mathbb{Y}} [1 - \exp(-s f(y))] \nu(ds, dy) \right) \quad (\text{A.1})$$

for any measurable function  $f : \mathbb{Y} \rightarrow \mathbb{R}$  such that  $\int_{\mathbb{Y}} |f| d\mu_c < \infty$  almost surely. We refer to the measure  $\nu$  as the *Lévy intensity* of  $\mu_c$ . For our purposes, it will be useful to rewrite  $\nu$  as

$$\nu(ds, dy) = \rho_y(s) ds cP_0(dy), \quad (\text{A.2})$$

where  $P_0$  is a probability measure on  $(\mathbb{Y}, \mathcal{Y})$  and  $(y, B) \mapsto \int_B \rho_y(s) ds$  is a transition kernel on  $\mathbb{Y} \times \mathcal{B}(\mathbb{R}^+)$ . If  $\rho_y = \rho$  in (A.2), for any  $y$  in  $\mathbb{Y}$ , the CRM  $\mu_c$  is said *homogeneous*. In the paper, we suppose that  $P_0$  is non-atomic and  $L = 0$ . Hence,  $\mu = \mu_c$  and there are no jumps at fixed discontinuities. A

well-known example corresponds to  $\rho_y(s) = \rho(s) = e^{-s}/s$ , which identifies a so-called *gamma CRM*  $\mu$ . With such a choice of the Lévy intensity, from (A.1) it can be seen that for any measurable function  $f : \mathbb{Y} \rightarrow \mathbb{R}$  such that  $\int \log(1 + |f|) dP_0 < \infty$  one has

$$\mathbb{E} \left[ \exp \left( - \int_{\mathbb{Y}} f(y) \mu(dy) \right) \right] = \exp \left( -c \int_{\mathbb{Y}} \log(1 + f(y)) P_0(dy) \right).$$

Hence, if  $f = \mathbb{1}_A$  is the indicator function of a set  $A$  in  $\mathcal{Y}$  such that  $P_0(A) > 0$ , the random variable  $\mu(A)$  has a  $\text{Ga}(1, cP_0(A))$ .

## B Correlation between GM-dependent RHRs

In the paper it has been mentioned that the determination of the correlation coefficient between the point values  $\tilde{S}_1(t_1)$  and  $\tilde{S}_2(t_2)$  of the survival functions, at any positive  $t_1$  and  $t_2$ , is straightforward and does not require any extra integrability condition for the jump part of the Lévy intensity  $\rho$ . The same cannot be said when it comes to evaluating the correlation coefficient between any two values of the hazard function  $\tilde{h}_1(t_1)$  and  $\tilde{h}_2(t_2)$ , at any  $t_1$  and  $t_2$ . Indeed, in this case, one needs to introduce additional integrability conditions one of which concerns  $\rho$ . From the Proposition below, it can be further noted that the correlation between  $\tilde{h}_1(t_1)$  and  $\tilde{h}_2(t_2)$  displays a linear dependence on  $z$  and, as expected, it is equal to 0 when  $z = 1$ .

PROPOSITION B.1. Let  $(\tilde{h}_1, \tilde{h}_2)$  be GM-dependent RHRs. Suppose also that the following conditions hold true:

- (i)  $\int_{\mathbb{Y}} k(t_1; y) k(t_2; y) P_0(dy) < \infty$  for any positive  $t_1$  and  $t_2$ ,
- (ii)  $\int_0^\infty s^2 \rho(s) ds < \infty$ .

Then the correlation between  $\tilde{h}_1(t_1)$  and  $\tilde{h}_2(t_2)$  is given by

$$\text{Corr}(\tilde{h}_1(t_1), \tilde{h}_2(t_2)) = (1 - z) \frac{\int_{\mathbb{R}^+ \times \mathbb{Y}} s^2 k(t_1; y) k(t_2; y) \rho(s) ds P_0(dy)}{\prod_{i=1}^2 \sqrt{\int_{\mathbb{R}^+ \times \mathbb{Y}} s^2 k(t_i; y)^2 \rho(s) ds P_0(dy)}}.$$

PROOF. To simplify the notation, for  $i = 0, 1, 2$ , we set  $\mu_i(k(t)) := \int_{\mathbb{Y}} k(t; y) \mu_i(dy)$  and note that the determination of the correlation between  $\tilde{h}_1(t_1)$  and  $\tilde{h}_2(t_2)$  boils down to computing moments of

the functionals  $\mu_i(k(t))$ . This can be accomplished by means of the following

$$\begin{aligned}\mathbb{E}[\mu_i(k(t))] &= -\frac{\partial}{\partial \lambda} \mathbb{E}[\exp(-\mu_i(\lambda k(t)))] \Big|_{\lambda=0} = c z^i (1-z)^{1-i} \left\{ \frac{\partial}{\partial \lambda} \psi(\lambda k(t)) \right\}_{\lambda=0}, \\ \mathbb{E}[\mu_i(k(t))^2] &= \frac{\partial^2}{\partial \lambda^2} \mathbb{E}[\exp(-\mu_i(\lambda k(t)))] \Big|_{\lambda=0} = -c z^i (1-z)^{1-i} \left\{ \frac{\partial^2}{\partial \lambda^2} \psi(\lambda k(t)) \right\}_{\lambda=0} \\ &\quad + c z^{2i} (1-z)^{2-2i} \left\{ \frac{\partial}{\partial \lambda} \psi(\lambda k(t)) \right\}_{\lambda=0}^2\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}[\mu_0(k(t_1)) \mu_0(k(t_2))] &= \frac{\partial^2}{\partial \lambda_1 \partial \lambda_2} \mathbb{E}[\exp(-\mu_0(\lambda_1 k(t_1) + \lambda_2 k(t_2)))] \Big|_{\lambda_1=\lambda_2=0} \\ &= -c(1-z) \left\{ \frac{\partial^2}{\partial \lambda_1 \partial \lambda_2} \psi(\lambda_1 k(t_1) + \lambda_2 k(t_2)) \right\}_{\lambda_1=\lambda_2=0} \\ &\quad + c^2 (1-z)^2 \prod_{j=1}^2 \left\{ \frac{\partial}{\partial \lambda_j} \psi(\lambda_j k(t_j)) \right\}_{\lambda_j=0}.\end{aligned}$$

One accordingly has

$$\text{Corr}(\tilde{h}(t_1) \tilde{h}_2(t_2)) = (1-z) \frac{-\left\{ \frac{\partial^2}{\partial \lambda_1 \partial \lambda_2} \psi(\lambda_1 k(t_1) + \lambda_2 k(t_2)) \right\}_{\lambda_1=\lambda_2=0}}{\prod_{j=1}^2 \sqrt{-\left\{ \frac{\partial^2}{\partial \lambda_j^2} \psi(\lambda_j k(t_j)) \right\}_{\lambda_j=0}}}.\quad (\text{B.1})$$

If  $\int_{\mathbb{Y}} k(t_1; y) k(t_2; y) P_0(dy) < \infty$  for any  $t_1$  and  $t_2$  and  $\int_0^\infty s^2 \rho(s) ds < \infty$ , one may write

$$\left\{ \frac{\partial^2}{\partial \lambda_1 \partial \lambda_2} \psi(\lambda_1 k(t_1) + \lambda_2 k(t_2)) \right\}_{\lambda_1=\lambda_2=0} = - \int_{\mathbb{R}^+ \times \mathbb{Y}} s^2 k(t_1; y) k(t_2; y) \rho(s) ds P_0(dy)\quad (\text{B.2})$$

and, similarly,

$$\left\{ \frac{\partial^2}{\partial \lambda^2} \psi(\lambda k(t)) \right\}_{\lambda=0} = - \int_{\mathbb{R}^+ \times \mathbb{Y}} s^2 k(t; y)^2 \rho(s) ds P_0(dy).\quad (\text{B.3})$$

The proof is complete when (B.2) and (B.3) are plugged in (B.1).  $\square$

## C Proof of Proposition 2

We shall resort to an approach set forth in Regazzini and Sazonov (2000) that makes use of a discretization process. First note that since  $\mathbb{Y}$  is separable there exists a sequence  $(\Pi_r)_{r \geq 1}$  of measurable

partitions of  $\mathbb{Y}$ , with  $\Pi_r = \{A_{r,i} : i = 1, \dots, \kappa_r + 1\}$ , such that: (a)  $\Pi_{r+1}$  is a refinement of  $\Pi_r$ ; (b) if  $\mathcal{G}_r = \sigma(\Pi_r)$ , then  $\mathcal{Y} = \sigma(\cup_{r \geq 1} \mathcal{G}_r)$ ; (c)  $\max_{1 \leq i \leq \kappa_r + 1} \text{diam}(A_{r,i}) \rightarrow 0$  as  $r \rightarrow \infty$ . Accordingly, define sequences  $(Y_{r,i}^{(1)})_{i \geq 1}$  and  $(Y_{r,i}^{(2)})_{i \geq 1}$  of  $\mathbb{Y}$ -valued random elements with

$$Y_{r,i}^{(\ell)} = \sum_{h=1}^{\kappa_r+1} y_{r,h,\ell}^* \delta_{A_{r,h}}(Y_{i,\ell})$$

for  $\ell = 1, 2$ , where  $y_{r,h,\ell}^*$  are points in  $A_{r,h}$ . It follows that

$$\begin{aligned} \mathbb{P}[Y_{r,i}^{(1)} \in B_1, Y_{r,j}^{(2)} \in B_2 \mid (\tilde{\mu}_1, \tilde{\mu}_2)] = \\ \sum_{h,g=1}^{\kappa_r+1} \frac{\tilde{\mu}_1(A_{r,h}) T(y_{r,h,1}^*; \tilde{\mu}_1)}{\int_{\mathbb{Y}} \tilde{\mu}_1(dy) T(y; \tilde{\mu}_1)} \frac{\tilde{\mu}_2(A_{r,g}) T(y_{r,g,2}^*; \tilde{\mu}_2)}{\int_{\mathbb{Y}} \tilde{\mu}_2(dy) T(y; \tilde{\mu}_2)} \delta_{y_{r,h,1}^*}(B_1) \delta_{y_{r,g,2}^*}(B_2) \end{aligned}$$

with  $T(y, \mu) = \int_0^\infty \exp(-\int_{\mathbb{Y}} \int_0^x k(s; y') ds \mu(dy')) k(x, y) dx$  for any  $y$  in  $\mathbb{Y}$  and  $\mu$  in the space  $M_{\mathbb{Y}}$  of boundedly finite measures on  $\mathbb{Y}$ . If we introduce vectors  $\mathbf{Y}_{r,n_1}^{(1)} = (Y_{r,1}^{(1)}, \dots, Y_{r,n_1}^{(1)})$  and  $\mathbf{Y}_{r,n_2}^{(2)} = (Y_{r,1}^{(2)}, \dots, Y_{r,n_2}^{(2)})$ , then, Proposition 2 in Regazzini and Sazonov (2000) implies

$$\begin{aligned} \mathcal{T}_L^{(r)}(t_1, t_2; \lambda_1, \lambda_2) &:= \mathbb{E}[\exp(-\lambda_1 \tilde{h}_1(t_1) - \lambda_2 \tilde{h}_2(t_2)) \mid \mathbf{X}, \mathbf{V}, \mathbf{Y}_{r,n_1}^{(1)}, \mathbf{Y}_{r,n_2}^{(2)}] \\ &\longrightarrow \mathbb{E}[\exp(-\lambda_1 \tilde{h}_1(t_1) - \lambda_2 \tilde{h}_2(t_2)) \mid \mathbf{X}, \mathbf{V}, \mathbf{Y}] =: \mathcal{T}_L(t_1, t_2; \lambda_1, \lambda_2) \quad (\text{C.1}) \end{aligned}$$

almost surely, as  $r \rightarrow \infty$ , with  $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$ ,  $\mathbf{V} = (\mathbf{V}_1, \mathbf{V}_2)$  and  $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2)$ . Hence, relying on this result, our main task is the determination of  $\mathcal{T}_L^{(r)}(t_1, t_2; \lambda_1, \lambda_2)$ , which is feasible, and of its limit  $\mathcal{T}_L(t_1, t_2; \lambda_1, \lambda_2)$  as  $r \rightarrow \infty$ . In (C.1) we shall also suppose one is also conditioning to  $c, z$  and possibly other hyperparameters involved in the elicitation of the model. In view of the properties of the sequence of nested partitions  $(\Pi_r)_{r \geq 1}$ , for  $r$  large enough the distinct values  $Y_{i,\ell}^*$  and  $Y_m^*$ , for  $i = 1, \dots, k_\ell$  and  $m = 1, \dots, k_0$ , belong to disjoint partition sets  $A_{r,h}$ . Hence, we denote as  $A_{r,i}^{(\ell)}$  and  $A_{r,m}$  the sets containing  $Y_{i,\ell}^*$  and  $Y_m^*$ , respectively. Moreover,  $\mathbb{Y}_r^*$  will denote  $\mathbb{Y}$  after excluding all those  $A_{r,h}$  sets including the distinct  $Y_{i,\ell}^*$  and  $Y_m^*$ . In order to evaluate  $\mathcal{T}_L^{(r)}(t_1, t_2; \lambda_1, \lambda_2)$ , define  $C_{i,\ell} := \{j : Y_{j,\ell} = Y_{i,\ell}^*\}$ ,  $C'_{m,\ell} := \{j : Y_{j,\ell} = Y_m^*\}$  and

$$Q(\mathbf{x}, \mathbf{y}) = \prod_{\ell=1}^2 \left\{ \prod_{i=1}^{k_\ell} \prod_{j \in C_{i,\ell}} k(x_{i,\ell}; y_{j,\ell}^*) \right\} \left\{ \prod_{m=1}^{k_0} \prod_{j \in C'_{m,\ell}} k(x_{i,\ell}; y_m^*) \right\}$$

It can, then, be easily seen that

$$\mathcal{T}_L^{(r)}(t_1, t_2; \lambda_1, \lambda_2) = \frac{\mathbb{E} \left[ \exp(-\lambda_1 \tilde{h}_1(t_1) - \lambda_2 \tilde{h}_2(t_2)) \mathcal{L}_r(\tilde{\mu}_1, \tilde{\mu}_2, \mathbf{x}, \mathbf{y}) \right]}{\mathbb{E}[\mathcal{L}_r(\tilde{\mu}_1, \tilde{\mu}_2, \mathbf{x}, \mathbf{y})]} \quad (\text{C.2})$$

where

$$\begin{aligned} \mathcal{L}_r(\tilde{\mu}_1, \tilde{\mu}_2; \mathbf{x}, \mathbf{y}) &= \\ &= Q(\mathbf{x}, \mathbf{y}) \prod_{\ell=1}^2 \exp \left( - \int_{\mathbb{Y}_r^*} K_\ell(y) \tilde{\mu}_\ell(dy) \right) \prod_{m=1}^{k_0} \exp(-K_\ell(y_{r,m}^*) \tilde{\mu}_\ell(A_{r,m})) (\tilde{\mu}_\ell(A_{r,m}))^{q_{m,\ell}} \\ &\quad \times \prod_{i=1}^{k_\ell} \exp(-K_1(y_{r,i,\ell}^*) \tilde{\mu}_1(A_{r,i}^{(\ell)}) - K_2(y_{r,i,\ell}^*) \tilde{\mu}_2(A_{r,i}^{(\ell)})) (\tilde{\mu}_\ell(A_{r,i}^{(\ell)}))^{n_{i,\ell}} \end{aligned}$$

is the likelihood corresponding to the discretized observation procedure. Since for any  $A \cap B = \emptyset$  the two vectors  $(\tilde{\mu}_1(A), \tilde{\mu}_2(A))$  and  $(\tilde{\mu}_1(B), \tilde{\mu}_2(B))$  are independent, conditional on the hyperparameters, one has for  $r$  large enough

$$\begin{aligned} \mathbb{E}[\mathcal{L}_r(\tilde{\mu}_1, \tilde{\mu}_2, \mathbf{x}, \mathbf{y})] &= Q(\mathbf{x}, \mathbf{y}) \mathbb{E} \left[ \exp \left( - \sum_{i=1}^2 \int_{\mathbb{Y}_r^*} K_i(y) \tilde{\mu}_i(dy) \right) \right] \\ &\quad \times \prod_{\ell=1}^2 \prod_{i=1}^{k_\ell} \mathbb{E} \left[ \exp \left( - \sum_{j=1}^2 K_j(y_{r,i,\ell}^*) \tilde{\mu}_j(A_{r,i}^{(\ell)}) \right) (\tilde{\mu}_\ell(A_{r,i}^{(\ell)}))^{n_{i,\ell}} \right] \\ &\quad \times \prod_{m=1}^{k_0} \mathbb{E} \left[ \prod_{i=1}^2 \exp(-K_i(y_m^*) \tilde{\mu}_i(A_{r,m})) (\tilde{\mu}_i(A_{r,m}))^{q_{m,i}} \right]. \end{aligned} \quad (\text{C.3})$$

As for the the terms in the right hand side of (C.3), note first that

$$\mathbb{E} \left[ \exp \left( - \sum_{i=1}^2 \int_{\mathbb{Y}_r^*} K_i(y) \tilde{\mu}_i(dy) \right) \right] = \exp(-c \psi_z(K_1 \mathbb{1}_{\mathbb{Y}_r^*}, K_2 \mathbb{1}_{\mathbb{Y}_r^*})), \quad (\text{C.4})$$

Moreover, for any  $i = 1, \dots, k_1$ ,

$$\begin{aligned} \mathbb{E} \left[ \exp \left( - \sum_{j=1}^2 K_j(y_{r,i,1}^*) \tilde{\mu}_j(A_{r,i}^{(1)}) \right) (\tilde{\mu}_\ell(A_{r,i}^{(1)}))^{n_{i,1}} \right] \\ = (-1)^{n_{i,1}} \frac{\partial^{n_{i,1}}}{\partial \gamma_\ell^{n_{i,1}}} \exp \left( -c \psi_z \left( \gamma_1 \mathbb{1}_{A_{r,i}^{(1)}}, K_2(y_{r,i,1}^*) \mathbb{1}_{A_{r,i}^{(1)}} \right) \right) \Big|_{\gamma_1 = K_1(y_{r,i,1}^*)}, \end{aligned} \quad (\text{C.5})$$



with a similar expression holding true for  $\ell = 2$ , and

$$\begin{aligned} & \mathbb{E} \left[ \exp \left( - \sum_{j=1}^2 K_j(y_m^*) \tilde{\mu}_j(A_{r,m}) \right) (\tilde{\mu}_1(A_{r,m}))^{q_{m,1}} (\tilde{\mu}_2(A_{r,m}))^{q_{m,2}} \right] \\ &= (-1)^{q_{m,1}+q_{m,2}} \frac{\partial^{q_{m,1}+q_{m,2}}}{\partial \gamma_1^{q_{m,1}} \partial \gamma_2^{q_{m,2}}} \exp \left( -c \psi_z \left( \gamma_1 \mathbb{1}_{A_{r,m}}, \gamma_2 \mathbb{1}_{A_{r,m}} \right) \right) \Big|_{\gamma_i = K_i(y_m^*)}, \end{aligned} \quad (\text{C.6})$$

with  $i = 1, 2$  and for each  $r \in \{1, \dots, k_0\}$ . The derivatives in (C.5)–(C.6) can be evaluated by resorting to an extension of the Faà di Bruno’s formula as proposed, e.g., in Constantine and Savits (1996). Such a formula is quite difficult to evaluate, but in our case it considerably simplifies due to diffuseness of  $P_0$ . Indeed, as  $r \rightarrow \infty$  one has  $P_0(A_{r,h}) \rightarrow 0$ , for any  $h$ , and one can rely on the first term of the formula in Constantine and Savits (1996), since the remaining terms converge to zero, as  $r \rightarrow \infty$ , more quickly than the first one and do not contribute to the evaluation of (C.2). In view of this helpful remark, one may write for any  $i = 1, \dots, k_1$

$$\begin{aligned} & \mathbb{E} \left[ \exp \left( - \sum_{j=1}^2 K_j(y_{r,i,1}^*) \tilde{\mu}_j(A_{r,i}^{(1)}) \right) (\tilde{\mu}_1(A_{r,i}^{(1)}))^{n_{i,1}} \right] \\ &= c P_0(A_{r,i}^{(1)}) \exp \left( -c \psi_z \left( K_1(y_{r,i,1}^*) \mathbb{1}_{A_{r,i}^{(1)}}, K_2(y_{r,i,1}^*) \mathbb{1}_{A_{r,i}^{(1)}} \right) \right) \\ & \quad \times \left[ (1-z) \int_0^{+\infty} s^{n_{i,1}} \exp \left( - (K_1(y_{r,i,1}^*) + K_2(y_{r,i,1}^*)) s \right) \rho(s) ds \right. \\ & \quad \left. + z \int_0^{+\infty} s^{n_{i,1}} \exp \left( -K_1(y_{r,i,1}^*) s \right) \rho(s) ds \right] + o \left( P_0(A_{r,i}^{(1)}) \right), \end{aligned} \quad (\text{C.7})$$

where, as usual, for any sequence  $(b_r)_{r \geq 1}$  such that  $b_r \rightarrow 0$  as  $r \rightarrow \infty$ , then  $b'_r = o(b_r)$  defines a sequence such that  $b'_r/b_r \rightarrow 0$  as  $r \rightarrow \infty$ . A similar expression can be deduced for  $\ell = 2$ . Moreover,

$$\begin{aligned} & \mathbb{E} \left[ \exp \left( - \sum_{j=1}^2 K_j(y_m^*) \tilde{\mu}_j(A_{r,m}) \right) (\tilde{\mu}_1(A_{r,m}))^{q_{m,1}} (\tilde{\mu}_2(A_{r,m}))^{q_{m,2}} \right] \\ &= c P_0(dy_m^*) \exp \left( -c \psi_z \left( K_1(y_m^*) \mathbb{1}_{A_{r,m}}, K_2(y_m^*) \mathbb{1}_{A_{r,m}} \right) \right) \\ & \quad \times (1-z) \int_0^{+\infty} s^{q_{m,1}+q_{m,2}} \exp \left( - (K_1(y_m^*) + K_2(y_m^*)) s \right) \rho(s) ds + o(P_0(A_{r,m})). \end{aligned} \quad (\text{C.8})$$

It is now useful to introduce the auxiliary variables  $(\mathbf{V}_1, \mathbf{V}_2)$  which allow, starting from (C.7), to deduce the following

$$\begin{aligned} & \mathbb{E} \left[ \exp \left( - \sum_{j=1}^2 K_j(y_{r,i,1}^*) \tilde{\mu}_j(A_{r,i}^{(1)}) \right) (\tilde{\mu}_1(A_{r,i}^{(1)}))^{n_{i,1}} \middle| \mathbf{V}_1, \mathbf{V}_2 \right] \\ &= c P_0(A_{r,i}^{(1)}) \exp \left( -c \psi_z \left( K_1(y_{r,i,1}^*) \mathbb{1}_{A_{r,i}^{(1)}}, K_2(y_{r,i,1}^*) \mathbb{1}_{A_{r,i}^{(1)}} \right) \right) \\ & \quad \times \int_0^{+\infty} s^{n_{i,1}} \exp \left( - (K_1(y_{r,i,1}^*) + K_2(y_{r,i,1}^*) V_{i,1}) s \right) \rho(s) ds + o \left( P_0(A_{r,i}^{(1)}) \right) \end{aligned} \quad (\text{C.9})$$

and, analogously,

$$\begin{aligned} & \mathbb{E} \left[ \exp \left( - \sum_{i=1}^2 K_i(y_{r,j,2}^*) \tilde{\mu}_i(A_{r,j}^{(2)}) \right) (\tilde{\mu}_2(A_{r,j}^{(2)}))^{n_{j,2}} \middle| \mathbf{V}_1, \mathbf{V}_2 \right] \\ &= c P_0(A_{r,j}^{(2)}) \exp \left( -c \psi_z \left( K_1(y_{r,j,2}^*) \mathbb{1}_{A_{r,j}^{(2)}}, K_2(y_{r,j,2}^*) \mathbb{1}_{A_{r,j}^{(2)}} \right) \right) \\ & \quad \times \int_0^{+\infty} s^{n_{j,2}} \exp \left( - (K_1(y_{r,j,2}^*) V_{j,2} + K_2(y_{r,j,2}^*)) s \right) \rho(s) ds + o \left( P_0(A_{r,j}^{(2)}) \right). \end{aligned} \quad (\text{C.10})$$

Using (C.4), (C.8), (C.9) and (C.10), and defining

$$B_r := \prod_{i=1}^{k_1} P_0(A_{r,i}^{(1)}) \prod_{j=1}^{k_2} P_0(A_{r,j}^{(2)}) \prod_{m=1}^{k_0} P_0(A_{r,m})$$

we can give a conditioned expression of the expected value in (C.3), that is

$$\begin{aligned} \mathbb{E} [\mathcal{L}_r(\tilde{\mu}_1, \tilde{\mu}_2, \mathbf{x}, \mathbf{y}) \mid \mathbf{V}_1, \mathbf{V}_2] &= Q(\mathbf{x}, \mathbf{y}) B_r e^{k_1+k_2+k_0} (1-z)^{k_0} \exp(-c \psi_z(K_1, K_2)) \\ & \quad \times \prod_{i=1}^{k_1} \int_0^{+\infty} s^{n_{i,1}} \exp \left( - (K_1(y_{r,i,1}^*) + K_2(y_{r,i,1}^*) V_{i,1}) s \right) \rho(s) ds \\ & \quad \times \prod_{j=1}^{k_2} \int_0^{+\infty} s^{n_{j,2}} \exp \left( - (K_1(y_{r,j,2}^*) V_{j,2} + K_2(y_{r,j,2}^*)) s \right) \rho(s) ds \\ & \quad \times \prod_{m=1}^{k_0} \int_0^{+\infty} s^{q_{m,1}+q_{m,2}} \exp \left( - (K_1(y_m^*) + K_2(y_m^*)) s \right) \rho(s) ds + o(B_r) \end{aligned} \quad (\text{C.11})$$

Thus, we got an expression for the denominator of the ratio (C.2) and, as  $r \rightarrow \infty$ , we can ignore the last summand on right hand side above since it will be negligible if compared to the other. The

determination of the numerator of (C.2) proceeds along the same lines, the only difference being the consideration of  $\tilde{h}_1(t_1)$  and  $\tilde{h}_2(t_2)$ . To this end, we set  $L_\ell(\cdot) = \lambda_\ell k(t_\ell; \cdot)$  for  $\ell = 1, 2$ , and using the independence increments property that has been used for determining the denominator, one has

$$\begin{aligned}
& \mathbb{E} \left[ \exp \left( -\lambda_1 \tilde{h}_1(t_1) - \lambda_2 \tilde{h}_2(t_2) \right) \mathcal{L}_r(\tilde{\mu}_1, \tilde{\mu}_2, \mathbf{x}, \mathbf{y}) \mid \mathbf{V}_1, \mathbf{V}_2 \right] \\
&= Q(\mathbf{x}, \mathbf{y}) B_r c^{k_1+k_2+k_0} (1-z)^{k_0} \exp(-c\psi_z(L_1+K_1, L_2+K_2)) \\
&\times \prod_{i=1}^{k_1} \int_0^{+\infty} s^{n_{i,1}} \exp \left( - (L_1(y_{r,i,1}^*) + K_1(y_{r,i,1}^*) + [L_2(y_{r,i,1}^*) + K_2(y_{r,i,1}^*)] V_{i,1}) s \right) \rho(s) ds \\
&\times \prod_{j=1}^{k_2} \int_0^{+\infty} s^{n_{j,2}} \exp \left( - ([L_1(y_{r,j,2}^*) + K_1(y_{r,j,2}^*)] V_{j,2} + L_2(y_{r,j,2}^*) + K_2(y_{r,j,2}^*)) s \right) \rho(s) ds \\
&\times \prod_{m=1}^k \int_0^{+\infty} s^{q_{m,1}+q_{m,2}} \exp \left( - (L_1(y_m^*) + K_1(y_m^*) + L_2(y_m^*)K_2(y_m^*)) s \right) \rho(s) ds + o(B_r). \quad (\text{C.12})
\end{aligned}$$

One can now deduce that, as  $r \rightarrow \infty$ , the ratio that defines the discretized Laplace transform in (C.2) converges to

$$\begin{aligned}
& \frac{\exp(-c\psi_z(L_1+K_1, L_2+K_2))}{\exp(-c\psi_z(K_1, K_2))} \\
&\times \prod_{i=1}^{k_1} \frac{\int_0^{+\infty} s^{n_{i,1}} \exp \left( - (L_1(y_{r,i,1}^*) + K_1(y_{r,i,1}^*) + [L_2(y_{r,i,1}^*) + K_2(y_{r,i,1}^*)] V_{i,1}) s \right) \rho(t) ds}{\int_0^{+\infty} s^{n_{i,1}} \exp \left( - (K_1(y_{r,i,1}^*) + K_2(y_{r,i,1}^*) V_{i,1}) s \right) \rho(s) ds} \\
&\times \prod_{j=1}^{k_2} \frac{\int_0^{+\infty} s^{n_{j,2}} \exp \left( - ([L_1(y_{r,j,2}^*) + K_1(y_{r,j,2}^*)] V_{j,2} + L_2(y_{r,j,2}^*) + K_2(y_{r,j,2}^*)) s \right) \rho(t) ds}{\int_0^{+\infty} s^{n_{j,2}} \exp \left( - (K_1(y_{r,j,2}^*) V_{j,2} + K_2(y_{r,j,2}^*)) s \right) \rho(s) ds} \\
&\times \prod_{m=1}^k \frac{\int_0^{+\infty} s^{q_{m,1}+q_{m,2}} \exp \left( - (L_1(y_m^*) + K_1(y_m^*) + L_2(y_m^*)K_2(y_m^*)) s \right) \rho(s) ds}{\int_0^{+\infty} s^{q_{m,1}+q_{m,2}} \exp \left( - (K_1(y_m^*) + K_2(y_m^*)) s \right) \rho(s) ds}. \quad (\text{C.13})
\end{aligned}$$

If  $\psi$  is the Laplace exponent of a CRM with Lévy intensity  $\nu(ds, dy) = c\rho(s)ds P_0(dy)$ , then  $\psi_\tau := \psi(\cdot + \tau) - \psi(\tau)$  is, for any non-negative and measurable function  $\tau$ , the Laplace exponent of a CRM with Lévy intensity  $c \exp(-\tau(y)s) \rho(s) ds P_0(dy)$ . One thus realizes that

$$\begin{aligned}
& \frac{\exp(-c\psi_z(L_1+K_1, L_2+K_2))}{\exp(-c\psi_z(K_1, K_2))} = \\
&= \frac{\exp(-c(1-z)\psi(L_1+K_1+L_2+K_2) - cz(\psi(L_1+K_1) + \psi(L_2+K_2)))}{\exp(-c(1-z)\psi(K_1+K_2) - cz(\psi(K_1) + \psi(K_2)))}
\end{aligned}$$

$$= \exp(-c(1-z)\psi_{K_1+K_2}(L_1+L_2) - cz(\psi_{K_1}(L_1) + \psi_{K_2}(L_2)))$$

is the Laplace transform of  $(\tilde{h}_1^*, \tilde{h}_2^*)$ . We focus now on the first of the three products in (C.13). In every factor we consider the integral that does not depend on  $L_1$  and  $L_2$ , that is on  $\lambda_1$  and  $\lambda_2$ , as a normalizing constant and recognize the Laplace transform of a random vector  $(\sum_{i=1}^{k_1} J_1^{(i,1)}, \sum_{i=1}^{k_1} J_2^{(i,1)})$ , where the random jumps

$$(J_1^{(i,1)}, J_2^{(i,1)}) \stackrel{d}{=} (1, V_{i,1}) k(\cdot; Y_{i,1}^*) J_{i,1}$$

and  $J_{i,1}$  has density function given by  $f(\cdot | n_{i,1}, K_1(Y_{i,1}^*) + K_2(Y_{i,1}^* V_{i,1}))$ . Similarly, the second product in (C.13) coincides with the Laplace transform of  $(\sum_{j=1}^{k_2} J_1^{(j,2)}, \sum_{j=1}^{k_2} J_2^{(j,2)})$ , where the random jumps

$$(J_1^{(j,2)}, J_2^{(j,2)}) \stackrel{d}{=} (V_{j,2}, 1) k(\cdot; Y_{j,2}^*) J_{j,2}$$

and  $J_{j,2}$  has density function given by  $f(\cdot | n_{j,2}, K_1(Y_{j,2}^*) V_{j,2} + K_2(Y_{j,2}^*))$ . And finally, the third product in (C.13) corresponds to the Laplace transform of a random vector  $(\sum_{m=1}^k J_1^{(m)}, \sum_{m=1}^k J_2^{(m)})$ , where the random jumps

$$(J_1^{(m)}, J_2^{(m)}) \stackrel{d}{=} \mathbf{1}_2 k(\cdot; Y_m^*) J_m,$$

$\mathbf{1}_2 = (1, 1)$  and  $J_m$  has density function given by  $f(\cdot | q_{m,1} + q_{m,2}, K_1(Y_m^*) + K_2(Y_m^*))$ . This concludes the proof.  $\square$

## D Proof of Proposition 3

Let us use the simplified notation  $\mathbb{E}^*[\cdot]$  to denote an expectation conditional on the vector of observations and latent variables  $(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}_1, \mathbf{Y}_2, \mathbf{V}_1, \mathbf{V}_2)$  and on other hyperparameters of the model. Under square loss function, the estimator of the survival function  $S_1$  evaluated at a point  $t > 0$  is equal to

$$\begin{aligned} \mathbb{E}^* [\tilde{S}_1(t)] &= \mathbb{E}^* \left[ \exp \left( - \int_0^t \tilde{h}_1(s) ds \right) \right] \\ &= \mathbb{E}^* \left[ \exp \left( - \int_0^t \tilde{h}_1^*(s) ds - \sum_{i=1}^{k_1} \int_0^t k(s; Y_{i,1}^*) J_{i,1} ds \right. \right. \\ &\quad \left. \left. - \sum_{j=1}^{k_2} \int_0^t k(s; Y_{j,2}^*) V_{j,2} J_{j,2} ds - \sum_{m=1}^k \int_0^t k(s; Y_m^*) J_m ds \right) \right] \end{aligned}$$

$$\begin{aligned}
&= \mathbb{E}^* \left[ \exp \left( - \int_0^t \tilde{h}_1^*(s) ds - \sum_{i=1}^{k_1} \bar{K}_t(Y_{i,1}^*) J_{i,1} \right. \right. \\
&\quad \left. \left. - \sum_{j=1}^{k_2} \bar{K}_t(Y_{j,2}^*) V_{j,2} J_{j,2} - \sum_{m=1}^k \bar{K}_t(Y_m^*) J_m \right) \right].
\end{aligned}$$

By virtue of independence this may be written as

$$\begin{aligned}
\mathbb{E} \left[ \exp \left( - \int_0^t \tilde{h}_1^*(s) ds \right) \right] &\prod_{i=1}^{k_1} \mathbb{E} \left[ \exp \left( - \bar{K}_t(Y_{i,1}^*) J_{i,1} \right) \right] \\
&\times \prod_{j=1}^{k_2} \mathbb{E} \left[ \exp \left( - \bar{K}_t(Y_{j,2}^*) V_{j,2} J_{j,2} \right) \right] \prod_{m=1}^k \mathbb{E} \left[ \exp \left( - \bar{K}_t(Y_m^*) J_m \right) \right]. \quad (\text{D.1})
\end{aligned}$$

We shall now consider separately the expected values that appear in (D.1). We have

$$\begin{aligned}
\mathbb{E}^* \left[ \exp \left( - \int_0^t \tilde{h}_1^*(s) ds \right) \right] &= \mathbb{E}^* \left[ \exp \left( - \int_0^t \int_{\mathbb{Y}} k(s; y) \tilde{\mu}_1^*(dy) ds \right) \right] \\
&= \mathbb{E}^* \left[ \exp \left( - \int_{\mathbb{Y}} \int_0^t k(s; y) ds \tilde{\mu}_1^*(dy) \right) \right] \\
&= \mathbb{E}^* \left[ \exp \left( - \int_{\mathbb{Y}} \bar{K}_t(y) \tilde{\mu}_1^*(dy) \right) \right],
\end{aligned}$$

where we have used the Fubini-Tonelli theorem. As stated in Proposition 3,  $\tilde{\mu}_1^*$  is equal in distribution to the sum of  $\mu_1^*$  and  $\mu_0^*$ , independent CRMs. Thus, the last expression may be written as

$$\begin{aligned}
&\mathbb{E}^* \left[ \exp \left( - \int_{\mathbb{Y}} \bar{K}_t(y) \mu_1^*(dy) \right) \right] \mathbb{E}^* \left[ \exp \left( - \int_{\mathbb{Y}} \bar{K}_t(y) \mu_0^*(dy) \right) \right] \\
&= \exp \left( - \int_{\mathbb{R}^+ \times \mathbb{Y}} (1 - \exp(-s \bar{K}_t(y))) \nu_1^*(ds, dy) - \int_{\mathbb{R}^+ \times \mathbb{Y}} (1 - \exp(-s \bar{K}_t(y))) \nu_0^*(ds, dy) \right),
\end{aligned}$$

where  $\nu_0^*$  and  $\nu_1^*$  are the Lévy intensities in Proposition 3. Thus we may conclude that

$$\begin{aligned}
\mathbb{E} \left[ \exp \left( - \int_0^t \tilde{h}_1^*(s) ds \right) \right] &= \\
&= \exp \left( -c z \int_{\mathbb{R}^+ \times \mathbb{Y}} (1 - \exp(-s \bar{K}_t(y))) \exp(-s K_1(y)) \rho(s) ds P_0(dy) \right) \quad (\text{D.2}) \\
&\times \exp \left( -c(1-z) \int_{\mathbb{R}^+ \times \mathbb{Y}} (1 - \exp(-s \bar{K}_t(y))) \exp(-s (K_1(y) + K_2(y))) \rho(s) ds P_0(dy) \right).
\end{aligned}$$

We consider now  $\mathbb{E}^* [\exp(-\bar{K}_t(Y_{i,1}^*)J_{i,1})]$  and remind that the jumps  $J_{i,1}$  have conditional density function  $f(\cdot | n_{i,1}, K_1(Y_{i,1}^*) + V_{i,1} K_2(Y_{i,1}^*))$ . Hence

$$\mathbb{E}^* [\exp(-\bar{K}_t(Y_{i,1}^*)J_{i,1})] = \frac{\int_0^\infty \exp(-(K_1(Y_{i,1}^*) + K_2(Y_{i,1}^*)V_{i,1} + \bar{K}_t(Y_{i,1}^*))s) s^{n_{i,1}} \rho(s) ds}{\int_0^\infty \exp(-(K_1(Y_{i,1}^*) + K_2(Y_{i,1}^*)V_{i,1})s) s^{n_{i,1}} \rho(s) ds} \quad (\text{D.3})$$

One similarly finds out that

$$\begin{aligned} \mathbb{E}^* [\exp(-\bar{K}_t(Y_{j,2}^*)V_{j,2} J_{j,2})] \\ = \frac{\int_0^\infty \exp(-(K_1(Y_{j,2}^*)V_{j,2} + K_2(Y_{j,2}^*) + \bar{K}_t(Y_{j,2}^*)V_{j,2})s) s^{n_{j,2}} \rho(s) ds}{\int_0^\infty \exp(-(K_1(Y_{j,2}^*)V_{j,2} + K_2(Y_{j,2}^*))s) s^{n_{j,2}} \rho(s) ds} \end{aligned} \quad (\text{D.4})$$

and  $\mathbb{E} [\exp(-\bar{K}_t(Y_m^*)J_m)]$  is equal to

$$\frac{1}{C_m} \int_0^\infty \exp(-(K_1(Y_m^*) + K_2(Y_m^*) + \bar{K}_t(Y_m^*))s) s^{q_{m,1} + q_{m,2}} \rho(s) ds. \quad (\text{D.5})$$

Plugging expressions (D.2), (D.3), (D.4) and (D.5) into (D.1) completes the proof.  $\square$

## E Posterior variance of $\tilde{S}_\ell(t)$

A direct application of Proposition 2 allows to derive an explicit analytic form of the posterior variance of  $\tilde{S}_\ell(t)$ , given the latent variables  $\mathbf{L} = (\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{V}_1, \mathbf{V}_2)$ . Indeed, the evaluation of  $\mathbb{E} \tilde{S}_\ell^2(t) - (\mathbb{E} \tilde{S}_\ell(t))^2$  amounts to the computation of the marginal Laplace transform of the bivariate cumulative hazard induced by (20). Though this can be carried out for any vector of GM-dependent CRMs  $(\tilde{\mu}_1, \tilde{\mu}_2)$  and for any choice of the mixed kernel  $k(\cdot; \cdot)$ , here we focus on the dependent extended gamma process. More specifically, we set  $(\tilde{\mu}_1, \tilde{\mu}_2)$  as a GM-dependent gamma process and  $k(t, y) = \beta \mathbf{1}_{(0,t]}(y)$  for which we obtain a result analogous to Corollary 1. That is, at every point  $t > 0$ , the variance of the random survival function  $\tilde{S}_1(t)$ , conditional on  $\mathbf{X}_1, \mathbf{X}_2$ , latent variables  $\mathbf{L}$  and hyperparameters  $\Theta$ , equals

$$\begin{aligned} & \left( \exp \left\{ -cz \sum_{i=1}^{n_1} G_{i,1}^{(2)}(t) - c(1-z) \sum_{i=1}^{n_1+n_2} \tilde{G}_j^{(2)}(t) \right\} \right. \\ & \left. - \exp \left\{ -2cz \sum_{i=1}^{n_1} G_{i,1}(t) - 2c(1-z) \sum_{i=1}^{n_1+n_2} \tilde{G}_j(t) \right\} \right) \prod_{i=1}^{k_1} \left( 1 + \frac{\beta(t - Y_{i,1}^*) \mathbf{1}_{[Y_{i,1}^*, +\infty)}(t)}{1 + \beta \zeta_{i,1}} \right)^{-2n_{i,1}} \end{aligned}$$

$$\times \prod_{j=1}^{k_2} \left( 1 + V_{j,2} \frac{\beta(t - Y_{j,2}^*) \mathbb{1}_{[Y_{j,2}^*, +\infty)}(t)}{1 + \beta \zeta_{j,2}} \right)^{-2n_{j,2}} \prod_{m=1}^{k_0} \left( 1 + \frac{\beta(t - Y_m^*) \mathbb{1}_{[Y_m^*, +\infty)}(t)}{1 + \beta \tilde{\zeta}_m} \right)^{-2(q_{m,1} + q_{m,2})}, \quad (\text{E.1})$$

where  $G_{i,1}$  and  $\tilde{G}_j(t)$  are defined as in (25) and (26), whereas  $G_{i,\ell}^{(2)}$  and  $\tilde{G}_j^{(2)}(t)$  are defined as follows,

$$G_{i,\ell}^{(2)}(t) = \int_{X_{i,\ell} \wedge t}^{X_{i-1,\ell} \wedge t} \log \left( 1 + \frac{2\beta(t-y)}{1 + \beta \xi_{i-1,\ell} - \beta(i-1)y} \right) P_0(dy)$$

for  $\ell = 1, 2$  and  $i = 1, \dots, n_\ell + 1$  provided  $X_{0,\ell} \equiv +\infty$  and  $X_{n_\ell+1,\ell} \equiv 0$ , and

$$\tilde{G}_j^{(2)}(t) = \int_{\tilde{X}_j \wedge t}^{\tilde{X}_{j-1} \wedge t} \log \left( 1 + \frac{2\beta(t-y)}{1 + \beta \tilde{\xi}_{i-1} - \beta(i-1)y} \right) P_0(dy)$$

for  $j = 1, \dots, n_1 + n_2 + 1$ , where  $\tilde{X}_0 \equiv +\infty$  and  $\tilde{X}_{n_1+n_2+1} \equiv 0$ .  $\square$

Corollary 1 and (E.1) provide the tools to compute  $\text{Var}_\mu(\tilde{S}_1(t) \mid \mathbf{X}_1, \mathbf{X}_2)$ , by means of the MCMC output. More specifically, we evaluated the ratio in (29) over a grid of equally distanced values in the interval  $[0, \max\{\mathbf{X}_1, \mathbf{X}_2\}]$  and, on average, we obtained about 0.95 for three datasets.

## F The Gibbs sampler

In this section we describe a Gibbs type algorithm that we will use to estimate  $\tilde{S}_\ell$ , for  $\ell = 1, 2$ , when  $(\tilde{h}_1, \tilde{h}_2)$  are dependent extended gamma processes as described in the previous section. It is to be noted that, despite we specialize the sampler to the case where  $\tilde{\mu} = (\tilde{\mu}_1, \tilde{\mu}_2)$  is a vector of GM-dependent CRMs, the posterior characterization we provide in Section 4 of the paper allows one to extend the description below to any vector of GM-dependent CRMs and any kernel  $k(\cdot, \cdot)$ .

In order to simplify notation, set  $\mathbf{L} = (\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{V}_1, \mathbf{V}_2)$  and  $\Theta = (c, z, \beta)$ . The marginalization of  $\mathbb{E}[\tilde{S}_\ell \mid \mathbf{X}_1, \mathbf{X}_2, \mathbf{L}, \Theta]$ , for  $\ell = 1, 2$ , with respect to the hyperparameters  $\Theta$  and the latent variables  $\mathbf{L}$  can be achieved through the following sampler

0. Start with admissible initial values  $(\mathbf{L}^{(0)}, \Theta^{(0)})$ ,
1. At iteration  $j \geq 1$ , update  $(\mathbf{Y}_1^{(j-1)}, \mathbf{Y}_2^{(j-1)})$  by using its full conditional,
2. At iteration  $j \geq 1$ , update  $(\mathbf{V}_1^{(j-1)}, \mathbf{V}_2^{(j-1)})$  by using its full conditional,

3. At iteration  $j \geq 1$ , update  $c^{(j-1)}$ ,  $z^{(j-1)}$  and  $\beta^{(j-1)}$  by using their full conditional.

After an adequate number  $j_0$  of burn-in iterations, we consider every realization of latent variables and hyperparameters as a sample from the posterior distribution of  $(\mathbf{L}, \Theta)$  and use them in order to determine an approximate evaluation of the expected value of (27). Therefore, we add one step to the algorithm.

4. At iteration  $j \geq j_0$ , plug  $(\mathbf{L}^{(j)}, \Theta^{(j)})$  into (27) and obtain a realization of

$$\mathbb{E}[\tilde{S}_1 | \mathbf{X}_1, \mathbf{X}_2, \mathbf{L}, \Theta] \text{ that we denote by } \mathbb{E}_{\mathbf{X}}[\tilde{S}_1]^{(j)} \text{ (analogously for } \mathbb{E}_{\mathbf{X}}[\tilde{S}_2]^{(j)}).$$

Ergodic means of  $\mathbb{E}_{\mathbf{X}}[\tilde{S}_1]^{(j)}$  and  $\mathbb{E}_{\mathbf{X}}[\tilde{S}_2]^{(j)}$ , then, provide an approximate evaluation of posterior estimates of  $\tilde{S}_1$  and  $\tilde{S}_2$ , namely  $\mathbb{E}[\tilde{S}_1 | \mathbf{X}_1, \mathbf{X}_2]$  and  $\mathbb{E}[\tilde{S}_2 | \mathbf{X}_1, \mathbf{X}_2]$ .

We only need to provide the details on the full conditionals we have used for implementing the steps above and they can be found in the next two Subsections.

## F.1 Full conditional distributions

Here we display the full conditionals of the components of  $\mathbf{Y}_1$  and  $\mathbf{V}_1$  and those of the hyperparameters entering the model specification. Such distributions can be obtained from the joint law of observations, latent variables and hyperparameters, that is easily derived from (C.12). We confine ourselves to considering the conditional distributions of latent variables that refer to the first group; deriving analogous expressions for the second group is straightforward. With  $\mathbf{Y}_1^{(-i)}$  we denote the vector obtained by deleting the  $i$ -th component of  $\mathbf{Y}_1$  and with  $\mathbf{V}_1^{(-i)}$  we indicate the vector of latent variables corresponding to the distinct values of  $\mathbf{Y}_1^{(-i)}$ . Similarly,  $k_0^{(-i)}$ ,  $k_1^{(-i)}$ ,  $k_2^{(-i)}$  denote the number of distinct values, shared or featured only by first and second group respectively, that are left once  $Y_{i,1}$  has been dismissed, whose frequencies are  $q_m^{(-i)} = q_{m,1}^{(-i)} + q_{m,2}^{(-i)}$ ,  $n_{j,1}^{(-i)}$  and  $n_{l,2}^{(-i)}$ , with  $m = 1, \dots, k_0^{(-i)}$ ,  $j = 1, \dots, k_1^{(-i)}$  and  $l = 1, \dots, k_2^{(-i)}$ . The full conditional of  $Y_{i,1}$  may be written as

$$\mathbb{P}[Y_{i,1} \in dy | \mathbf{Y}_1^{(-i)}, \dots] = w_0 G_0(dy) + \sum_{j=1}^{k_1^{(-i)}} w_{j,1} \delta_{Y_{j,1}^*}(dy) + \sum_{l=1}^{k_2^{(-i)}} w_{l,2} \delta_{Y_{l,2}^*}(dy) + \sum_{m=1}^{k_0^{(-i)}} w_m \delta_{Y_m^*}(dy), \quad (\text{F.1})$$

where

$$w_0 \propto \frac{c}{T} \left[ (1 - V_{i,1}) z \sum_{j=1}^i \frac{1}{j} \log \left( \frac{\xi_{j,1} - j(X_{j+1,1}) + \beta^{-1}}{\xi_{j,1} - j(X_{j,1}) + \beta^{-1}} \right) \right]$$



$$\begin{aligned}
& + V_{i,1} (1 - z) \sum_{j=1}^{n_1+n_2} \frac{1}{j} \log \left( \frac{\tilde{\xi}_j - j(\tilde{X}_{j+1} \wedge X_{i,1}) + \beta^{-1}}{\tilde{\xi}_j - j(\tilde{X}_j \wedge X_{i,1}) + \beta^{-1}} \right) \Big] \\
w_{j,1} & \propto \frac{n_{j,1}^{(-i)} \mathbb{1}_{\{y_{j,1}^* \leq x_{i,1}\}} \mathbb{1}_{\{V_{i,1} = V_{j,1}^*\}}}{\zeta_{j,1} + \beta^{-1}}, \\
w_{l,2} & \propto \frac{n_{l,2}^{(-i)} \mathbb{1}_{\{y_{l,2}^* \leq x_{i,1}\}} \mathbb{1}_{\{V_{i,1} = V_{l,2}^* = 1\}}}{\zeta_{l,2} + \beta^{-1}}, \\
w_m & \propto \frac{q_m^{(-i)} \mathbb{1}_{\{y_m^* \leq x_{i,1}\}} \mathbb{1}_{\{V_{i,1} = 1\}}}{\tilde{\zeta}_m + \beta^{-1}}
\end{aligned}$$

and

$$G_0(dy) \propto \left( \frac{(1 - V_{i,1}) \mathbb{1}_{\{y \leq x_{i,1}\}}}{\sum_{j=1}^{n_1} (x_{j,1} - y) \mathbb{1}_{\{y \leq x_{j,1}\}} + \beta^{-1}} + \frac{V_{i,1} \mathbb{1}_{\{y \leq x_{i,1}\}}}{\sum_{j=1}^{n_1+n_2} (\tilde{x}_j - y) \mathbb{1}_{\{y \leq \tilde{x}_j\}} + \beta^{-1}} \right) dy. \quad (\text{F.2})$$

As suggested by Ishwaran and James (2004) we propose a rejection algorithm to sample from  $G_0$ : the details are provided in Appendix F.2. In order to fasten up significantly the mixing of the algorithm we add a further step. The idea, suggested e.g. in Ishwaran and James (2004), consists in re-sampling, at the end of every iteration, the distinct values of the latent variables  $\mathbf{Y}_1$  and  $\mathbf{Y}_2$  from their full conditional distribution that, in the case of dependent extended gamma processes we are considering, equals

$$\mathcal{L}(Y_{i,1}^* | \mathbf{Y}_1 \setminus Y_{i,1}^*, \dots) \propto \mathbb{1}_{(0, \min_{j \in C_{i,1}}(X_{j,1})]}(Y_{i,1}) (\beta \zeta_{i,1} + 1)^{-n_{i,1}}, \quad (\text{F.3})$$

for  $i = 1, \dots, k_1$ . Similar expressions hold for  $Y_{j,2}^*$ , with  $j = 1, \dots, k_2$ , and  $Y_m^*$ , with  $m = 1, \dots, k_0$ . It is possible to sample from (F.3) by means of a rejection algorithm that essentially proceeds as the one we used to sample a new value of  $Y_{i,1}$ , for which we still refer to Appendix F.2.

As far as the label  $V_{i,1}$  is concerned, we have

$$\mathbb{P}[V_{i,1} = v | \mathbf{V}_1^{(-i)}, \dots] \propto \frac{(1 - z)^v z^{-v}}{\left( \zeta_{i,1} |_{V_{i,1}=v} + \beta^{-1} \right)^{n_{i,1}+1}}. \quad (\text{F.4})$$

The full conditional distribution of  $z$  coincides with

$$\mathcal{L}(z | \dots) \propto \mathcal{L}(z) (1 - z)^{k+\bar{k}_1+\bar{k}_2} z^{k_1+k_2-\bar{k}_1-\bar{k}_2} \exp\left(-\frac{zC}{T}(X_{1,1} \wedge X_{1,2})\right)$$

$$\begin{aligned}
& \times \prod_{\ell=1}^2 \prod_{i=1}^{n_\ell} \frac{(1 + \beta \xi_{i,\ell} - i\beta X_{i+1,\ell})^{-\frac{z_\ell c}{T} \left( \frac{\xi_{i,\ell} + \beta^{-1}}{i} - X_{i+1,\ell} \right)}}{(1 + \beta \xi_{i,\ell} - i\beta X_{i,\ell})^{-\frac{z_\ell c}{T} \left( \frac{\xi_{i,\ell} + \beta^{-1}}{i} - X_{i,\ell} \right)}} \\
& \times \prod_{i=1}^{n_1+n_2} \frac{(1 + \beta \tilde{\xi}_i - i\beta \tilde{X}_i)^{-\frac{z c}{T} \left( \frac{\tilde{\xi}_i + \beta^{-1}}{i} - \tilde{X}_i \right)}}{(1 + \beta \tilde{\xi}_i - i\beta \tilde{X}_{i+1})^{-\frac{z c}{T} \left( \frac{\tilde{\xi}_i + \beta^{-1}}{i} - \tilde{X}_{i+1} \right)}}, \quad (\text{F.5})
\end{aligned}$$

where  $\mathcal{L}(z)$  is the prior distribution of  $z$  and  $\tilde{k}_\ell = \sum_{i=1}^{k_\ell} V_{i,\ell}^*$ , for  $\ell = 1, 2$ . The full conditional distribution of  $c$  may be written as

$$\begin{aligned}
\mathcal{L}(c | \dots) & \propto \mathcal{L}(c) e^{k_1+k_2+k} \exp \left( -\frac{z c}{T} (X_{1,1} \wedge X_{1,2}) - \frac{c}{T} (X_{1,1} \vee X_{1,2}) \right) \\
& \times \prod_{\ell=1}^2 \prod_{i=1}^{n_\ell} \frac{(1 + \beta \xi_{i,\ell} - i\beta X_{i+1,\ell})^{-\frac{z_\ell c}{T} \left( \frac{\xi_{i,\ell} + \beta^{-1}}{i} - X_{i+1,\ell} \right)}}{(1 + \beta \xi_{i,\ell} - i\beta X_{i,\ell})^{-\frac{z_\ell c}{T} \left( \frac{\xi_{i,\ell} + \beta^{-1}}{i} - X_{i,\ell} \right)}} \\
& \times \prod_{i=1}^{n_1+n_2} \frac{(1 + \beta \tilde{\xi}_i - i\beta \tilde{X}_i)^{-\frac{(z-1)c}{T} \left( \frac{\tilde{\xi}_i + \beta^{-1}}{i} - \tilde{X}_i \right)}}{(1 + \beta \tilde{\xi}_i - i\beta \tilde{X}_{i+1})^{-\frac{(z-1)c}{T} \left( \frac{\tilde{\xi}_i + \beta^{-1}}{i} - \tilde{X}_{i+1} \right)}}, \quad (\text{F.6})
\end{aligned}$$

with  $\mathcal{L}(c)$  prior distribution of  $c$ . Finally, if we denote with  $\mathcal{L}(\beta)$  the prior distribution of  $\beta$ , we have

$$\begin{aligned}
\mathcal{L}(\beta | \dots) & \propto \mathcal{L}(\beta) \prod_{\ell=1}^2 \left[ \prod_{i=1}^{n_\ell} \frac{(1 + \beta \xi_{i,\ell} - i\beta X_{i+1,\ell})^{-\frac{z_\ell c}{T} \left( \frac{\xi_{i,\ell} + \beta^{-1}}{i} - X_{i+1,\ell} \right)}}{(1 + \beta \xi_{i,\ell} - i\beta X_{i,\ell})^{-\frac{z_\ell c}{T} \left( \frac{\xi_{i,\ell} + \beta^{-1}}{i} - X_{i,\ell} \right)}} \right. \\
& \times \left. \prod_{i=1}^{k_\ell} (\zeta_{i,\ell} + \beta^{-1})^{-n_{i,\ell}} \right] \\
& \times \prod_{i=1}^{n_1+n_2} \frac{(1 + \beta \tilde{\xi}_i - i\beta \tilde{X}_i)^{-\frac{(z-1)c}{T} \left( \frac{\tilde{\xi}_i + \beta^{-1}}{i} - \tilde{X}_i \right)}}{(1 + \beta \tilde{\xi}_i - i\beta \tilde{X}_{i+1})^{-\frac{(z-1)c}{T} \left( \frac{\tilde{\xi}_i + \beta^{-1}}{i} - \tilde{X}_{i+1} \right)}} \\
& \times \prod_{m=1}^k (\tilde{\zeta}_m + \beta^{-1})^{-(q_{m,1} + q_{m,2})}
\end{aligned}$$

## F.2 Sampling a new latent variable $Y_{i,1}$

Equation (F.2) provides, up to a proportionality constant, the full conditional distribution of a latent variable  $Y_{i,1}$  conditionally on the fact that  $Y_{i,1}$  does not coincide with any already observed value. Here we show how to sample a new value for  $Y_{i,1}$  conditionally on  $V_{i,1} = 0$ , the extension to the case

where  $V_{i,1} = 1$  being straightforward. According to (F.2),  $Y_{i,1}$  is, with probability 1, in the interval  $(0, X_{i,1}]$ . For  $j = i, \dots, n_1$ , we define

$$I_j = (X_{j+1,1}, X_{j,1}]$$

and obtain a partition  $\{I_i, \dots, I_{n_1}\}$  of  $(0, X_{i,1}]$ . Moreover, for each  $j = i, \dots, n_1$ , we define  $\pi_j$  as the probability that a new value for  $Y_{i,1}$  is in the interval  $I_j$  and we obtain

$$\pi_j = \frac{1}{j} \log \left( \frac{\xi_{j,1} - jX_{j+1,1} + \beta^{-1}}{\xi_{j,1} - jX_{j,1} + \beta^{-1}} \right).$$

Thus, if we define, for  $j = i, \dots, n_1$ ,

$$p_j = \frac{\pi_j}{\sum_{r=i}^{n_1} \pi_r},$$

we can rewrite

$$G_0(dy) = \sum_{j=i}^{n_1} p_j f_j(y) dy, \tag{F.7}$$

where

$$f_j(y) := \frac{1}{\pi_j} \frac{\mathbb{1}_{I_j}(y)}{\xi_{j,1} - jy + \beta^{-1}}.$$

A trivial rejection algorithm allows to sample from each  $f_j$ . Since, for every  $j = i, \dots, n_1$ , the density  $f_j$  vanishes outside the interval  $I_j$  and is bounded by

$$M_j := \frac{1}{\pi_j(\xi_{j,1} - jx_{j,1} + \beta^{-1})},$$

$M_j \mathbb{1}_{I_j}$  can be used as dominating curve. Thus, if we define a  $(n_1 - i + 1)$ -dimensional random vector  $\mathbf{Q}$  with multinomial distribution of parameters  $(1, p_i, \dots, p_{n_1})$ , we can sample a new value for  $Y_{i,1}$  by following the next steps.

1. Generate  $\mathbf{Q}$ ,
2. define  $j_0$  as the only element of  $\{i, \dots, n_1\}$  s.t.  $Q_{j_0} = 1$ ,
3. generate  $Y_{i,1}$  from  $f_{j_0}$  by means of the described rejection algorithm.

## References

- Aalen, O. O., Borgan, Ø., and Gjessing, H. K. (2008). *Survival and event history analysis*. Springer, New York.
- Antoniak, C. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Ann. Statist.*, 2:1152–1174.
- Choi, S. and Huang, X. (2012). A general class of semiparametric transformation frailty models for nonproportional hazards survival data. *Biometrics*, 68:1126–1135.
- Constantine, G. and Savits, T. (1996). A multivariate Faà di Bruno formula with applications. *Trans. Amer. Math. Soc.*, 348:503–520.
- Cox, D. (1972). Regression models and life tables (with discussion). *J. Roy. Statist. Soc. Ser. A*, 34:187–202.
- Daley, D. and Vere-Jones, D. (2003). *An introduction to the theory of point processes. Vol. I*. Springer-Verlag, New York.
- Damien, P. and Walker, S. (2002). Bayesian nonparametric nonproportional hazards survival modeling. *Scand. J. Statist.*, 29:51–56.
- De Blasi, P., Peccati, G., and Prünster, I. (2009). Asymptotics for posterior hazards. *Ann. Statist.*, 37(4):1906–1945.
- De Iorio, M., Johnson, W., Müller, P., and Rosner, G. (2009). Bayesian nonparametric nonproportional hazards survival modeling. *Biometrics*, 65:762–771.
- De Iorio, M., Müller, P., Rosner, G., and MacEachern, S. (2004). An ANOVA model for dependent random measures. *J. Amer. Statist. Assoc.*, 99:205–215.
- Doksum, K. (1974). Tailfree and neutral random probabilities and their posterior distributions. *Ann. Probab.*, 2:183–201.
- Dykstra, R. and Laud, P. (1981). A Bayesian nonparametric approach to reliability. *Ann. Statist.*, 9:356–367.
- Epifani, I. and Lijoi, A. (2010). Nonparametric priors for vectors of survival functions. *Statist. Sinica*, 20:1455–1484.

- Ferguson, T. (1974). Prior distributions on spaces of probability measures. *Ann. Statist.*, 2:615–629.
- Ferguson, T. and Phadia, E. (1979). Bayesian nonparametric estimation based on censored data. *Ann. Statist.*, 7:163–186.
- Gjessing, H. K., Aalen, O. O., and Hjort, N. L. (2003). Frailty models based on Lévy processes. *Adv. in Appl. Probab.*, 35(2):532–550.
- Griffin, J., Kolossatis, M., and Steel, M. (2013). Comparing distributions by using dependent normalized random-measure mixtures. *J. Roy. Statist. Soc. Ser. B*, 75:499–529.
- Griffin, J. and Steel, M. (2008). Order-based dependent Dirichlet processes. *J. Amer. Statist. Assoc.*, 101:179–194.
- Griffiths, R. and Milne, R. (1978). A class of bivariate Poisson processes. *J. Multivariate Anal.*, 8:380–395.
- Hanson, T., Jara, A., and Zhao, L. (2012). A Bayesian semiparametric temporally stratified proportional hazards model with spatial frailties. *Bayesian Anal.*, 7:147–188.
- Hougaard, P. (2000). *Analysis of Multivariate Survival Data*. Springer, New York.
- Ishwaran, H. and James, L. (2004). Computational methods for multiplicative intensity models using weighted gamma processes: proportional hazards, marked point processes, and panel count data. *J. Amer. Statist. Assoc.*, 99:175–190.
- James, L. (2003). Bayesian calculus for gamma processes with applications to semiparametric intensity models. *Sankhyā*, 65:179–206.
- James, L. (2005). Bayesian Poisson process partition calculus with an application to Bayesian Lévy moving averages. *Ann. Statist.*, 33:1771–1799.
- Jara, A. and Hanson, T. (2011). A class of mixtures of dependent tail-free processes. *Biometrika*, 98:553–566.
- Kalbfleish, J. and Prentice, R. (2002). *The statistical analysis of failure time data*. Wiley, New York.
- Kingman, J. (1967). Completely random measures. *Pacific J. Math.*, 21:59–78.

- Klein, J. and Moeschberger, M. (1997). *Survival Analysis. Techniques for Censored and Truncated Data*. Springer-Verlag, New York.
- Lawless, J. (1982). *Statistical models and methods for lifetime data*. Wiley, Chichester.
- Lee, E., Wei, L., and Ying, Z. (1993). Linear regression analysis for highly stratified failure time data. *J. Amer. Statist. Assoc.*, 88:557–565.
- Lijoi, A., Nipoti, B., and Prünster, I. (2013). Bayesian inference with dependent normalized completely random measures. *Bernoulli*. To appear.
- Lijoi, A. and Prünster, I. (2010). Models beyond the dirichlet process. In Hjort, N., Holmes, C., Müller, P., and Walker, S., editors, *Bayesian Nonparametrics*, pages 80–136. Cambridge University Press, Cambridge.
- Lo, A. and Weng, C. (1989). On a class of Bayesian nonparametric estimates. II. Hazard rate estimates. *Ann. Inst. Statist. Math.*, 41:227–245.
- MacEachern, S. (1999). Dependent nonparametric processes. In *ASA Proceedings of the Section on Bayesian Statistical Science*. American Statistical Association, Alexandria.
- MacEachern, S. (2000). Dependent Dirichlet processes. *Technical Report*.
- Peccati, G. and Prünster, I. (2008). Linear and quadratic functionals of random hazard rates: an asymptotic analysis. *Ann. Appl. Probab.*, 18(5):1910–1943.
- Regazzini, E. and Sazonov, V. (2000). Approximation of distributions of random probabilities by mixtures of dirichlet distributions with applications to nonparametric bayesian statistical inferences. *Theory Probab. Appl.*, 45:93–110.
- Rodríguez, A., Dunson, D., and Gelfand, A. (2008). The nested Dirichlet process. *J. Amer. Statist. Assoc.*, 103:1566–1581.
- Susarla, V. and Van Ryzin, J. (1976). Nonparametric Bayesian estimation of survival curves from incomplete observations. *J. Amer. Statist. Assoc.*, 71:897–902.
- Teh, Y., Jordan, M., Beal, M. J., and Blei, D. (2006). Hierarchical Dirichlet processes. *J. Amer. Statist. Assoc.*, 101:1566–1581.

- Trippa, L., Müller, P., and Johnson, W. (2011). The multivariate beta process and an extension of the Pólya tree model. *Biometrika*, 98:17–34.
- Wolpert, R. L., Ickstadt, K., and Hansen, M. B. (2003). A nonparametric Bayesian approach to inverse problems. In *Bayesian Statistics 7*, pages 403–417. Oxford Univ. Press, New York.
- Wong, W. and Ma, L. (2010). Optional Pólya tree and Bayesian inference. *Ann. Statist.*, 38:1433–1459.
- Zeng, D. and Lin, D. (2007). Maximum likelihood estimation in semiparametric regression models with censored data. *J. R. Stat. Soc. Ser. B*, 69:507–564.