Survival analysis via hierarchically dependent mixture hazards

Federico Camerlenghi[1,3], Antonio Lijoi[2,3] and Igor Prünster[2,3]

[1] Department of Economics, Management and Statistics, University of Milano–Bicocca, Piazza dell'Ateneo Nuovo 1, 20126 Milano, Italy.
[2] Department of Decision Sciences and Boconi Institute for Data Science and Analytics (BIDSA), Bocconi University, via Röntgen 1, 20136 Milano, Italy.
[3] Collegio Carlo Alberto, Piazza Vincenzo Arbarello 8, 10122 Torino, Italy.

Abstract

Hierarchical nonparametric processes are popular tools for defining priors on collections of probability distributions, which induce dependence across multiple samples. In survival analysis problems one is typically interested in modeling the hazard rates, rather than the probability distributions themselves, and the currently available methodologies are not applicable. Here we fill this gap by introducing a novel, and analytically tractable, class of multivariate mixtures whose distribution acts as a prior for the vector of sample–specific baseline hazard rates. The dependence is induced through a hierarchical specification for the mixing random measures that ultimately corresponds to a composition of random discrete combinatorial structures. Our theoretical results allow to develop a full Bayesian analysis for this class of models, which can also account for right–censored survival data and covariates, and we also show posterior consistency. In particular, we emphasize that the posterior characterization we achieve is the key for devising both marginal and conditional algorithms for evaluating Bayesian inferences of interest. The effectiveness of our proposal is illustrated through some synthetic and real data examples.

Keywords: Bayesian Nonparametrics, completely random measures, generalized gamma processes, hazard rate mixtures, hierarchical processes, meta–analysis, partial exchangeability.
AMS Classification 2010: 62F15; 62N02; 60G57.

1 Introduction

Hierarchical processes are hugely popular Bayesian nonparametric models, which have seen successful applications in linguistics, information retrieval, topic modeling and genomics, among others. They are ideally suited to model relationships across multiple samples, which may share distinct observations' values (or latent features, if used in a mixture setup). For instance, a topic might be shared by different documents of a *corpus* or a specific sub-sequence of a cDNA sequence might be recorded at different tissues of an organism. The prototype of this class of models is the *hierarchical Dirichlet process* (HDP) introduced in [42], which can be seen as an infinite–dimensional extension of the latent Dirichlet allocation model in [4]. Besides being a clever construction, the availability of suitable algorithms allow to perform Bayesian inference. The general structure of hierarchical processes amounts to

$$\begin{aligned}(\tilde{p}_1,\ldots,\tilde{p}_d)\,|\,\tilde{p}_0 &\stackrel{\text{iid}}{\sim} \tilde{\mathcal{L}}_0\\ \tilde{p}_0 &\sim \mathcal{L}_0\end{aligned} \tag{1}$$

where $\tilde{\mathcal{L}}_0$ is the probability distribution of each random probability measure $\tilde{p}_i$ such that $\mathbb{E}[\tilde{p}_i\,|\,\tilde{p}_0] = \int p\,\tilde{\mathcal{L}}_0(\mathrm{d}p) = \tilde{p}_0$, and $\mathcal{L}_0$ is such that $\mathbb{E}[\tilde{p}_0] = \int p\,\mathcal{L}_0(\mathrm{d}p) = P_0$, for some fixed non–atomic probability measure P_0. The vector of random probability measures in (1) defines a prior for the probability distributions of d partially exchangeable samples with dependence across samples being induced by $\tilde{p}_0$. Since $\tilde{p}_0$ and $(\tilde{p}_1,\ldots,\tilde{p}_d)$ are taken to be discrete random probability measures, distinct values are shared within and across the d samples. Clearly, in the extreme case of $\tilde{p}_0$ degenerating on P_0, the $\tilde{p}_i$'s are unconditionally independent corresponding to independence across samples with no shared values. See [42], [43] and, for a recent study of its distributional properties, [6].

Though currently available hierarchical processes are effective in modeling probability measures, in survival analysis one needs a flexible tool that is able to model directly the hazard rates. Here we successfully address the issue and introduce the natural counterpart, for hazard rate functions, of hierarchical random probability measures. This corresponds to a class of dependent hazard models defined as mixtures with mixing measures displaying a hierarchical dependence structure. Such a prior is ideally suited to model hazard rate functions associated to different, though related, populations in a similar fashion as hierarchical processes do for related probability distributions. Moreover, both censored observations and covariates are easily incorporated. More specifically, our proposal assumes a prior for the hazard rate functions of each individual sample identified by the distribution of the random mixture

$$\tilde{h}_\ell(t) = \int_{\mathbb{Y}} k(t;y)\,\tilde{\mu}_\ell(\mathrm{d}y) \qquad \ell = 1,\ldots,d \tag{2}$$

where $k(\,\cdot\,;\,\cdot\,)$ is a suitable kernel and $\tilde{\mu}_\ell$ a random measure. Note that an extension of such a proposal that accommodates also for subject–specific covariates can be achieved by resorting, e.g., to a semiparametric representation through a multiplicative term as in Cox proportional hazards models. Dependence among the hazard rates is then created at the level of the random measures through a hierarchical structure

$$\begin{aligned}(\tilde{\mu}_1\ldots\tilde{\mu}_d)\,|\,\tilde{\mu}_0 &\stackrel{\text{iid}}{\sim} \tilde{\mathcal{G}}_0\\ \tilde{\mu}_0 &\sim \mathcal{G}_0\end{aligned} \tag{3}$$

where $\tilde{\mathcal{G}}_0$ is the distribution of each random measure $\tilde{\mu}_\ell$ and depends on $\tilde{\mu}_0$, which in turn is distributed according to $\mathcal{G}_0$. An important feature of the proposed construction is that it allows for nonproportional hazard rates across samples. In the sequel $\tilde{\mu}_0$ and $\tilde{\mu}_\ell|\tilde{\mu}_0$, for $\ell = 1,\ldots,d$, will be assumed to be completely random measures (CRMs), which play a key role in Bayesian Nonparametrics as effectively described in [30]. This will allow, in the following sections, to derive some key distributional properties of $(\tilde{\mu}_0,\tilde{\mu}_1,\ldots,\tilde{\mu}_d)$, both *a priori* and *a posteriori*. Based on the latter, and conditional on a suitable latent structure, closed form expressions for posterior estimates of both the hazard rates and the survival functions are obtained as well as a sampling scheme for performing full Bayesian inference.

Survival analysis has been one of the driving application areas for Bayesian Nonparametrics since its early days and among the many seminal papers in the simple exchangeable setup we mention [14, 19, 20, 23, 16, 31]. The latter two papers are of particular importance to the present contribution since they introduced the mixture hazard structure (2) with a gamma process as mixing measure. Still in the exchangeable case, they were extended to general mixing CRMs in [26] and further investigated in e.g. [24, 36, 37, 39, 10, 38, 13]. Beyond the exchangeable case, a popular modeling strategy, which yields the desired heterogeneity across samples, relies on the construction of dependent and sample–specific random measures that, suitably transformed, allow to model partially exchangeable data. An example is the mixture transformation in (10), which is the focus of this contribution. This technique has been widely used for creating dependent random *probability* measures in the literature and examples are available in, e.g., [12, 11, 17, 35, 29, 34, 5]. For a review see [33]. The only contribution aiming to model directly dependent hazard rates, is [28], which however has the drawback of incurring into combinatorial, and hence computational, issues for $d > 2$ samples. In contrast, the methodology proposed here leads to simple and tractable expressions for $d > 2$ groups of observations.

The outline of the paper is as follows. In Section 2 the basic building blocks of our model, namely vectors of hierarchical CRMs, are defined. Hierarchically dependent random hazard rates are introduced and investigated in Section 3. Posterior consistency, in a partially exchangeable framework, is established in Section 4, while the posterior distribution is identified in Section 5. The theoretical findings form the basis for devising suitable marginal and conditional sampling schemes, which are described in Section 6. Finally a simulation study is considered in Section 7. Proofs, other relevant technical details, extensions accounting for right-censored data and covariates as well as additional illustrations are deferred to the Supplementary Material.

2 Hierarchical random measures

The key ingredient for our dependent hazard rates model is a vector of random measures $(\tilde\mu_1, \dots, \tilde\mu_d)$ with dependence induced by a hierarchical structure as in (3). First we introduce the notion of completely random measure (CRM). A CRM is a random element $\tilde\mu$, defined on some probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and taking values in the space $\mathsf{M}_{\mathbb{Y}}$ of boundedly finite measures on some Polish space $\mathbb{Y}$, such that $\tilde\mu(A_1), \dots, \tilde\mu(A_n)$ are mutually independent random variables for any choice of bounded and pairwise disjoint Borel sets $A_1, \dots, A_n$, and any $n \geq 1$. In the following we consider CRMs without fixed jumps and no deterministic component. They are characterized by the respective Laplace functional transform at any measurable function $g : \mathbb{Y} \to \mathbb{R}^+$

$$\mathbb{E}\Big[\mathrm{e}^{-\int_{\mathbb{Y}} g(y)\tilde\mu(\mathrm{d}y)}\Big] = \exp\left\{-\int_{\mathbb{R}^+\times\mathbb{Y}} (1 - e^{-sg(y)})\nu(\mathrm{d}s, \mathrm{d}y)\right\}, \tag{4}$$

where ν, known as the Lévy intensity measure, uniquely identifies $\tilde\mu$. Hence, the notation $\tilde\mu \sim \mathrm{CRM}(\nu)$. For completeness, we remind that $\tilde\mu$ may be seen as a functional of a Poisson random measure $\tilde N = \sum_{i\geq 1} \delta_{(J_i, Y_i)}$ on $\mathbb{R}^+ \times \mathbb{Y}$ characterized by a mean intensity measure ν such that for any Borel set A in $\mathbb{R}^+ \times \mathbb{Y}$ with $\nu(A) < \infty$ one has $\tilde N(A) \sim \mathrm{Po}(\nu(A))$. The CRM $\tilde\mu$ can be then represented as $\sum_{i\geq 1} J_i \delta_{Y_i}$, hence its realizations are a.s. discrete with both jumps and locations random. See [27] for an exhaustive account on CRMs.

In analogy to the hierarchical construction for random probability measures we define a hierarchical structure for CRMs as

$$\begin{aligned} \tilde\mu_\ell \,|\, \tilde\mu_0 &\overset{\text{ind}}{\sim} \mathrm{CRM}(\tilde\nu_\ell) \qquad \ell = 1, \dots, d \\ \tilde\mu_0 &\sim \mathrm{CRM}(\nu_0). \end{aligned} \tag{5}$$

with Lévy intensities of the form

$$\tilde{\nu}_\ell(\mathrm{d}s, \mathrm{d}y) = \rho_\ell(s)\, \mathrm{d}s\, \tilde{\mu}_0(\mathrm{d}y), \qquad \nu_0(\mathrm{d}s, \mathrm{d}y) = \rho_0(s)\, \mathrm{d}s\, c_0 P_0(\mathrm{d}y), \tag{6}$$

where P_0 is a diffuse probability measure on $\mathbb{Y}$, while ρ_ℓ (for $\ell = 1, \dots, d$) and ρ_0 are non–negative measurable functions such that $\int_0^\infty \min\{1, s\}\, \rho_\ell(s)\, \mathrm{d}s < \infty$ and $\int_0^\infty \min\{1, s\}\, \rho_0(s)\, \mathrm{d}s < \infty$.

Remark 1. The specification in (5) entails that the $\tilde{\mu}_\ell$'s may have different distributions, which yields greater flexibility in applications. Nonetheless, for the sake of simplicity we henceforth stick to the case where $\rho_\ell = \rho$ for any $\ell = 1, \dots, d$. The results we obtain can be adapted to recover the case where the ρ_ℓ's differ and we defer this to future work.

The fundamental tool to work with a hierarchical CRM vector $(\tilde{\mu}_1, \cdots, \tilde{\mu}_d)$ is represented by its Laplace functional transform, which has a simple structure. For $\ell = 1, \dots, d$, set $\tilde{\mu}_\ell(g_\ell) = \int_{\mathbb{Y}} f_\ell(y) \tilde{\mu}_\ell(\mathrm{d}y)$ with g_ℓ a non–negative real valued function. Moreover, we let $\psi^{(0)}(u) := \int_0^\infty (1 - e^{-su}) \rho_0(s) \mathrm{d}s$ and $\psi^{(\ell)}(u) := \int_0^\infty (1 - e^{-su}) \rho_\ell(s) \mathrm{d}s$, as $\ell = 1, \dots, d$. By exploiting conditional independence and (4), the Laplace functional transform is given by

$$\mathbb{E}[\mathrm{e}^{-\tilde{\mu}_1(g_1) - \cdots - \tilde{\mu}_d(g_d)}] = \exp\Big\{ - c_0 \int_{\mathbb{Y}} \psi^{(0)} \Big[\sum_{\ell=1}^d \psi^{(\ell)}(g_\ell(y)) \Big]\, P_0(\mathrm{d}y) \Big\} \tag{7}$$

Now we introduce two special cases to be considered throughout and which represent natural choices for hierarchical CRM constructions.

Example 1. (Hierarchical gamma CRM model). If both $\tilde{\mu}_\ell$'s and $\tilde{\mu}_0$ are gamma CRMs, which corresponds to $\rho_0(s) = \rho_\ell(s) = \mathrm{e}^{-s}\, s^{-1}$ with $c_0 = 1$, we obtain the hierarchical gamma CRM model, whose Laplace functional reduces to

$$\mathbb{E}[\mathrm{e}^{-\tilde{\mu}_1(g_1) - \, \cdots \, - \tilde{\mu}_d(g_d)}] = \exp\Big\{ - \int_{\mathbb{Y}} \log\Big(1 + \sum_{\ell=1}^d \log(1 + g_\ell(y)\Big)\, P_0(\mathrm{d}y) \Big\}.$$

Example 2. (Hierarchical generalized gamma CRM model). Let us assume that both $\tilde{\mu}_\ell$'s and $\tilde{\mu}_0$ are generalized gamma CRMs, i.e. $\rho_0(s) = e^{-s}\, s^{-1-\sigma_0}/\Gamma(1-\sigma_0)$ and $\rho_\ell(s) = e^{-s}\, s^{-1-\sigma}/\Gamma(1-\sigma)$ for some σ and σ_0 in $(0, 1)$, we obtain the hierarchical generalized gamma CRM model with Laplace functional

$$\begin{aligned} &\mathbb{E}[\mathrm{e}^{-\tilde{\mu}_1(g_1) - \cdots - \tilde{\mu}_d(g_d)}] \\ &\quad = \exp\Big\{ - \frac{c_0}{\sigma_0} \int_{\mathbb{Y}} \Big[\Big(\sum_{\ell=1} \frac{(g_\ell(y) + 1)^\sigma - 1}{\sigma} + 1 \Big)^{\sigma_0} - 1 \Big] P_0(\mathrm{d}y) \Big\}. \end{aligned}$$

Remark 2. The construction of hierarchical CRMs in (5) is closely related to the popular Bochner subordination in the theory of Lévy processes. Indeed, with $\mathbb{Y} = [0, 1]$, define $t \mapsto \tau_t = \tilde{\mu}_0((0, t])$ and $t \mapsto \xi_t^{(\ell)} = \tilde{\mu}_\ell((0, t])$. Moreover, let $P_0(\mathrm{d}t) = \mathrm{d}t$. In this case, $\{\tau_t :\ 0 \le t \le 1\}$ is known as a *subordinator* and the time–changed process $\{\xi_{\tau_t}^{(\ell)} :\ 0 \le t \le 1\}$ corresponds to the so–called Bochner's subordination. See, e.g., [2] and [41]. Hence, our proposal can be seen as an extension of such a construction to more abstract spaces. Subordination of Lévy processes has been widely used in Finance and the first contributions in this direction can be found in [32] and [7]. Very much in the spirit of the present paper, subordination has been applied to a vector of independent Brownian motions $(B_{\tau_t}^{(1)}, \dots, B_{\tau_t}^{(d)})$ in order to define dependent Lévy processes for financial applications. Examples can be found in the monograph by Cont et al. [8].

3 Hierarchical mixture hazard rates

If an exchangeable sequence of lifetimes, or time-to-event data, is from an abosolutely continuous distribution, with hazard rate h, the pioneering papers by [16] and [31] define a prior for h in terms of a mixture model. Recast in a more general setting, as done in [26], their prior equals the distribution of the random hazard

$$t \mapsto \tilde{h}(t) := \int_{\mathbb{Y}} k(t; y)\tilde{\mu}(\mathrm{d}y) \tag{8}$$

where $\tilde{\mu}$ is a mixing CRM and $k : \mathbb{R}^+ \times \mathbb{Y} \to \mathbb{R}^+$ is a suitable transition kernel. Dykstra and Laud [16] considered the case of $\tilde{h}$ being an *extended gamma process*, which corresponds to $\tilde{\mu}$ being a gamma CRM and $k(t; y) = \mathbb{1}_{(0,t]}(y)\beta(y)$ for some positive and right–continuous function β. Lo and Weng [31] investigated the *weighted gamma process*, which arises with $\tilde{\mu}$ still a gamma CRM and k a general kernel. The posterior characterization for the general case was first derived in [26].

Assuming the lifetimes are exchangeable is clearly restrictive for practical purposes. A typical example are data collected under d different experimental conditions that identify d samples or groups such as patients suffering from the same illness but undergoing different treatments resulting in homogeneity within the groups of patients undergoing the same treatment and heterogeneity across groups. It is worth noting that the d samples may be identified also by a vector of categorical covariates, as long as they take on finitely many values. For example, the real data illustration in the Supplementary Material (Section A.7) discusses the case where the patients' groups are determined by the hospital where they are treated and by the type of tumor they suffer from. In these cases, partial exchangeability [9] is the appropriate notion of dependence, since it implies that observations are exchangeable within the same group and conditionally independent across the different groups. More precisely, for an array $\{(X_{\ell,i})_{i\geq 1} : \ell = 1, \dots, d\}$ of $\mathbb{X}$–valued random elements we assume that for any $\ell_1 \neq \cdots \neq \ell_k$ in $\{1, \dots, d\}$, $(i_1, \dots, i_k) \in \mathbb{N}^k$ and $k \geq 1$

$$\begin{aligned} (X_{\ell_1,i_1}, \dots, X_{\ell_k,i_k}) \,|\, (\tilde{p}_1, \dots, \tilde{p}_d) \;&\overset{\text{iid}}{\sim}\; \tilde{p}_{\ell_1} \times \cdots \times \tilde{p}_{\ell_k} \\ (\tilde{p}_1, \dots, \tilde{p}_d) \;&\sim\; Q. \end{aligned} \tag{9}$$

with Q is a distribution on $\mathsf{P}_{\mathbb{X}}^d$ dictating the dependence across the different groups. The literature on dependent nonparametric priors essentially boils down to the definition of probability measures Q inducing natural dependence structures across groups while still preserving mathematical tractability. And, hierarchical processes represent one of the most popular instances.

In the survival context, i.e. $\mathbb{X} = \mathbb{R}^+$, it is often more convenient to devise inferential procedures that are suited for models based on hazard rates and this motivates our approach; in addition our proposal allows for a straightforward inclusion of possible censoring mechanisms and individuals' covariates as discussed in the Supplementary Material (see Section A.7). The prior specification we suggest resorts to a vector of dependent hazard rates $\tilde{h}_1, \dots, \tilde{h}_d$ based on hierarchically dependent CRMs according to the following

Definition 1. A partially exchangeable sequence $\{(X_{\ell,i})_{i\geq 1} : \ell = 1, \dots, d\}$ is directed by a *hierarchically dependent mixture hazard model* if it is characterized by (9) with

$$\tilde{p}_\ell((-\infty, t]) = \mathbb{1}_{(0,+\infty)}(t) \, \exp\Big(- \int_0^t \tilde{h}_\ell(s)\,\mathrm{d}s \Big) \qquad (\ell = 1, \dots, d) \tag{10}$$

where the hazard rates $\tilde{h}_\ell$ admit the mixture representation in (2) and the mixing hierarchical CRMs are defined as in (3).

Note that our model allows for non–proportional hazards across samples. Moreover, the survival function associated to the hazard rate (10) is

$$\tilde{S}_\ell(t) = \exp\Big(- \int_{\mathbb{Y}} K_t^{(\ell)}(y)\tilde{\mu}_\ell(\mathrm{d}y)\Big), \qquad \ell = 1, \cdots, d \tag{11}$$

where $K_t^{(\ell)}(y) = \int_0^t k(s;y)\,\mathrm{d}s$ for each $\ell = 1, \cdots, d$. Two important properties of the dependent survival functions are the determination of sufficient conditions guaranteeing they are proper and of their pairwise correlation structure. These are provided in the following result.

Theorem 1. *Consider a hierarchical mixture hazard model as in* Definition 1*.*

(i) *If* $\int_0^\infty \rho_\ell(s)\mathrm{d}s = +\infty$*, for* $\ell = 0, \ldots, d$*, and* $\lim_{t\to\infty} \int_0^t k(s;y)\mathrm{d}s = \infty$*,* P_0*–a.s., then with probability 1*

$$\lim_{t\to\infty} \tilde{S}_\ell(t) = 0 \qquad (\ell = 1, \ldots, d). \tag{12}$$

(ii) *The covariance between any two survival functions* $\ell_1 \neq \ell_2$*, time points* $t_1, t_2 \in \mathbb{R}^+$ *equals*

$$\begin{aligned} &\mathrm{Cov}(\tilde{S}_{\ell_1}(t_1), \tilde{S}_{\ell_2}(t_2)) \\ &\quad = \exp\Big(- c_0 \int_{\mathbb{Y}} \psi^{(0)}\Big(\sum_{i=1}^2 \psi^{(\ell_i)}(K_{t_i}^{(\ell_i)}(y))\Big)\, P_0(\mathrm{d}y)\Big) \\ &\qquad - \exp\Big(- c_0 \sum_{i=1}^2 \int_{\mathbb{Y}} \psi^{(0)}(\psi^{(\ell_i)}(K_{t_i}^{(\ell_i)}(y))\, P_0(\mathrm{d}y)\Big) \geq 0. \end{aligned} \tag{13}$$

As for (ii), the non–negativity of the covariance between random probabilities $\tilde{p}_i(A)$ and $\tilde{p}_j(A)$, for any $i \neq j$, is a common feature of a number of dependent processes' proposals in the literature. In particular, there are several instances where $\mathrm{Cov}(\tilde{p}_i(A), \tilde{p}_j(A))$ does not even depend on the specific set A and is then often interpreted as a measure of overall dependence between $\tilde{p}_i$ and $\tilde{p}_j$.

Theorem 1 can be nicely illustrated in two special cases that correspond to the hierarchical CRMs already considered in Section 2. Note further that we shall consider the $\tilde{\mu}_\ell$'s conditionally identically distributed, i.e. $\rho_\ell = \rho$ for any ℓ, and hence it is not surprising that in both examples $\mathrm{Cov}(\tilde{S}_{\ell_1}(t_1), \tilde{S}_{\ell_2}(t_1))$ does not depend on the specific $\ell_1 \neq \ell_2$.

Example 3. (Hierarchical gamma mixture hazard model). Consider the Dykstra–Laud kernel $k(t;y) := \mathbb{1}_{(0,t]}(y)\alpha(y)$ and, for simplicity, take $\alpha(y) \equiv \alpha > 0$. Combining this kernel with a hierarchical vector of gamma CRMs on $\mathbb{Y} = \mathbb{R}^+$ leads to a hierarchical gamma mixture hazard model. The associated survival functions are proper, given they clearly satisfy the conditions of (i) in Theorem 1. As for the pairwise correlation structure, by (ii) in Theorem 1 and calculations detailed in the Supplementary Material, one obtains for any $\ell_1 \neq \ell_2$ and $t_1 \leq t_2$ in $\mathbb{R}^+$

$$\begin{aligned} &\mathrm{Cov}(\tilde{S}_{\ell_1}(t_1), \tilde{S}_{\ell_2}(t_2)) = \\ &\quad \exp\Big(- c_0 \int_{t_1}^{t_2} \log(1 + \log(1 + \alpha\,(t_2 - y)))\, P_0(\mathrm{d}y)\Big) \\ &\quad \times \Big\{ \exp\Big(- c_0 \int_0^{t_1} \log\Big(1 + \sum_{i=1}^2 \log\Big(1 + \alpha\,(t_i - y)\Big)\Big)\, P_0(\mathrm{d}y)\Big) \\ &\qquad - \exp\Big(- \sum_{i=1}^2 c_0 \int_0^{t_1} \log(1 + \log(1 + \alpha\,(t_i - y)))\, P_0(\mathrm{d}y)\Big) \Big\}. \end{aligned} \tag{14}$$

Example 4. (Hierarchical generalized gamma mixture hazard model). If the Dykstra and Laud kernel in Example 3 is combined with a hierarchical vector of generalized gamma CRMs on $\mathbb{Y} = \mathbb{R}^+$, we obtain a hierarchical generalized gamma mixture hazard model. The conditions of (i) in Theorem 1 are met also in this case leading to proper survival functions. The pairwise correlation between survival functions, for any $\ell_1 \neq \ell_2$ and $t_1 \leq t_2$ in $\mathbb{R}^+$ is

$$\begin{aligned}
&\mathrm{Cov}(\tilde{S}_{\ell_1}(t_1), \tilde{S}_{\ell_2}(t_2)) = \\
&\exp\Big(-\frac{c_0}{\sigma_0}\int_{t_1}^{t_2}\Big[\Big(1+[(1+A_2(y))^\sigma - 1]/\sigma\Big)^{\sigma_0} - 1\Big]P_0(\mathrm{d}y)\Big) \\
&\times\Big\{\exp\Big(-\frac{c_0}{\sigma_0}\int_0^{t_1}\Big[\Big(\sum_{i=1}^{2}[(1+A_i(y))^\sigma - 1]/\sigma + 1\Big)^{\sigma_0} - 1\Big]P_0(\mathrm{d}y)\Big) \\
&-\exp\Big(-\frac{c_0}{\sigma_0}\sum_{i=1}^{2}\int_0^{t_1}\Big[\Big(1+[(1+A_i(y))^\sigma - 1]/\sigma\Big)^{\sigma_0} - 1\Big]P_0(\mathrm{d}y)\Big)\Big\}.
\end{aligned} \tag{15}$$

where $A_i(y) = \alpha(t_i - y)\mathbb{1}_{(0,t_i)}(y)$. This is again a consequence of (ii) in Theorem 1 and calculations detailed in the Supplementary Material.

4 Posterior consistency

In this section we investigate frequentist posterior consistency of the proposed class of priors. Most of the existing literature on the topic deals with sequences of exchangeable data and other specific forms of dependence (see [21] for an account). But very few results are available for partially exchangeable observations (see, e.g., [1]) and general theorems to study posterior consistency in a multiple samples framework, as the one we are studying here, are still missing. Here we provide a strategy to face consistency in a partially exchangeable setting and, then, adapt it to show consistency of hierarchical generalized gamma CRMs. Consider the partially exchangeable framework (9) for $\mathbb{R}^+$–valued random elements $X_{\ell,i}$'s and note that the hierarchically dependent mixture hazard model in Definition 1 implies that their distribution is absolutely continuous with respect to the Lebesgue measure. If $(\tilde{f}_1, \ldots, \tilde{f}_d)$ is the vector of random dependent densities associated with $(\tilde{p}_1, \ldots, \tilde{p}_d)$, one has

$$\tilde{f}_\ell(t) = \tilde{h}_\ell(t)\tilde{S}_\ell(t) = \tilde{h}_\ell(t)\, \mathrm{e}^{-\tilde{H}_\ell(t)} \tag{16}$$

where $t \mapsto \tilde{H}_\ell(t) = \int_0^t \tilde{h}_\ell(s)\mathrm{d}s$ is the cumulative hazard of the ℓ-th sample, with $\ell = 1, \ldots, d$. Let us denote by $\mathsf{F}_\mathbb{R}$ the space of all probability density functions, with respect to the Lebesgue measure on $\mathbb{R}$. We shall now assume the data from each sequence $(X_{\ell,i})_{i\geq 1}$, with $\ell = 1, \ldots, d$, are independently generated from a true and fixed density $f_\ell^{(0)}$ and we, then, check whether the posterior distribution of the vector $(\tilde{f}_1, \ldots, \tilde{f}_d)$ accumulates in a neighborhood of $\boldsymbol{f}^{(0)} := (f_1^{(0)}, \ldots, f_d^{(0)})$ in a suitable topology on the product space $\mathsf{F}_\mathbb{R}^d$. In the sequel we deal with *weak* consistency, hence we endow $\mathsf{F}_\mathbb{R}$ with the *weak* topology and for any $f \in \mathsf{F}_\mathbb{R}$ we denote by $A(f)$ a weak neighborhood of f. Moreover, the space $\mathsf{F}_\mathbb{R}^d$ is naturally endowed with the product topology.

In order to state the main results, we need to introduce some additional notation. First, we let $P_\ell^{(0)}$ denote the probability distribution associated with $f_\ell^{(0)}$ and $P_\ell^{(0),\infty}$ be the infinite product measure. Similarly, $P^{(0),\infty} := P_1^{(0),\infty} \times \ldots \times P_d^{(0),\infty}$. If Π is the prior distribution on $\mathsf{F}_\mathbb{R}^d$ induced by Q, we let $\Pi_{\boldsymbol{n}}(\,\cdot\,|\boldsymbol{X}_1, \ldots, \boldsymbol{X}_d)$ be the corresponding posterior and $\boldsymbol{n} = (n_1, \ldots, n_d)$. Hence, the goal is to identify sufficient conditions on the intensities ν_ℓ of the underlying CRMs ($\ell = 0, 1, \ldots, d$) such that, as $n_1, \ldots, n_d \to +\infty$, one has

$$\Pi_{\boldsymbol{n}}(A(\boldsymbol{f}^{(0)})|\boldsymbol{X}_1, \ldots, \boldsymbol{X}_d) \to 1 \qquad P^{(0),\infty} - \text{a.s.} \tag{17}$$

for any neighborhood $A(\boldsymbol{f}_0)$ of $\boldsymbol{f}^{(0)} = (f_1^{(0)}, \ldots, f_d^{(0)}) \in \mathsf{F}_{\mathbb{R}}^d$ in the weak topology. If (17) holds true, we say that Π is consistent at $\boldsymbol{f}^{(0)}$. One can, then, show the following

Theorem 2. *Let $(f_1^{(0)}, \ldots, f_d^{(0)})$ be an element of $\mathsf{F}_{\mathbb{R}}^d$, if for any weak neighborhood $A_\ell(f_\ell^{(0)})$ of $f_\ell^{(0)}$*

$$\Pi_{\ell, n_\ell}\Big(A_\ell(f_\ell^{(0)}) \,|\, \boldsymbol{X}_\ell\Big) := \mathbb{P}\Big[\tilde{f}_\ell \in A_\ell(f_\ell^{(0)}) \,|\, \boldsymbol{X}_\ell\Big] \to 1 \qquad P_\ell^{(0),\infty}\text{–a.s} \tag{18}$$

as $n_\ell \to \infty$ for any $\ell = 1, \ldots, d$, then (17) holds true and Π is weakly consistent at $(f_1^{(0)}, \ldots, f_d^{(0)})$.

It is worth remarking two important points related to this result. Firstly, from Theorem 2 it is apparent that showing (17) boils down to proving consistency for each $\tilde{f}_\ell$ given the ℓ–th sample $\boldsymbol{X}_\ell$ and, hence, allows us to address the problem in an exchangeable setting. Secondly, Theorem 2 may be rephrased for vectors of random probability measures $(\tilde{p}_1, \ldots, \tilde{p}_d)$ and its validity is not limited just to random dependent densities.

In view of Theorem 2, in order to show consistency with partially exchangeable arrays, we need to identify conditions for which (18) holds true, for any $\ell = 1, \ldots, d$. Henceforth we assume that the support of ρ_ℓ is the whole positive real line $\mathbb{R}^+$ and that

$$\int_{\mathbb{R}^+} \rho_\ell(s) \mathrm{d}s = +\infty$$

for each $\ell = 0, 1, \ldots, d$. Moreover, P_0 is such that its weak support is the whole space $\mathbb{Y}$. We tackle the problem by relying on [10, Theorem 2], which is proved for mixture hazards, and on a result on the support of hierarchical CRMs discussed in Section A.2.3 of the Supplementary Material. To fix the notation, $h_\ell^{(0)}$ is the true hazard rate for the ℓ–sample, namely $f_\ell^{(0)}(t) = h_\ell^{(0)}(t) \exp\{-\int_0^t h_\ell^{(0)}(s)\mathrm{d}s\}$. Let us now recall the following:

Theorem 3 ([10])**.** *Let $\tilde{f}_\ell$ be a random density function induced by a random hazard mixture model and denote its distribution by Π_ℓ. If the following conditions hold*

(i) *$f_\ell^{(0)} > 0$ on $(0, \infty)$ and $\int_{\mathbb{R}^+} \max\{\mathbb{E}\tilde{H}_\ell(t), t\} f_\ell^{(0)}(t)\mathrm{d}t < +\infty$,*

(ii) *either there exists $r > 0$ such that $\liminf_{t\to 0} \tilde{h}_\ell(t)/t^r = \infty$, a.s., or $h_\ell^{(0)}(0) > 0$,*

a sufficient condition for weak consistency of Π_ℓ at $f_\ell^{(0)}$ is

$$\Pi_\ell\Big(\Big\{h : \sup_{t \le T} |h(t) - h_\ell^{(0)}(t)| < \delta\Big\}\Big) > 0 \quad \forall\ T, \delta \in (0, \infty). \tag{19}$$

Theorem 3 provides us with a sufficient condition to prove weak consistency of random densities $\tilde{f}_\ell$. In order to verify condition (ii), one should study the small time behavior of $\tilde{h}_\ell$ otherwise one has to assume that the true hazard rate is such that $h_\ell^{(0)}(0) > 0$ as in [15].

We now focus on the case $\mathbb{Y} = \mathbb{R}^+$, and identify easier sufficient conditions ensuring the validity of Theorem 3(ii). See, e.g., Proposition 3 in [10]. We shall do this with hierarchies of the generalized gamma random measure

$$\rho_0(s) = \frac{1}{\Gamma(1-\sigma_0)} \frac{e^{-s}}{s^{1+\sigma_0}} \quad \text{and} \quad \rho_\ell(s) \equiv \rho(s) = \frac{1}{\Gamma(1-\sigma)} \frac{e^{-s}}{s^{1+\sigma}} \tag{20}$$

as $\ell = 1, \dots, d$, being $\sigma, \sigma_0 \in (0,1)$. This prior specification will be adopted in the numerical illustrations in Section 7.

Proposition 1. *Let $k(\,\cdot\,;\,\cdot\,)$ be the Dykstra and Laud kernel and let $\tilde\mu_\ell$ be the hierarchical generalized gamma process as in (20), with $\sigma, \sigma_0 \in (0,1)$ and P_0 such that $P_0((0,t))/t^{r'} \to C \in (0,\infty)$ as $t \to 0$ for some $r' > 0$. Then the corresponding mixture hazard $\tilde h_\ell$ satisfies condition* (ii) *of Theorem 3.*

We can state a theorem concerning marginal consistency of a hierarchical mixture hazard based on generalized gamma processes.

Theorem 4. *Let $\tilde h_\ell$ be a hierarchical mixture hazard as in* Proposition *1. Then Π_ℓ is consistent at any $f_\ell^{(0)} \in \mathsf{F}_1$, where F_1 is the set of densities satisfying $\int_{\mathbb{R}^+} \mathbb{E}[\tilde H(t)] f_\ell^{(0)}(t)\mathrm{d}t < \infty$, $h_\ell^{(0)}(0) = 0$ and $h_\ell^{(0)}(t)$ is strictly positive and non-decreasing for any $t > 0$.*

Combining Theorems 2 and 4 we obtain a consistency theorem for Π. With $\mathsf{F}_1^d = \times_1^d \mathsf{F}_1$, one easily deduces the following

Corollary 1. *Let $k(\,\cdot\,;\,\cdot\,)$ be the Dykstra and Laud kernel and $(\tilde\mu_1, \dots, \tilde\mu_d)$ a vector of hierarchical generalized gamma processes that meets the conditions in* Proposition 1. *Then Π is consistent at any $(f_1^{(0)}, \dots, f_d^{(0)}) \in \mathsf{F}_1^d$.*

Along similar lines one may prove consistency for other kernels k, as the ones considered by De Blasi et al. [10].

Remark 3. If $\sigma = \sigma_0 = 0$, we obtain the gamma process, but condition ii) is not satisfied in such a case and it should be replaced with a condition on the true hazard rate ($h_\ell^{(0)}(0) > 0$). Indeed, Drăghici and Ramamoorthi [15] show the consistency under this assumption and they consider a random hazard rate of the following type

$$\tilde h_\ell(t) = \tilde h_\ell(0) + \tilde\mu_\ell(0,t)$$

where $\tilde h_\ell(0)$ is a random element with support on $\mathbb{R}^+$, to consistently estimate the value of the hazard rate in $t = 0$.

5 Random partitions and Bayesian inference

The primary goal we wish to pursue here is the determination of a conditional probability distribution of the vector $(\tilde\mu_1, \cdots, \tilde\mu_d)$ in (5), given data from a partially exchangeable array and given a suitable collection of latent variables. Indeed, it will be assumed that the data are from an array $\{(X_{\ell,j})_{j\ge1} : \ell = 1, \dots, d\}$ as in (9) with

$$\tilde p_\ell((t, +\infty)) = \exp\Big(- \int_{\mathbb{Y}} \int_0^t k(s,y)\,\mathrm{d}s\,\tilde\mu_\ell(\mathrm{d}y)\Big) \tag{21}$$

for any $\ell = 1, \dots, d$. The hierarchical specification of $\tilde{\boldsymbol\mu} = (\tilde\mu_1, \dots, \tilde\mu_d)$, combined with the almost sure discreteness of the base measure $\tilde\mu_0$, poses some serious analytical challenges that need to be addressed if one wants to determine Bayesian inferences for survival data modeled through the random probability measure in (21). Hence, from a technical standpoint achieving a posterior characterization is much more difficult than analogous results for exchangeable (see [26]) and for alternative partially exchangeable priors (see, e.g., [28]). Here we successfully tackle this issue and show that the results we get are not only of theoretical interest per se, but they are also fundamental for devising efficient sampling schemes for an approximate evaluation of Bayesian inferences in this framework. The key is the introduction of two collections of latent variables that describe a sampling procedure somehow reminiscent of the Chinese restaurant franchise scheme introduced in [42] and extended to a more general framework in [6].

5.1 Latent variables and their partition structure

For simplicity in the sequel we will assume that all data are exact. The possible presence of censored observations, with censoring times being independent from all other random components of the model, can be easily accommodated for and this will be explicitly seen in Section 7 through an illustrative example. In order to keep the notation concise, we set $\tilde{\boldsymbol{\mu}} = (\tilde{\mu}_\ell)_\ell$ and $\boldsymbol{X} = (\boldsymbol{X}_\ell)_\ell$. From [25] it is seen that the likelihood function associated to a multiplicative intensity model as the one we are considering here is

$$\mathscr{L}(\tilde{\boldsymbol{\mu}}; \boldsymbol{X}) = \mathrm{e}^{-\sum_{\ell=1}^d \int_{\mathbb{Y}} K_\ell(y)\tilde{\mu}_\ell(\mathrm{d}y)} \prod_{\ell=1}^d \prod_{i=1}^{n_\ell} \int_{\mathbb{Y}} k(X_{\ell,i}; y)\, \tilde{\mu}_\ell(\mathrm{d}y) \tag{22}$$

where

$$K_\ell(y) := \sum_{i=1}^{n_\ell} K^{(\ell)}_{X_{\ell,i}}(y) = \sum_{i=1}^{n_\ell} \int_0^{X_{\ell,i}} k(s, y)\, \mathrm{d}s \tag{23}$$

for $\ell = 1, \dots, d$. A considerable simplification of (22) occurs if one removes the integrals by introducing a suitable sequence of latent variables $\boldsymbol{Y} = (\boldsymbol{Y}_\ell)_\ell$. This leads to

$$\mathscr{L}(\tilde{\boldsymbol{\mu}}; \boldsymbol{X}, \boldsymbol{Y}) = \mathrm{e}^{-\sum_{\ell=1}^d \int_{\mathbb{Y}} K_\ell(y)\, \tilde{\mu}_\ell(\mathrm{d}y)} \prod_{\ell=1}^d \prod_{i=1}^{n_\ell} k(X_{\ell,i}; Y_{\ell,i})\, \tilde{\mu}_\ell(\mathrm{d}Y_{\ell,i}). \tag{24}$$

It is worth noting that $\boldsymbol{Y} = (\boldsymbol{Y}_\ell)_\ell$ forms an array of partially exchangeable random elements and (24) suggests they are generated by a discrete random probability measure. Hence, with positive probability there will be tied values in $\boldsymbol{Y}$ both within each sample $\boldsymbol{Y}_\ell$ and across samples $\boldsymbol{Y}_{\ell_1}$ and $\boldsymbol{Y}_{\ell_2}$, for $\ell_1 \neq \ell_2$. In other terms, $\boldsymbol{Y}$ generates a random partition of the integers $[N] = \{1, \dots, N\}$, where $N = n_1 + \cdots + n_d$. This can be described by ordering the $Y_{\ell,i}$'s in such a way that $Y_{\ell,i} = Y_{\ell',j}$ entails $\bar{n}_{\ell-1} + i$ and $\bar{n}_{\ell'-1} + j$ are in the same partition set, where $\bar{n}_\ell = n_1 + \cdots + n_\ell$ for each $\ell = 1, \dots, d$ with the proviso $\bar{n}_0 = 0$. According to this notation, for any partition $\mathcal{C} = \{\bar{C}_1, \dots, \bar{C}_k\}$ of $[N]$, one can further decompose $\bar{C}_j = \bar{C}_{1,j} \cup \cdots \cup \bar{C}_{d,j}$ with $\bar{C}_{\ell,j} = \{\bar{n}_{\ell-1} + i \in \bar{C}_j : i = 1, \dots, n_\ell\}$. In the sequel we denote by Ψ_N the random partition of $[N]$ generated by $\boldsymbol{Y}$, and by $\boldsymbol{Y}^* := (Y_1^*, \dots, Y_k^*)$ the k distinct values associated with such a partition. With this notation, $\bar{C}_{\ell,j}$ refers to all elements of the ℓ–th sample that coincide with Y_j^*. It is easy to realize that the distribution of $(\boldsymbol{X}, \boldsymbol{Y}^*, \Psi_N)$ is equivalent to the distribution of the vector $(\boldsymbol{X}, \boldsymbol{Y})$, which are characterized in the following theorem, where we agree that $\sum_{i=1}^0 a_i \equiv 0$. In order to simplify the notation, we also set

$$Q(\boldsymbol{X}, \boldsymbol{Y}^*) = \prod_{\ell=1}^d \prod_{j=1}^k \prod_{i \in \bar{C}_{\ell,j}} k(X_{\ell,i}; Y_{\ell,i}), \quad \tau_q^{(\ell)}(u) = \int_0^\infty s^q \mathrm{e}^{-su} \rho_\ell(s) \mathrm{d}s$$

for any $u > 0$, $q \in \mathbb{N}$ and $\ell = 0, 1, \dots, d$. The subsequent expressions (25) and (27) we determine depend on sums over $\boldsymbol{q}$ and $\boldsymbol{i}$, and they may appear difficult to evaluate and interpret at a first glance. Nonetheless a considerable simplification, and additional intuition on these results, is achieved if one resorts to a version of the Chinese restaurant franchise metaphor suited to this particular setting. The franchise is made of d restaurants and the $Y_{\ell,i}$ variables identify the specific dish chosen from the menu by customer i at the ℓ–th restaurant. The values $Y_1^*,\dots,Y_k^*$ are the k distinct dishes selected by the N customers at the d restaurants of the franchise and $n_{\ell,j} = \mathrm{card}\{i : Y_{\ell,i} = Y_j^*\}$ is the number of customers being served dish j at the ℓ–th restaurant. Of course, one may have that $\{i : Y_{\ell,i} = Y_j^*\} = \emptyset$, in which case $n_{\ell,j} = 0$. With reference to previous notation, the set $C_{\ell,j} := \{i : Y_{\ell,i} = Y_j^*\}$ plays the role of $\bar{C}_{\ell,j}$, whose elements are equal to those of $C_{\ell,j}$ after subtracting the additive constant $\bar{n}_{\ell-1}$. At this stage it is, then, convenient to introduce latent variables $\boldsymbol{T}_\ell := (T_{\ell,1},\dots,T_{\ell,n_\ell})$ where, for each $\ell = 1,\dots,d$, $T_{\ell,j}$ is the label of the table where the j–th customer of restaurant ℓ is sitting and with the convention that customers seating at the same tables are being served the same dish. The $T_{\ell,i}$'s are generated by a discrete random probability measure and hence admit ties, namely $\{T_{\ell,i} : i \in C_{\ell,j}\}$ display $i_{\ell,j} \le n_{\ell,j}$ distinct values $T^*_{\ell,j,1},\dots,T^*_{\ell,j,i_{\ell,j}}$ that induce a partition of $[n_{\ell,j}] = \{1,\dots,n_{\ell,j}\}$ into $i_{\ell,j}$ sets

$$C_{\ell,j,t} = \{i \in C_{\ell,j} : T_{\ell,i} = T^*_{\ell,j,t}\} \quad (t = 1,\dots,i_{\ell,j})$$

with $q_{\ell,j,t} = \mathrm{card}(C_{\ell,j,t})$ such that $\sum_{t=1}^{i_{\ell,j}} q_{\ell,j,t} = n_{\ell,j}$. We are now ready to prove the main results of this section, that can be read in terms of the metaphor just outlined.

Theorem 5. *The probability distribution of* $(\boldsymbol{X},\boldsymbol{Y}^*)$ *and of* Ψ_N *equals*

$$\begin{aligned}
&\pi(\boldsymbol{X},\boldsymbol{Y}^*,\Psi_N)\\
&\quad\propto Q(\boldsymbol{X},\boldsymbol{Y}^*) c_0^k\, \mathrm{e}^{-c_0 \int_{\mathbb{Y}} \psi^{(0)}(\sum_{\ell=1}^d \psi^{(\ell)}(K_\ell(y)))\, P_0(\mathrm{d}y)}\\
&\quad\times \sum_{\boldsymbol{i}} \sum_{\boldsymbol{q}} \prod_{j=1}^k \tau^{(0)}_{i_{\bullet j}}\Big(\sum_{\ell=1}^d \psi^{(\ell)}(K_\ell(Y_j^*))\Big)\, P_0(\mathrm{d}Y_j^*)\\
&\quad\times \prod_{\ell=1}^d \prod_{j=1}^k \frac{1}{i_{\ell,j}!} \binom{n_{\ell,j}}{q_{\ell,j,1},\cdots,q_{\ell,j,i_{\ell,j}}} \prod_{t=1}^{i_{\ell,j}} \tau^{(\ell)}_{q_{\ell,j,t}}(K_\ell(Y_j^*))
\end{aligned} \tag{25}$$

where the sum with respect to $\boldsymbol{q}$ *runs over all the vectors of positive integers* $\{q_{\ell,j,t}\}$ *such that* $\sum_{t=1}^{i_{\ell,j}} q_{\ell,j,t} = n_{\ell,j}$ *for any* $\ell = 1,\dots,d$ *and* $j = 1,\dots,k$ *and* $\boldsymbol{i} = \{i_{\ell,j}\}$ *is through the set of all integers* $i_{\ell,j} \in \{1,\cdots,n_{\ell,j}\}$, *with* $i_{\bullet j} = \sum_{\ell=1}^d i_{\ell,j}$, *for any* $j = 1,\dots,k$.

This result is the backbone of theoretical and computational developments in the remainder of the paper. For example, one can deduce from (25) the following augmented joint distribution

$$\begin{aligned}
&\pi(\boldsymbol{X},\boldsymbol{Y}^*,\boldsymbol{T},\Psi_N)\\
&\quad\propto c_0^k\, Q(\boldsymbol{X},\boldsymbol{Y}^*)\, \mathrm{e}^{-c_0 \int_{\mathbb{Y}} \psi^{(0)}(\sum_{\ell=1}^d \psi^{(\ell)}(K_\ell(y)))\, P_0(\mathrm{d}y)}\\
&\quad\times \prod_{j=1}^k \tau^{(0)}_{i_{\bullet j}}\Big(\sum_{\ell=1}^d \psi^{(\ell)}(K_\ell(Y_j^*))\Big)\, P_0(\mathrm{d}Y_j^*) \prod_{\ell=1}^d \prod_{j=1}^k \prod_{t=1}^{i_{\ell,j}} \tau^{(\ell)}_{q_{\ell,j,t}}(K_\ell(Y_j^*))\, P_0(\mathrm{d}T^*_{\ell,j,t}),
\end{aligned} \tag{26}$$

where $\boldsymbol{i} = \{i_{\ell,j}\}$ and $\boldsymbol{q} = \{q_{\ell,j,t}\}$ are now fixed. Note that, since the specific table's label is not relevant, for simplicity and with no loss of generality we have assumed that each table's label still takes values in $\mathbb{Y}$. This formula provides some nice intuition on the Chinese restaurant franchise metaphor briefly discussed before Theorem 5. Indeed, if we agree that $\boldsymbol{D}_{-(\ell,i)}$ is the vector $(\boldsymbol{X}, \boldsymbol{Y}^*, \boldsymbol{T}, \boldsymbol{q}, \boldsymbol{i})$ after the removal of $(X_{\ell,i}, Y_{\ell,i}, T_{\ell,i})$, one has that the dish chosen by the ith customer in the ℓth restaurant and the table where she seats are determined through the following distribution

$$\mathbb{P}[Y_{\ell,i} = Y_j^*,\ T_{\ell,i} = T_{\ell,j,t}^* \mid \boldsymbol{D}_{-(\ell,i)}] \propto k(X_{\ell,i}; Y_j^*)\, \frac{\tau^{(\ell)}_{q_{\ell,j,t}+1}(K_\ell(Y_j^*))}{\tau^{(\ell)}_{q_{\ell,j,t}}(K_\ell(Y_j^*))}$$

which corresponds to the customer choosing the jth dish already ordered from the menu and seating at the already existing tth table, whereas

$$\mathbb{P}[Y_{\ell,i} = Y_j^*,\ T_{\ell,i} \notin \mathcal{T}_{-(\ell,i)} \mid \boldsymbol{D}_{-(\ell,i)}] \propto k(X_{\ell,i}; Y_j^*)\, \frac{\tau^{(0)}_{i_{\bullet,j}+1}(\sum_{\ell=1}^d \psi^{(\ell)}(Y_j^*))}{\tau^{(0)}_{i_{\bullet,j}}(\sum_{\ell=1}^d \psi^{(\ell)}(Y_j^*))}$$

is the probability that the customer still chooses the jth dish though decides to sit at a new table whose label will be drawn from P_0. Finally, the customer may choose a new dish and, then, seat at a new table with probability

$$\mathbb{P}[Y_{\ell,i} \notin \mathcal{Y}_{-(\ell,i)}, T_{\ell,i} \notin \mathcal{T}_{-(\ell,i)} \mid \boldsymbol{D}_{-(\ell,i)}] \propto c_0 \int_{\mathbb{Y}} k(X_{\ell,i}; y)\, \tau_1^{(\ell)}(K_\ell(y))\, \tau_1^{(0)}(\sum_{\ell=1}^d \psi^{(\ell)}(K_\ell(y))) P_0(\mathrm{d}y).$$

and the label of both the new dish and the new table are generated from P_0. This structure of the Chinese restaurant franchise will be used in the algorithm that will be used for determining Bayesian inferences. For details see the Supplementary Material.

Theorem 5 is also relevant for deducing the distribution of the random partition Ψ_N conditionally on the other variables of the model.

Corollary 2. *Let $\mathcal{C} = \{\bar{C}_1, \ldots, \bar{C}_k\}$ denote a partition of $[N]$ into k sets such that $\bar{C}_j = \bar{C}_{1,j} \cup \cdots \cup \bar{C}_{d,j}$. Moreover, we let $\mathrm{card}(\bar{C}_j) = n_j \geq 1$ and $\mathrm{card}(\bar{C}_{\ell,j}) = n_{\ell,j} \geq 0$. If Ψ_N denotes the random partition of $[N]$ induced by $\boldsymbol{Y}$, then $\mathbb{P}[\Psi_N = \mathcal{C} \mid \boldsymbol{X}]$ is proportional to*

$$\begin{aligned} c_0^k \sum_{\boldsymbol{i}} \sum_{\boldsymbol{q}} \Big\{ \prod_{\ell=1}^d \prod_{j=1}^k \frac{1}{i_{\ell,j}!} \binom{n_{\ell,j}}{q_{\ell,j,1}, \cdots, q_{\ell,j,i_{\ell,j}}} \Big\} \\ \times \prod_{j=1}^k \int_{\mathbb{Y}} \Big\{ \prod_{\ell=1}^d \prod_{i \in \bar{C}_{\ell,j}} k(X_{\ell,i-\bar{n}_{\ell-1}}; y) \Big\} \tau^{(0)}_{i_{\bullet j}} \Big(\sum_{\ell=1}^d \psi^{(\ell)}(K_\ell(y)) \Big) \\ \times \prod_{\ell=1}^d \prod_{t=1}^{i_{\ell,j}} \tau^{(\ell)}_{q_{\ell,j,t}}(K_\ell(y))\, P_0(\mathrm{d}y) \end{aligned} \tag{27}$$

where the sums with respect to $\boldsymbol{q}$ and $\boldsymbol{i}$ are as in Theorem 5.

At this point one may wonder whether there exist CRMs such that

$$\pi(\boldsymbol{Y}^*, \boldsymbol{T}, \Psi_N \mid \boldsymbol{X}) = \int \pi(\boldsymbol{Y}^*, \boldsymbol{T}, \Psi_N \mid \boldsymbol{X}, \boldsymbol{\mu})\, \mathbb{P}(\mathrm{d}\boldsymbol{\mu} \mid \boldsymbol{X}), \tag{28}$$

the answer is provided by the following result, which is based on Corollary 2.

Theorem 6. *Let* $m_0 \sim \mathrm{CRM}(\nu_0)$ *and* $m_{\ell,y} \sim \mathrm{CRM}(\nu_{\ell,y})$ *be independent and such that*

$$\nu_0(\mathrm{d}s, \mathrm{d}y) = c_0\, \mathrm{e}^{-\sum_{\ell=1}^d \psi^{(\ell)}(K_\ell(y))}\, \rho_0(s)\, \mathrm{d}s\, P_0(\mathrm{d}y)$$

$$\nu_{\ell,y}(\mathrm{d}s, \mathrm{d}w) = \mathrm{e}^{-K_\ell(y)\, w} \rho_\ell(s)\, \mathrm{d}s\, P_0(\mathrm{d}w)$$

Then,

$$\pi(\boldsymbol{Y}^*, \boldsymbol{T}, \Psi_N \mid \boldsymbol{X}, m_0, \boldsymbol{m}) \propto \prod_{\ell=1}^d \prod_{j=1}^k \Big\{ \prod_{i \in C_{\ell,j}} k(X_{\ell,i}, Y_j^*) \Big\} \prod_{t=1}^{i_{\ell,j}} m_0(\mathrm{d}Y_j^*) \prod_{r \in C_{\ell,j,t}} m_{\ell,Y_j^*}(\mathrm{d}T_{\ell,j,t}^*) \quad (29)$$

where $\boldsymbol{m}$ *denotes the vector containing the* m_{ℓ,Y_j^*}*'s.*

We have previously defined the random partition Ψ_N in such a way that there is a one to one correspondence between the two vectors $(\boldsymbol{Y}^*, \Psi_N)$ and $\boldsymbol{Y}$. Henceforth, we equivalently write $\boldsymbol{Y}$ instead of $(\boldsymbol{Y}^*, \Psi_N)$.

5.2 A posterior characterization

The result in Theorem 5, combined with the related augmented probability distribution in (26), paves the way to the determination of the posterior distribution of $\tilde{\boldsymbol{\mu}} = (\tilde\mu_1, \cdots, \tilde\mu_d)$, given the data and the latent variables. This is described in the next theorem, which shows that a structural conjugacy property holds true and the vector $\tilde{\boldsymbol{\mu}}$ is still hierarchical *a posteriori*. In order to state the result, let $I_1, \ldots, I_k$ be independent non–negative random variables with I_j having density function

$$f_j(s) \propto s^{i_{\bullet j}} e^{-s \sum_{\ell=1}^d \psi^{(\ell)}(K_\ell(Y_j^*))}\, \rho_0(s)\, \mathrm{d}s \quad (30)$$

and $\eta_0^* \sim \mathrm{CRM}(\nu_0^*)$ where

$$\nu_0^*(\mathrm{d}s, \mathrm{d}y) = \mathrm{e}^{-s \sum_{\ell=1}^d \psi^{(\ell)}(K_\ell(y))}\, \rho_0(s)\, \mathrm{d}s\, c_0\, P_0(\mathrm{d}y). \quad (31)$$

Finally, set $\tilde\mu_0^* = \eta_0^* + \sum_{j=1}^k I_j \delta_{Y_j^*}$ and correspondingly

$$\tilde\nu_\ell^*(\mathrm{d}s, \mathrm{d}y) = \mathrm{e}^{-sK_\ell(y)}\, \rho_\ell(s)\, \mathrm{d}s\, \tilde\mu_0^*(\mathrm{d}y) \quad (32)$$

for each $\ell = 1, \ldots, d$.

Theorem 7. *The posterior distribution of* $\tilde{\boldsymbol{\mu}}$*, conditional on* $(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{T})$*, equals the distribution of the vector of CRMs*

$$(\tilde\mu_1^*, \cdots, \tilde\mu_d^*) + \Big(\sum_{j=1}^{k_1} \sum_{t=1}^{i_{1,j}} J_{1,j,t} \delta_{Y_{1,j}^*}, \cdots, \sum_{j=1}^{k_d} \sum_{t=1}^{i_{d,j}} J_{d,j,t} \delta_{Y_{d,j}^*} \Big) \quad (33)$$

where $\tilde\mu_\ell^* \mid \tilde\mu_0^* \stackrel{\mathrm{ind}}{\sim} CRM(\tilde\nu_\ell^*)$*, for* $\ell = 1, \ldots, d$*, and the jumps* $J_{\ell,j,t}$ *are independent and non–negative random elements with corresponding density functions*

$$f_{\ell,j,t}(s) \propto s^{q_{\ell,j,t}} e^{-sK_\ell(Y_{\ell,j}^*)}\, \rho_\ell(s)\, \mathrm{d}s. \quad (34)$$

Moreover, the random elements $\tilde{\boldsymbol{\mu}}^* = (\tilde\mu_1^*, \cdots, \tilde\mu_d^*)$*,* η_0^**,* $(I_1, \ldots, I_k)$ *and* $\{J_{\ell,j,t}\}$ *are mutually independent.*

It is worth stressing that according to Theorem 7, the random measures $\tilde{\mu}_\ell^* + \sum_{j=1}^{k_\ell} \sum_{t=1}^{i_{\ell,j}} J_{\ell,j,t}\, \delta_{Y^*_{\ell,j}}$ are conditionally independent CRMs, given $\tilde{\mu}_0^*$. The main differences with respect to the prior (5) are the presence of the jumps $J_{\ell,j,t}$ at fixed locations identified by the distinct values $Y^*_{\ell,j}$ of the latent variables $\boldsymbol{Y}_\ell$ and the exponential updating of the conditional Lévy intensity as described in (32). Moreover, unlike (5) the CRM $\tilde{\mu}_0^*$ at the top hierarchy does also have jumps at fixed locations corresponding to the overall distinct latent variables' values $Y_1^*, \dots, Y_k^*$. Hence the hierarchical structure is preserved also *a posteriori* and a property of structural conjugacy holds true. Note that the posterior characterization given in Theorem 7 shares some interesting features with the one of [6, Theorem 10]. Indeed, at the CRM level, the posteriors exhibit structural analogies displaying an updated CRM independent of positive fixed jumps I_j's, that are shared across the groups, and of positive fixed jumps $J_{\ell,j,t}$'s, that are specific to group ℓ. Otherwise the posteriors are very different in distribution, given the different transformation of the CRMs they rely on. This, somehow surprising and interesting, finding is reminiscent of the structural analogies pointed out by [30] for the simple exchangeable case.

A posterior characterization, as the one in Theorem 7, is the starting point for evaluating Bayesian inferences of interest on functionals of $(\tilde{p}_1, \dots, \tilde{p}_d)$. For example, one can resort to (33) for determining a posterior estimate of the sample–specific survival functions. Indeed one can deduce the following

Corollary 3. *For any $t > 0$, the posterior estimate of the survival function $\tilde{S}_\ell(t)$ with respect to a quadratic loss is*

$$\mathbb{E}[\tilde{S}_\ell(t) \mid \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{T}] = \exp\Big(- c_0 \int_{\mathbb{Y}} \psi_*^{(0)}(\psi_*^{(\ell)}(K_t^{(\ell)}(y)))\, P_0(\mathrm{d}y) \Big)$$
$$\times \prod_{j=1}^{k} \frac{\int_0^\infty \mathrm{e}^{-s\{\psi_*^{(\ell)}(K_t^{(\ell)}(Y_j^*)) + \sum_{h=1}^d \psi^{(h)}(K_h(Y_j^*))\}}\, s^{i_{\bullet j}} \rho_0(s) \mathrm{d}s}{\int_0^\infty \mathrm{e}^{-s \sum_{h=1}^d \psi^{(h)}(K_h(Y_j^*))} s^{i_{\bullet j}} \rho_0(s)}$$
$$\times \prod_{j=1}^{k_\ell} \prod_{t=1}^{i_{\ell,j}} \frac{\int_0^\infty \mathrm{e}^{-s(K_t^{(\ell)}(Y^*_{\ell,j}) + K_\ell(Y^*_{\ell,j}))} s^{q_{\ell,j,t}} \rho_\ell(s) \mathrm{d}s}{\int_0^\infty \mathrm{e}^{-sK_\ell(Y^*_{\ell,j})} s^{q_{\ell,j,t}} \rho_\ell(s) \mathrm{d}s} \tag{35}$$

where

$$\psi_*^{(0)}(u) = \int_0^\infty (1 - \mathrm{e}^{-su})\, \mathrm{e}^{-s \sum_{\ell=1}^d \psi^{(\ell)}(K_\ell(y))} \rho_0(s) \mathrm{d}s$$
$$\psi_*^{(\ell)}(u) = \int_0^\infty (1 - \mathrm{e}^{-su})\, \mathrm{e}^{-sK_\ell(y)} \rho_\ell(s) \mathrm{d}s \quad \ell = 1, \dots, d.$$

The posterior expectation of $\tilde{S}_\ell$, as described in Corollary 3, is conditional on the data $\boldsymbol{X}$ and on all the latent variables. Therefore, in order to obtain a posterior estimate of the survival function, one marginalizes (35) with respect to the posterior distribution of $(\boldsymbol{Y}, \boldsymbol{T})$, given $\boldsymbol{X}$. This goal may be achieved resorting to a Gibbs sampler, since such a marginalization cannot be handled analytically, and this will be the subject of Section 6.1. Analogously one can obtain the posterior expectation of the hazard rate for each group of survival times.

Corollary 4. *For any $t > 0$, the posterior estimate of the hazard rate $\tilde{h}_\ell(t)$ conditionally given*

$\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{T}$, *under a square loss function is*

$$\int_{\mathbb{Y}} k(t;y) \int_0^\infty s e^{-sK_\ell(y)} \rho_\ell(s) \mathrm{d}s \int_0^\infty w e^{-w \sum_{h=1}^d K_h(y)} \rho_0(w) \mathrm{d}w c_0 P_0(\mathrm{d}y)$$
$$+ \sum_{j=1}^k k(t;Y_j^*) \int_0^\infty s e^{-sK_\ell(Y_j^*)} \rho_\ell(s) \mathrm{d}s \int_0^\infty w f_j(w) \mathrm{d}w + \sum_{j=1}^{k_\ell} \sum_{t=1}^{i_{\ell,j}} k(t;Y_{\ell,j}^*) \int_0^\infty w f_{\ell,j,t}(w) \mathrm{d}w.$$

6 Marginal and conditional samplers

We now rely on the theoretical finding of the previous sections to describe a marginal and a conditional sampler, which can be used, for instance, to obtain posterior estimates of the survival functions. More precisely in Section 6.1 we develop a MCMC procedure to estimate dependent survival functions, based on Theorem 7 and Corollary 3. We in fact use the output of such an MCMC procedure to estimate the posterior expected values of the random survival functions, marginalizing out the CRMs. Hence in Section 6.1 we devise a so–called *marginal* algorithm. On the other hand, we also describe a *conditional* strategy in Section 6.2, which allows us to simulate the trajectories of the hierarchical CRMs.

6.1 A marginal MCMC sampler

Though here we will focus on a special case where each $\tilde{\mu}_\ell$ is a generalized gamma process, more specifically $\rho_\ell(s) = e^{-s}s^{-1-\sigma}/\Gamma(1-\sigma)$ for $\ell = 1, \ldots, d$ and $\rho_0(s) = e^{-s}s^{-1-\sigma_0}/\Gamma(1-\sigma_0)$, our analysis can be extended to any other choice of Lévy intensities that satisfy the conditions (i)–(ii) in Theorem 1. We further assume that $\mathbb{Y} = \mathbb{R}^+$ and set $k(t;y) := \mathbb{1}_{(0,t]}(y)\alpha(y)$ for some positive right–continuous real–valued function $\alpha(\,\cdot\,)$. Note that when $\sigma = \sigma_0 = 0$, each random hazard rate is an extended gamma process, given $\tilde{\mu}_0$, with parameters $(\tilde{\mu}_0, \alpha)$. The extended gamma process has been introduced in [16] to model monotone increasing hazard rates and it has been extensively used in an exchangeable setting. We will further assume that the function $\alpha(y) \equiv \alpha$ is constant and we specify a prior for α. We are able to deal with the general case $d \geq 2$, without analytical difficulties with respect the situation in which $d = 2$. One can now specialize Corollary 3 to this case and get the following result.

Corollary 5. *Let* $\rho_\ell(s) = e^{-s}s^{-1-\sigma}/\Gamma(1-\sigma)$ *for* $\ell = 1, \ldots, d$, $\rho_0(s) = e^{-s}s^{-1-\sigma_0}/\Gamma(1-\sigma_0)$, *and* $k(t;y) := \alpha\, \mathbb{1}_{(0,t]}(y)$. *Then the posterior expected value* $\mathbb{E}[\tilde{S}_\ell(t) \mid \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{T}]$ *equals*

$$\exp\Big(-\frac{c_0}{\sigma_0}\int_{\mathbb{Y}} F_\ell(y)\, P_0(\mathrm{d}y)\Big) \prod_{j=1}^k \Big\{1 + \frac{[1+K_\ell(Y_j^*)+K_t^{(\ell)}(Y_j^*)]^\sigma - [1+K_\ell(Y_j^*)]^\sigma}{\sum_{h=1}^d [(K_h(Y_j^*)+1)^\sigma - 1]}\Big\}^{-(i_{\bullet j}-\sigma_0)}$$
$$\times \prod_{j=1}^{k_\ell} \Big\{1 + \frac{K_t^{(\ell)}(Y_{\ell,j}^*)}{1+K_\ell(Y_{\ell,j}^*)}\Big\}^{-(n_{\ell,j}-i_{\ell,j}\sigma)}$$

being

$$F_\ell(y) = \left[\sum_{h=1}^d \frac{(K_h(y)+1)^\sigma - 1}{\sigma} + 1\right]^{\sigma_0} \left\{\left[1 + \frac{[1+K_\ell(y)+K_t^{(\ell)}(y)]^\sigma - [1+K_\ell(y)]^\sigma}{\sigma\left[\sum_{h=1}^d \frac{(K_h(y)+1)^\sigma - 1}{\sigma} + 1\right]}\right]^{\sigma_0} - 1\right\}$$

for each $\ell = 1, \ldots, d$.

As far as the hazard is concerned, an application of Corollary 4 in the case of hierarchies of generalized gamma process leads to the following:

Corollary 6. *Let* $\rho_\ell(s) = e^{-s}s^{-1-\sigma}/\Gamma(1-\sigma)$ *for* $\ell = 1, \dots, d$, $\rho_0(s) = e^{-s}s^{-1-\sigma_0}/\Gamma(1-\sigma_0)$, *and* $k(t; y) := \alpha\, \mathbb{1}_{(0,t]}(y)$, *then the posterior expected value* $\mathbb{E}[\tilde{h}_\ell(t)|\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{T}]$ *equals*

$$\int_{\mathbb{Y}} \frac{k(t;y)}{\left(1+\sum_{h=1}^d K_h(y)\right)^{1-\sigma_0}(1+K_\ell(y))^{1-\sigma}} c_0 P_0(\mathrm{d}y)$$
$$+ \sum_{j=1}^{k} \frac{k(t;Y_j^*)(i_{\bullet j}-\sigma_0)}{\left(1+\sigma^{-1}\sum_{h=1}^d[(1+K_h(Y_j^*))^\sigma - 1]\right)\left(1+K_\ell(Y_j^*)\right)^{1-\sigma}} \tag{36}$$
$$+ \sum_{j=1}^{k_\ell}\sum_{t=1}^{i_{\ell,j}} \frac{k(t;Y_{\ell,j}^*)(q_{\ell,j,t}-\sigma)}{(1+K_\ell(Y_{\ell,j}^*))}$$

for any $\ell = 1, \dots, d$.

It is now immediate to check that, with our choices of the kernel $k(t;y)$ and of the Lévy intensities ρ_ℓ, one has

$$K_t^{(\ell)}(y) = \alpha\,(t-y)\mathbb{1}_{(0,t]}(y)$$
$$K_\ell(y) = \alpha \sum_{\{i:\ y\le X_{\ell,i}\}} X_{\ell,i} - y\alpha \cdot \#\{i:\ y \le X_{\ell,i}\}.$$

An approximate evaluation of the posterior estimate $\mathbb{E}[\tilde{S}_\ell(t)|\boldsymbol{X}]$ of the survival function $\tilde{S}_\ell$ may be obtained by means of an MCMC procedure. At the same time, one can obtain a posterior estimate of the hazard rate $\tilde{h}_\ell(t)$. To this end, the full conditional distributions of the latent variables must be identified. The detailed description of the Gibbs sampler and all the full conditional distributions is reported in Section A.5. To complete the picture, we assume α and c_0 are independent and further specify the following priors

$$\alpha \sim \mathrm{Ga}(a,b),\ c_0 \sim \mathrm{Ga}(a_0, b_0). \tag{37}$$

For computational convenience σ and σ_0 are assumed to be fixed values.

6.2 Conditional sampler

The key distinctive feature of the MCMC procedure detailed in previous Section 6.1 is the marginalization with respect to the mixing dependent CRMs $\tilde{\mu}_\ell$'s. This is very useful when one is mainly interested in determining estimates of the survival functions, at any point $t > 0$, as posterior expected values of $\tilde{S}_\ell(t)$'s. See Corollary 5. On the contrary, in the present section we are going to develop a *conditional* algorithm that generates trajectories of the $\tilde{\mu}_\ell$'s from their respective posterior distributions. Our strategy relies on Theorem 7 and a proper adaptation of the algorithm by Wolpert and Ickstadt [44]. A conditional sampler is very useful for many practical reasons, and we here confine ourselves to just mentioning that: (i) it allows to estimate the actual posterior distribution of the survival functions or, equivalently, of the hazard rates at any time point t; (ii) it yields estimates of the posterior credible intervals for the estimated quantities that are more reliable than those arising from marginal samplers.

As for (i), if $\boldsymbol{\Phi}_1, \dots, \boldsymbol{\Phi}_N$ denote N posterior samples of the latent vector $\boldsymbol{\Phi}$ obtained through an MCMC procedure, and if in addition $\tilde{\mu}_\ell^{(i,1)}, \dots, \tilde{\mu}_\ell^{(i,M)}$ are M trajectories of $\tilde{\mu}_\ell$, conditionally given

$\mathbf{\Phi}_i, \boldsymbol{X}$, then the posterior distribution function of survival function $\tilde{S}_\ell(t)$, at t, is approximated by

$$\mathbb{P}(\tilde{S}_\ell(t) \le s | \boldsymbol{X}) \approx \frac{1}{NM} \sum_{i=1}^{N} \sum_{k=1}^{M} \mathbb{1}_{\left\{\tilde{S}_\ell^{(i,k)}(t) \le s\right\}}, \tag{38}$$

where

$$\tilde{S}_\ell^{(i,k)}(t) := \exp\Big(- \int_0^t \int_{\mathbb{Y}} k(s;y) \tilde{\mu}_\ell^{(i,k)}(\mathrm{d}y) \mathrm{d}s \Big).$$

See Section A.6.1 for details on the derivation of (38). Being provided with the approximation of the posterior distribution of the survival functions (38), it is now clear one can also address the aforementioned point (ii).

Hence, we need to devise an algorithm that samples the trajectories of the $\tilde{\mu}_\ell$'s from their posterior distribution and plug them in (38). Having fixed $\ell = 1, \ldots, d$, the computational procedure is based on Theorem 7 and a non–trivial adaptation of the algorithm developed by Wolpert and Ickstadt [44]. The technique suggested in [44] is easier to implement in the presence of non–homogeneous CMRs, as those considered here, if compared to an alternative and popular sampler yielded by the Ferguson and Klass [18] representation. Nonetheless, the jumps of the CRMs are not sampled in a decreasing order and this may be a serious drawback when one wants to obtain an approximate draw of a CRM by discarding the infinite number of jumps whose size does not exceed a fixed threshold, $\varepsilon > 0$ say. We have been able to successfully address this issue for the CRMs

$$\eta_0^* = \sum_{h \ge 1} J_h^{(0)} \delta_{y_h^{(0)}}, \qquad \tilde{\mu}_\ell^* = \sum_{h \ge 1} J_h^{(\ell)} \delta_{y_h^{(\ell)}}, \tag{39}$$

for $\ell = 1, \ldots, d$, where $y_h^{(0)} \overset{\text{iid}}{\sim} P_0$ and $y_h^{(\ell)} \,|\, \tilde{\mu}_0^* \overset{\text{iid}}{\sim} \tilde{\mu}_0^* / \tilde{\mu}_0^*(\mathbb{Y})$. See Theorem 7. Let, now, $S_h^{(0)}$ and $S_h^{(\ell)}$ be the points of a unit rate Poisson process, for $h \ge 1$, namely $S_h^{(0)} - S_{h-1}^{(0)}$ and $S_h^{(\ell)} - S_{h-1}^{(\ell)}$ are independent random variables whose distribution is negative exponential with unit mean. Hence, according to [44] the jumps $J_h^{(0)}$ and $J_h^{(\ell)}$ in (39) are such that

$$S_h^{(0)} = c_0 \int_{J_h^{(0)}}^{\infty} \mathrm{e}^{-s \sum_{\ell=1}^d \psi_\ell (K_\ell(y_h^{(0)}))} \rho_0(s) \, \mathrm{d}s \tag{40}$$

$$S_h^{(\ell)} = \tilde{c} \int_{J_h^{(\ell)}}^{\infty} \mathrm{e}^{-s K_\ell(y_h^{(\ell)})} \rho_\ell(s) \, \mathrm{d}s, \tag{41}$$

where $\tilde{c} := \sum_{h \ge 1} J_h^{(0)} \delta_{y_h^{(0)}} + \sum_{j=1}^k I_j \delta_{Y_j^*}$ is the total mass of $\tilde{\mu}_0^*$. One can, then, show the following

Proposition 2. *The sequences of non–negative random variables* $(\tilde{J}_h^{(0)})_{h \ge 1}$ *and* $(\tilde{J}_h^{(\ell)})_{h \ge 1}$*, for* $\ell = 1, \ldots, d$*, defined through the equations*

$$S_h^{(0)} = c_0 \int_{\tilde{J}_h^{(0)}}^{\infty} \rho_0(s) \, \mathrm{d}s, \qquad S_h^{(\ell)} = \tilde{c} \int_{\tilde{J}_h^{(\ell)}}^{\infty} \rho_\ell(s) \, \mathrm{d}s$$

are monotonically decreasing and dominate $(J_h^{(0)})_{h \ge 1}$ *and* $(J_h^{(\ell)})_{h \ge 1}$ *in* (40)*–*(41)*, namely* $J_h^{(0)} \le \tilde{J}_h^{(0)}$ *and* $J_h^{(\ell)} \le \tilde{J}_h^{(\ell)}$*, for any* $h \ge 1$ *and* $\ell = 1, \ldots, d$*, almost surely.*

In order to make precise the use of this result, let us focus attention on η_0^* in (39). If one wants to determine a finite–sum approximation of η_0^*, one can fix $\varepsilon > 0$ and discard jumps whose size is smaller than ε by setting the truncation level H^ε as the minimum value of h such that the additional jump satisfies $\tilde{J}_h^{(0)} \le \varepsilon$, i.e.

$$\eta_0^* \approx \sum_{h=1}^{H^\varepsilon} J_h^{(0)} \delta_{y_h^{(0)}}. \tag{42}$$

By virtue of Proposition 2, such a procedure guarantees that the heights of the discarded jumps $J_h^{(0)}$'s in (42) are smaller than ε. Similar arguments apply to $\tilde{\mu}_\ell^*$.

We now have all the necessary ingredients to describe the conditional algorithm.

(1) Generate $\tilde{\mu}_0$ from its posterior distribution, which is described right before Theorem 7, namely proceed as follows.

(1.a) Generate the random jumps I_j, as $j = 1, \dots, k$, whose density is proportional to (30);

(1.b) generate η_0^* using the algorithm developed by Wolpert and Ickstadt [44], i.e. fix a threshold level $\varepsilon > 0$ and proceed with the following steps:

- generate the atom of the CRM $y_h^{(0)} \sim P_0$;
- generate the waiting times $S_h^{(0)}$ of a standard Poisson process, that is to say $S_h^{(0)} - S_{h-1}^{(0)}$ are independent and identically distributed exponential random variables with unit mean;
- determine the jump $J_h^{(0)}$ by inverting the Lévy intensity (31), i.e. by solving (40);
- stop the procedure at $H_0^\varepsilon := \min\{h : \tilde{J}_h^{(0)} \le \varepsilon\}$ where the auxiliary jumps $\tilde{J}_h^{(0)}$ are determined according to Proposition 2.

(1.c) Evaluate an approximate draw of $\tilde{\mu}_0^*$

$$\tilde{\mu}_{0,\varepsilon}^* = \sum_{h=1}^{H_0^\varepsilon} J_h^{(0)} \delta_{y_h^{(0)}} + \sum_{j=1}^{k} I_j \delta_{Y_j^*}.$$

(2) Generate $\tilde{\mu}_\ell$, given $(\tilde{\mu}_0^*, \boldsymbol{X})$, as follows.

(2.a) Generate the $J_{\ell,j,t}$'s from the density in (34);

(2.b) with $\varepsilon > 0$ fixed, generate the $\tilde{\mu}_\ell^*$'s using the posterior representation in Theorem 7 and the algorithm by Wolpert and Ickstadt [44], namely:

- generate the atoms of the CRMs $y_h^{(\ell)} \sim P_{0,\varepsilon}^* = \tilde{\mu}_{0,\varepsilon}^* / \tilde{\mu}_{0,\varepsilon}^*(\mathbb{Y})$;
- generate the waiting times $S_h^{(\ell)}$ of a standard Poisson process, that is to say $S_h^{(\ell)} - S_{h-1}^{(\ell)}$ are independent and identically distributed exponential random variables with unit mean;
- determine the jump $J_h^{(\ell)}$ by inverting the Lévy intensity (41);
- stop the procedure at $H_\ell^\varepsilon := \min\{h : \tilde{J}^{(\ell)} \le \varepsilon\}$, where the auxiliary jumps $\tilde{J}_h^{(\ell)}$ are determined according to Proposition 2;

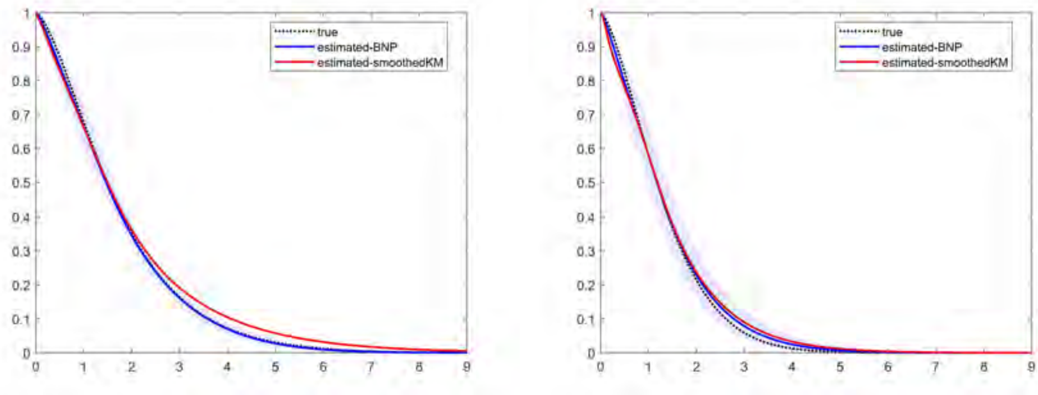
(2.c) evaluate an approximate draw from the posterior of $\tilde{\mu}_\ell$, by putting

$$\tilde{\mu}_\ell \approx \sum_{h=1}^{H_\ell^\varepsilon} J_h^{(\ell)} \delta_{y_h^{(\ell)}} + \sum_{j=1}^{k_\ell} \sum_{t=1}^{i_{\ell,j}} J_{\ell,j,t} \delta_{Y_{\ell,j}^*}.$$

For the numerical experiments of Section 7, we have implemented the conditional algorithm when $\tilde{\mu}_\ell$ is such that $\rho_\ell(s) = (\Gamma(1-\sigma))^{-1} e^{-s}/s^{1+\sigma}$ for $\ell = 1, \dots, d$, and $\rho_0(s) = (\Gamma(1-\sigma_0))^{-1} e^{-s}/s^{1+\sigma_0}$, i.e. the CRMs are generalized gamma processes exactly as in the previous section. For the readers' convenience we have specialized the conditional sampler for such a noteworthy example in Section A.6.3.

7 Illustrations

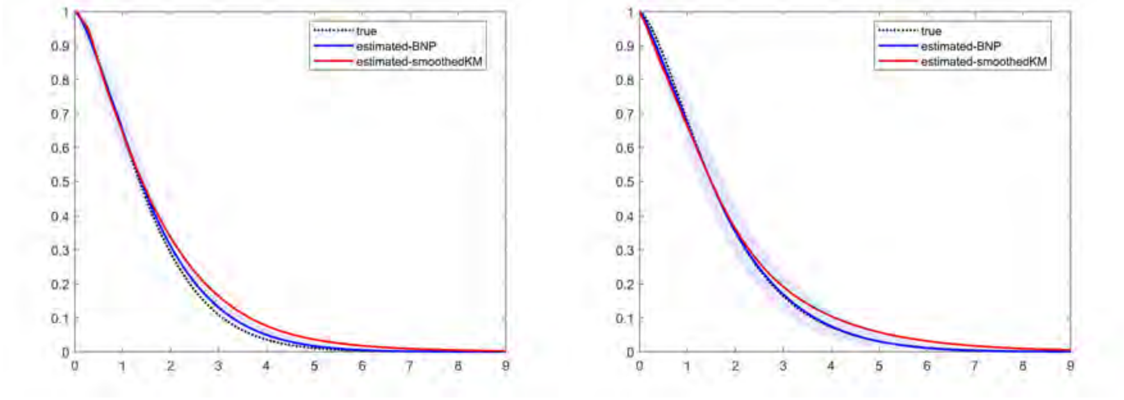
In this section we display an illustrative example that provide evidence of the effectiveness of our proposal on simulated datasets. In Section A.7 we also discuss how to adapt our strategy in presence of covariates and right–censored survival times, dealing with tumor survival data.



(a) (b)

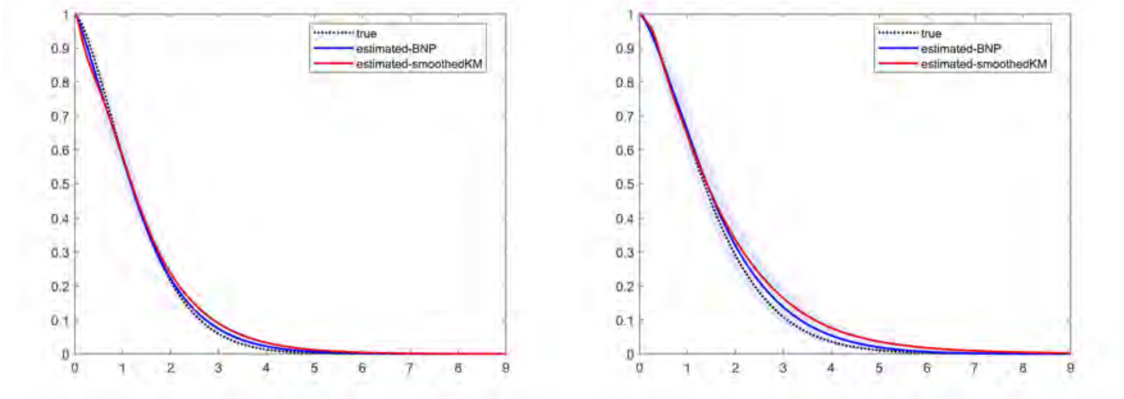
Figure 1: True survival functions (dotted) versus estimated (straight blue) for the first sample and 95% posterior credible intervals estimated using the marginal method (a) and conditional method (b). The BNP estimates are also compared with the corresponding frequentist estimate (straight red).

In the following simulated scenario we have run the MCMC algorithm for 50,000 iterations with a burn–in period of 20,000 sweeps: the number of iterations allow the convergence of all the MCMC procedures. We have considered a synthetic dataset composed by $d = 3$ groups of observations, more specifically the data are generated by three different mixtures of Weibull distributions $f_1 \sim \text{Wbl}(1.5, 3/2)$, $f_2 \sim 0.5\text{Wbl}(2.5, 3/2)+0.5\text{Wbl}(1.5, 3/2)$ and $f_3 \sim 0.5\text{Wbl}(2, 3/2)+0.5\text{Wbl}(1.5, 3/2)$. For any $\ell = 1, 2, 3$, we have set $n_\ell = 100$. For the sake of simplifying notation we denote by S_ℓ the survival function associated to f_ℓ, for any $\ell = 1, 2, 3$. As priors are concerned, we have used the specifications $a = a_0 = 1$, $b^{-1} = b_0^{-1} = 10$ and $T = 30$ which correspond to a non–informative



(a) (b)

Figure 2: True survival functions (dotted) versus estimated (straight blue) for the second sample and 95% posterior credible intervals estimated using the marginal method (a) and conditional method (b). The BNP estimates are also compared with the corresponding frequentist estimate (straight red).



(a) (b)

Figure 3: True survival functions (dotted) versus estimated (straight blue) for the third sample and 95% posterior credible intervals estimated using the marginal method (a) and conditional method (b). The BNP estimates are also compared with the corresponding frequentist estimate (straight red).

specification. While we have fixed the values of $\sigma = \sigma_0 = 0.25$. We compare our BNP estimates (obtained via the marginal and conditional algorithm) with a frequentist estimate which is based on a smoothed version of the Kaplan–Meier estimator (using the optimized Matlab command `ksdensity`). Since the data have been generated independently across groups, we evaluate the

Table 1: The Kolmogorov distance between the true and estimated survival functions for the different methodologies.

	$d_K(S_1, \hat{S}_1)$	$d_K(S_2, \hat{S}_2)$	$d_K(S_3, \hat{S}_3)$
marginal BNP	0.033	0.027	0.023
conditional BNP	0.031	0.024	0.031
smoothed KM	0.060	0.042	0.056

frequentist estimator separately for each group. Figure 1 compares the true survival curve, the estimated survival function obtained via the marginal algorithm (in blue) and the frequentist estimate (in red). In Figure 1(a) the estimated credible intervals are obtained through the outputs of the marginal procedure, while in Figure 1(b) they are evaluated through the conditional method. More precisely the BNP estimated survival function in Figure 1(a) has been obtained generating $N = 30000$ values of $\mathbb{E}[\tilde{S}_\ell(t) \,|\, \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{T}]$ (see Corollary 5) through the Gibbs sampler of Section 6.1, besides the respective credible intervals have been approximated evaluating the quantiles of the corresponding vector of outputs. On the other hand, as for Figure 1(b), we have implemented the conditional sampler depicted in Section 6.2 and we have estimated the posterior distribution (38) of $\tilde{S}_1$ to derive both the credible intervals and the estimator (posterior mean). Analogously Figures 2–3 compare the two survival functions for the second and third sample, respectively. First of all we observe that both the BNP marginal and the BNP conditional method outperform with respect to the frequentist approach, providing estimates which are closer to the truth. This is also emphasized by the Kolmogorov distances $d_K(S_\ell, \tilde{S}_\ell)$ between the true survival curve and the estimated one reported in Table 1 for all $\ell = 1, 2, 3$ and for the different kinds of estimators. Second, it is apparent that the point estimates are similar with both the marginal and conditional procedures, but the marginal method underestimates the credible intervals: indeed in such a situation the infinite dimensional random elements are integrated out. We finally note that the conditional algorithm is able to provide more reliable credible intervals which always contain the true survival functions, but as for the traditional marginal procedure this does not happen (see Figure 3).

Acknowledgments

F. Camerlenghi gratefully acknowledges funding from the European Research Council (ERC), grant No 817257, and from the Italian Ministry of Education, University and Research (MIUR), "Dipartimenti di Eccellenza" grant 2018-2022. A. Lijoi and I. Prünster are supported by MIUR, PRIN Project 2015SNS29B.

A Supplementary material

A.1 Proofs of Section 3

A.1.1 Proof of Theorem 1

We start by proving statement (i) and first show that (12) holds true if and only if

$$\lim_{t\to\infty} \mathbb{E}[\tilde{S}_\ell(t)] = 0. \tag{43}$$

In fact, an application of the dominated convergence theorem yields

$$\lim_{t\to\infty} \mathbb{E}[\tilde{S}_\ell(t)] = \lim_{t\to\infty} \mathbb{E}\Big[\exp\Big(-\int_0^t \tilde{h}_\ell(s)\mathrm{d}s\Big)\Big] = \mathbb{E}\Big[\lim_{t\to\infty} \exp\Big(-\int_0^t \tilde{h}_\ell(s)\mathrm{d}s\Big)\Big] = \mathbb{E}[\lim_{t\to\infty} \tilde{S}_\ell(t)].$$

Therefore, (43) follows, being $\tilde{S}_\ell(t) \geq 0$ with probability 1.

Hence, to prove the validity of (12), it is enough to show that

$$\lim_{t\to\infty} \mathbb{E}[\tilde{S}_\ell(t)] = 0.$$

The computation of the expected value of $\tilde{S}_\ell(t)$ is then as follows

$$\begin{aligned}\mathbb{E}[\tilde{S}_\ell(t)] &= \mathbb{E}\Big[\exp\Big(-\int_{\mathbb{Y}} K_t^{(\ell)}(y)\tilde{\mu}_\ell(\mathrm{d}y)\Big)\Big] = \mathbb{E}\Big[\mathbb{E}\Big[\exp\Big(-\int_{\mathbb{Y}} K_t^{(\ell)}(y)\tilde{\mu}_\ell(\mathrm{d}y)\Big)\Big|\tilde{\mu}_0\Big]\Big]\\ &= \mathbb{E}\Big[\exp\Big(-\int_{\mathbb{Y}} \psi^{(\ell)}(K_t^{(\ell)}(y))\tilde{\mu}_0(\mathrm{d}y)\Big)\Big] = \exp\Big(-\int_{\mathbb{Y}} \psi^{(0)}(\psi^{(\ell)}(K_t^{(\ell)}(y)))c_0P_0(\mathrm{d}y)\Big),\end{aligned}$$

where we have exploited the hierarchical structure of $\tilde{\mu}_\ell$.
Consequently $\lim_{t\to\infty} \mathbb{E}[\tilde{S}_\ell(t)] = 0$ is equivalent to

$$\lim_{t\to\infty} \int_{\mathbb{Y}} \psi^{(0)}(\psi^{(\ell)}(K_t^{(\ell)}(y)))P_0(\mathrm{d}y) = +\infty.$$

It is then easily seen that

$$\begin{aligned}&\lim_{t\to\infty} \int_{\mathbb{Y}} \psi^{(0)}(\psi^{(\ell)}(K_t^{(\ell)}(y)))P_0(\mathrm{d}y)\\ &\quad= \lim_{t\to\infty} \int_{\mathbb{Y}}\int_0^\infty \Big(1-\exp\Big(-s\int_0^\infty (1-e^{-uK_t^{(\ell)}(y)})\rho_\ell(u)\mathrm{d}u\Big)\Big)\rho_0(s)\mathrm{d}sP_0(\mathrm{d}y).\end{aligned}$$

Since $K_t^{(\ell)}(y)$ increases as t grows to infinity, the monotone convergence theorem is applied thrice to obtain

$$\begin{aligned}&\lim_{t\to\infty} \int_{\mathbb{Y}} \psi^{(0)}(\psi^{(\ell)}(K_t^{(\ell)}(y)))P_0(\mathrm{d}y)\\ &\quad= \int_{\mathbb{Y}}\int_0^\infty \Big(1-\exp\Big(-s\int_0^\infty (1-e^{-u\lim_{t\to\infty} K_t^{(\ell)}(y)})\rho_\ell(u)\mathrm{d}u\Big)\Big)\rho_0(s)\mathrm{d}sP_0(\mathrm{d}y).\end{aligned} \tag{44}$$

Under the stated conditions on the Lévy intensities and the kernel, the previous expression clearly diverges. Therefore (43) and consequently also (12) are proven.

Now we prove statement (ii) concerning the covariance structure. By definition we have

$$Cov(\tilde{S}_{\ell_1}(t_1), \tilde{S}_{\ell_2}(t_2)) = \mathbb{E}[\tilde{S}_{\ell_1}(t_1)\tilde{S}_{\ell_2}(t_2)] - \mathbb{E}[\tilde{S}_{\ell_1}(t_1)]\mathbb{E}[\tilde{S}_{\ell_2}(t_2)]. \tag{45}$$

Proceeding along the same lines as before, we find

$$\mathbb{E}[\tilde{S}_{\ell_i}(t_i)] = \exp\Big(- \int_{\mathbb{Y}} \psi^{(0)}(\psi^{(\ell_i)}(K_{t_i}^{(\ell_i)}(y)))c_0 P_0(\mathrm{d}y)\Big)$$

for each $i = 1, 2$. An analogous computation leads to determine the first expectation in (45) as

$$\mathbb{E}[\tilde{S}_{\ell_1}(t_1)\tilde{S}_{\ell_2}(t_2)] = \exp\Big(- \int_{\mathbb{Y}} \psi^{(0)}\Big(\sum_{i=1}^{2} \psi^{(\ell_i)}(K_{t_i}^{(\ell_i)}(y))\Big)c_0 P_0(\mathrm{d}y)\Big).$$

The desired expression in (13) then follows. In order to show non-negativity, i.e. $Cov(\tilde{S}_{\ell_1}(t_1), \tilde{S}_{\ell_2}(t_2)) \geq 0$, one has

$$\begin{aligned}&Cov(\tilde{S}_{\ell_1}(t_1), \tilde{S}_{\ell_2}(t_2))\\ &= \frac{\exp\Big(- \int_{\mathbb{Y}} \psi^{(0)}\Big(\sum_{i=1}^{2} \psi^{(\ell_i)}(K_{t_i}^{(\ell_i)}(y))\Big) + \sum_{i=1}^{2} \psi^{(0)}(\psi^{(\ell_i)}(K_{t_i}^{(\ell_i)}(y)))c_0 P_0(\mathrm{d}y)\Big) - 1}{\exp\Big(\sum_{i=1}^{2}\int_{\mathbb{Y}} \psi^{(0)}(\psi^{(\ell_i)}(K_{t_i}^{(\ell_i)}(y))c_0 P_0(\mathrm{d}y)\Big)}.\end{aligned}$$

The previous expression can then be rewritten as

$$Cov(\tilde{S}_{\ell_1}(t_1), \tilde{S}_{\ell_2}(t_2)) = \frac{\exp\Big(\int_{\mathbb{Y}}\int_0^\infty \prod_{i=1}^{2}(1 - e^{-s\psi^{(\ell_i)}(K_{t_i}^{(\ell_i)}(y))})\rho_0(s)\mathrm{d}s P_0(\mathrm{d}y)\Big) - 1}{\exp\Big(\sum_{i=1}^{2}\int_{\mathbb{Y}} \psi^{(0)}(\psi^{(\ell_i)}(K_{t_i}^{(\ell_i)}(y))c_0 P_0(\mathrm{d}y)\Big)} \geq 0,$$

since

$$\int_{\mathbb{Y}}\int_0^\infty \prod_{i=1}^{2}(1 - e^{-s\psi^{(\ell_i)}(K_{t_i}^{(\ell_i)}(y))})\rho_0(s)\mathrm{d}s P_0(\mathrm{d}y) \geq 0.$$

□

A.1.2 Details for the determination of (14) and (15)

Let us first prove (14) in Example 3. Since $\tilde{\mu}_\ell$ is a gamma CRM, i.e. $\rho_\ell(s) = e^{-s}/s$ for $\ell = 0, \dots, d$, one has $\psi^{(\ell)}(u) = \log(1 + u)$ for any $u > 0$ and $\ell = 0, \dots, d$. Hence, we are able to obtain

$$\begin{aligned}\int_{\mathbb{Y}} \psi^{(0)}\Big(\sum_{i=1}^{2} \psi^{(\ell_i)}(K_{t_i}^{(\ell_i)}(y))\Big)\, P_0(\mathrm{d}y) &= \int_0^\infty \log\left(1 + \sum_{i=1}^{2} \log\left(1 + \alpha\,(t_i - y)\,\mathbb{1}_{(0,t_i]}(y)\right)\right) P_0(\mathrm{d}y)\\ &= \int_0^{t_1} \log\left(1 + \sum_{i=1}^{2} \log\left(1 + \alpha\,(t_i - y)\right)\right)\, P_0(\mathrm{d}y)\\ &\quad + \int_{t_1}^{t_2} \log(1 + \log(1 + \alpha\,(t_2 - y)))\, P_0(\mathrm{d}y).\end{aligned}$$

In a similar fashion, one shows that

$$\begin{aligned}\sum_{i=1}^{2}\int_{\mathbb{Y}} \psi^{(0)}(\psi^{(\ell_i)}(K_{t_i}^{(\ell_i)}(y)))\, P_0(\mathrm{d}y) &= \sum_{i=1}^{2}\int_0^{t_1} \log(1 + \log(1 + \alpha\,(t_i - y)))\, P_0(\mathrm{d}y)\\ &\quad + \int_{t_1}^{t_2} \log(1 + \log(1 + \alpha\,(t_2 - y)))\, P_0(\mathrm{d}y)\end{aligned}$$

whenever $t_1 \le t_2$. Therefore the covariance in (13) reduces to (14) as desired.

Now we show that (15) in Example 4 holds. Here we deal with generalized gamma CRMs i.e. $\rho_\ell(s) = e^{-s}s^{-1-\sigma}/\Gamma(1-\sigma)$ for $\ell = 1, \ldots, d$ and $\tilde\mu_0$ is also generalized gamma with parameter σ_0. In this setup, one has $\psi^{(0)}(u) = ((u+1)^{\sigma_0} - 1)/\sigma_0$ and $\psi^{(\ell)}(u) = ((u+1)^{\sigma} - 1)/\sigma$. Moreover, one can derive

$$\int_{\mathbb{Y}} \psi^{(0)}\Big(\sum_{i=1}^{2} \psi^{(\ell_i)}(K_{t_i}^{(\ell_i)}(y))\Big)\, P_0(\mathrm{d}y)$$
$$= \frac{1}{\sigma_0}\int_0^{t_1}\Big[\Big(\sum_{i=1}^{2}[(1+\alpha(t_i-y))^{\sigma}-1]/\sigma+1\Big)^{\sigma_0}-1\Big]\,P_0(\mathrm{d}y)$$
$$+\frac{1}{\sigma_0}\int_{t_1}^{t_2}\Big[\Big([(1+\alpha(t_2-y))^{\sigma}-1]/\sigma+1\Big)^{\sigma_0}-1\Big]\,P_0(\mathrm{d}y)$$

and

$$\sum_{i=1}^{2}\int_{\mathbb{Y}} \psi^{(0)}(\psi^{(\ell_i)}(K_{t_i}^{(\ell_i)}(y)))\, P_0(\mathrm{d}y)$$
$$= \frac{1}{\sigma_0}\sum_{i=1}^{2}\int_0^{t_1}\Big[\Big([(1+\alpha(t_i-y))^{\sigma}-1]/\sigma+1\Big)^{\sigma_0}-1\Big]\,P_0(\mathrm{d}y)$$
$$+\frac{1}{\sigma_0}\int_{t_1}^{t_2}\Big[\Big([(1+\alpha(t_2-y))^{\sigma}-1]/\sigma+1\Big)-1\Big]\,P_0(\mathrm{d}y).$$

By resorting to (13) one then obtains (15). □

A.2 Proofs of Section 4

A.2.1 Proof of Theorem 2

Our proof is divided in two main parts. First of all we reduce the study of consistency of the whole posterior distribution $\mathbb{P}((\tilde f_1, \ldots, \tilde f_d) \in \cdot\,|\boldsymbol{X}_1, \ldots, \boldsymbol{X}_d)$ to the study of consistency of all the marginals $\mathbb{P}(\tilde f_\ell \in \cdot\,|\boldsymbol{X}_1, \ldots, \boldsymbol{X}_d)$ as $\ell = 1, \ldots, d$. Secondly we prove that if the d posteriors $\mathbb{P}(\tilde f_\ell \in \cdot\,|\boldsymbol{X}_\ell)$ are consistent at the respective $f_\ell^{(0)}$, for any $\ell = 1, \ldots, d$, as in the assumptions of the theorem, then also $\mathbb{P}(\tilde f_\ell \in \cdot\,|\boldsymbol{X}_1, \ldots, \boldsymbol{X}_d)$ are consistent at $f_\ell^{(0)}$, for any $\ell = 1, \ldots, d$.
It should be first stressed that if

$$\Pi_{\boldsymbol{n}}(\mathsf{F}_{\mathbb{R}} \times \cdots \times A_\ell(f_\ell^{(0)}) \times \cdots \times \mathsf{F}_{\mathbb{R}}\,|\boldsymbol{X}_1, \ldots, \boldsymbol{X}_d) \to 1 \qquad \text{a.s.–}P^{(0),\infty} \tag{46}$$

for any weak neighbourhood $A_\ell(f_\ell^{(0)})$ of the true density $f_\ell^{(0)}$, and for any $\ell = 1, \ldots, d$, then (17) holds true. To this end, it is enough to note that (46) implies

$$\mathbb{P}((\tilde f_1, \ldots, \tilde f_d) \in A_1(f_1^{(0)}) \times \ldots \times A_d(f_d^{(0)})|\boldsymbol{X}_1, \ldots, \boldsymbol{X}_d) \to 1$$

as $\boldsymbol{n} := (n_1, \ldots, n_d) \to \infty$, $P^{(0),\infty}$ – a.s.. Indeed,

$$\mathbb{P}((\tilde f_1, \ldots, \tilde f_d) \in A_1(f_1^{(0)}) \times \ldots \times A_d(f_d^{(0)})|\boldsymbol{X}_1, \ldots, \boldsymbol{X}_d)$$

$$= \mathbb{P}\Big(\bigcap_{\ell=1}^{d} \Big\{ \tilde{f}_\ell \in A_\ell(f_\ell^{(0)}) \Big\} | \boldsymbol{X}_1, \dots, \boldsymbol{X}_d \Big) = 1 - \mathbb{P}\Big(\bigcup_{\ell=1}^{d} \Big\{ \tilde{f}_\ell \in (A_\ell(f_\ell^{(0)}))^c \Big\} | \boldsymbol{X}_1, \dots, \boldsymbol{X}_d \Big)$$

$$\geq 1 - \sum_{\ell=1}^{d} \mathbb{P}\Big(\tilde{f}_\ell \in (A_\ell(f_\ell^{(0)}))^c | \boldsymbol{X}_1, \dots, \boldsymbol{X}_d \Big)$$

and the right–hand side converges to 1 as a consequence of (46).

As remarked at the beginning of the proof, we now need to show that (18) entails (46). To fix the notation we will define the following σ–fields

$$\mathscr{F}_{\ell,n_\ell} = \sigma(\boldsymbol{X}_\ell) = \sigma(X_{\ell,1}, \dots, X_{\ell,n_\ell})$$

the σ–field generated by $\boldsymbol{X}_\ell$, and by

$$\mathscr{F}_{\ell,\infty} = \sigma(X_{\ell,1}, X_{\ell,2}, \dots)$$

the σ–field generated by the whole sequence of random variables of group ℓ. Analogously, define

$$\mathscr{G}_{n_1,\dots,n_d} = \sigma(\boldsymbol{X}_1, \dots, \boldsymbol{X}_d)$$

and the limit σ–field as $n_1, \dots, n_d \to \infty$

$$\mathscr{G}_\infty = \sigma((X_{\ell,1}, X_{\ell,2} \dots);\ \ell = 1, \dots, d).$$

It is clear that, as $n_\ell \uparrow \infty$, then the sequence $\mathscr{F}_{\ell,n_\ell}$ is increasing and we have $\mathscr{F}_{\ell,n_\ell} \uparrow \mathscr{F}_{\ell,\infty}$; besides $\mathscr{G}_{n_1,\dots,n_d} \uparrow \mathscr{G}_\infty$ as $n_1, \dots, n_d \to \infty$. By [3, Theorem 35.6] we have that

$$\mathbb{P}(\tilde{f}_\ell \in A_\ell^c(f_\ell^{(0)}) | \mathscr{F}_{\ell,n_\ell}) \to \mathbb{P}(\tilde{f}_\ell \in A_\ell^c(f_\ell^{(0)}) | \mathscr{F}_{\ell,\infty}) \tag{47}$$

$P_\ell^{(0),\infty}$–almost surely as $n_\ell \to \infty$. Therefore, since the almost sure limit is unique, by (18) and (47), we get that

$$\mathbb{P}(\tilde{f}_\ell \in A_\ell^c(f_\ell^{(0)}) | \mathscr{F}_{\ell,\infty}) = 0 \tag{48}$$

$P_\ell^{(0),\infty}$–almost surely. Applying again [3, Theorem 35.6], we obtain

$$\mathbb{P}(\tilde{f}_\ell \in A_\ell^c(f_\ell^{(0)}) | \mathscr{G}_{n_1,\dots,n_d}) \to \mathbb{P}(\tilde{f}_\ell \in A_\ell^c(f_\ell^{(0)}) | \mathscr{G}_\infty) \tag{49}$$

as $n_1, \dots, n_d \to \infty$.
Therefore the limit of the conditional probability of interest $\mathbb{P}(\tilde{f}_\ell \in A_\ell^c(f_\ell^{(0)}) | \mathscr{G}_{n_1,\dots,n_d}) = \mathbb{P}(\tilde{f}_\ell \in A_\ell^c(f_\ell^{(0)}) | \boldsymbol{X}_1, \dots, \boldsymbol{X}_d)$ exists and it equals $\mathbb{P}(\tilde{f}_\ell \in A_\ell^c(f_\ell^{(0)}) | \mathscr{G}_\infty)$. It remains to prove that the latter term equals zero almost surely. To this end note that the properties of conditional expectations and (48) imply

$$0 = \mathbb{P}(\tilde{f}_\ell \in A_\ell^c(f_\ell^{(0)}) | \mathscr{F}_{\ell,\infty}) = \mathbb{E}[\mathbb{P}(\tilde{f}_\ell \in A_\ell^c(f_\ell^{(0)}) | \mathscr{G}_\infty) | \mathscr{F}_{\ell,\infty}]$$

almost surely. Hence since $\mathbb{P}(\tilde{f}_\ell \in A_\ell^c(f_\ell^{(0)}) | \mathscr{G}_\infty) \geq 0$, unless in a set of measure 0, we get that

$$\mathbb{P}(\tilde{f}_\ell \in A_\ell^c(f_\ell^{(0)}) | \mathscr{G}_\infty) = 0$$

$P_\ell^{(0),\infty}$–almost surely and our result follows. □

A.2.2 Proof of Proposition 1

Let k be the kernel by Dykstra and Laud, i.e. $k(t;y) = \alpha \mathbb{1}_{(0,t]}(y)$, then the random hazard rate $\tilde{h}_\ell(t)$ equals $\alpha \tilde{\mu}_\ell(0,t]$. Now fix $u > 0$ and evaluate the Laplace functional of $\tilde{\mu}_\ell(0,t]$:

$$\begin{aligned}\mathbb{E}e^{-u\tilde{\mu}_\ell(0,t]} &= \exp\left\{-c_0 P_0(0,t)\psi^{(0)}\Big(\frac{(u+1)^\sigma - 1}{\sigma}\Big)\right\} \\ &= \exp\left\{-\frac{c_0 P_0(0,t)}{\sigma_0}\left[\Big(\frac{(u+1)^\sigma - 1}{\sigma} + 1\Big)^{\sigma_0} - 1\right]\right\}.\end{aligned}$$

Therefore for any $u > 0$ and $r > 0$, we have

$$\mathbb{E}e^{-u\tilde{\mu}_\ell(0,t]/t^r} = \exp\left\{-\frac{c_0 P_0(0,t)}{\sigma_0}\left[\Big(\frac{(u/t^r + 1)^\sigma - 1}{\sigma} + 1\Big)^{\sigma_0} - 1\right]\right\}$$

and, since, as $t \to 0$, $P_0(0,t)$ tends to zero at the same speed as $t^{r'}$ does, there exists some $r > 0$ such that

$$\lim_{t\to 0} \mathbb{E}e^{-u\tilde{\mu}_\ell(0,t]/t^r} = 0$$

for some $r > 0$. Next, by the Fatou lemma

$$\mathbb{E}\left[\liminf_{t\to 0} \mathrm{e}^{-u\tilde{\mu}_\ell(0,t]/t^r}\right] \leq \liminf_{t\to 0} \mathbb{E}\mathrm{e}^{-u\tilde{\mu}_\ell(0,t]/t^r} = 0$$

which also implies that

$$\mathbb{E}\left[\liminf_{t\to 0} \mathrm{e}^{-u\tilde{\mu}_\ell(0,t]/t^r}\right] = 0$$

and from this one deduces that

$$\liminf_{t\to 0} \mathrm{e}^{-u\tilde{\mu}_\ell(0,t]/t^r} = 0$$

almost surely. Hence, condition ii) of Theorem 3 is satisfied. □

A.2.3 Support of hierarchical CRMs

In this section we consider a hierarchical CRM of the type

$$\begin{aligned}\tilde{\mu} \,|\, \tilde{\mu}_0 &\sim \mathrm{CRM}(\tilde{\nu}) \\ \tilde{\mu}_0 &\sim \mathrm{CRM}(\nu_0),\end{aligned}$$

where $\tilde{\nu}(\mathrm{d}s, \mathrm{d}y) = \rho(s)\,\mathrm{d}s\,\tilde{\mu}_0(\mathrm{d}y)$ and $\nu_0(\mathrm{d}s, \mathrm{d}y) = \rho_0(s)\,\mathrm{d}s\,c_0 P_0(\mathrm{d}y)$. We suppose that the following assumptions are fulfilled:

(H1) $\rho_0 > 0$ and $\rho > 0$ on $\mathbb{R}^+$,

(H2) ρ and ρ_0 have infinite integrals on $\mathbb{R}^+$, i.e.

$$\int_{\mathbb{R}^+} \rho(s)\mathrm{d}s = \int_{\mathbb{R}^+} \rho_0(s)\mathrm{d}s = +\infty,$$

(H3) ρ_0 has finite mean and second moment, namely $\int_0^\infty s\,\rho_0(s)\,\mathrm{d}s < \infty$ and $\int_0^\infty s^2\,\rho_0(s)\,\mathrm{d}s < \infty$.

We study the support of the hierarchical CRM, proving a result similar to [10, Theorem 2], which is used in the proof of Theorem 4 to show the consistency of marginal densities $\tilde{f}_\ell$ for each $\ell \in \{1, \ldots, d\}$.

Proposition 3. *Let $\tilde{\mu}|\tilde{\mu}_0 \sim \mathrm{CRM}(\tilde{\nu})$ and $\tilde{\mu}_0 \sim \mathrm{CRM}(\nu_0)$. If (H1)–(H3) hold true for ρ and ρ_0 and the support of P_0 is the interval $[0,T]$, for some $T>0$, then $\tilde{\mu}(0,t]$ is supported by the whole positive real line, for any $t \le T$.*

Proof. The proof goes along similar lines as the proof of [15, Proposition 1]: we need to show that for any $0<a<b$, $\mathbb{P}(\tilde{\mu}(0,t] \in (a,b)) > 0$. We denote by $\tilde{\mu}_t := \tilde{\mu}(0,t]$ and $\tilde{\mu}_t(\delta)$ the sum of all the jumps in $(0,t]$ of size less or equal than δ. Since $\tilde{\mu}$ is a CRM it can be represented as a functional of a Poisson process $\tilde{N}$ on $\mathbb{R}^+ \times [0,T]$, indeed

$$\tilde{\mu}(A) = \int_{\mathbb{R}^+ \times A} s\tilde{N}(\mathrm{d}s, \mathrm{d}x),$$

and the same holds true for $\tilde{\mu}_0$, in particular we denote by $\tilde{N}_0$ the respective Poisson point process on $\mathbb{R}^+ \times [0,T]$. Observe that

$$\tilde{\mu}_t(\delta) = \int_{(0,\delta)\times(0,t]} s\tilde{N}(\mathrm{d}s, \mathrm{d}x)$$

and its expected value is

$$\begin{aligned}
\mathbb{E}\tilde{\mu}_t(\delta) &= \mathbb{E}\mathbb{E}\left[\int_{(0,\delta)\times(0,t]} s\tilde{N}(\mathrm{d}s,\mathrm{d}x)\Big|\tilde{\mu}_0\right] \\
&= \mathbb{E}\left[\int_0^\delta s\rho(s)\mathrm{d}s\tilde{\mu}_0(0,t)\right] = \int_0^\delta s\rho(s)\mathrm{d}s\mathbb{E}\int_{\mathbb{R}^+\times(0,t)} s\tilde{N}_0(\mathrm{d}s,\mathrm{d}x) \\
&= c_0 P_0(0,t)\int_0^\delta s\rho(s)\mathrm{d}s\int_0^\infty s\rho_0(s)\mathrm{d}s
\end{aligned}$$

where we have repeatedly applied the Campbell Theorem. Hence, $\mathbb{E}\tilde{\mu}_t(\delta) < \infty$, for any $0 < t \le T$ and $\delta > 0$, and an argument similar to the one used in the proof of [15, Proposition 1] can be employed to show that for any $c>0$ there exists $\delta_c > 0$ such that $\mathbb{P}(\tilde{\mu}_t(\delta_c) < c) > 0$. Now set $c = (b-a)/k$, where k is such that $(b-a)/k < a$, and choose $\delta > 0$ such that the latter condition on $\tilde{\mu}_t(\delta)$ holds true. One, then, defines the following sets

$$\begin{aligned}
A_1 &:= \left\{\tilde{N}((a,(a+b)/2)\times(0,t]) = 1\right\} \\
A_2 &:= \left\{\tilde{N}((\delta,a)\times(0,t]) = 0\right\} \\
A_3 &:= \{\tilde{\mu}_t(\delta) < c\} \\
A_4 &:= \left\{\tilde{N}(((a+b)/2,+\infty)\times(0,t]) = 0\right\}
\end{aligned}$$

and notes that $A_1 \cap A_2 \cap A_3 \cap A_4$ is the event "$\tilde{\mu}$ has exactly one jump in $(0,t]$ of size in $(a,(a+b)/2)$ and all the other jumps have size less than δ with sum at most $c = (b-a)/k$", hence $\mathbb{P}(\tilde{\mu}(0,t] \in (a,b)) \ge \mathbb{P}(A_1 \cap A_2 \cap A_3 \cap A_4)$. We prove that this probability is strictly positive by a direct calculation. Using the conditional independence of the events A_i and the fact that $\tilde{N}(C \times (0,t))|\tilde{\mu}_0$ is, for any $C \in \mathscr{B}(\mathbb{R}^+)$, a Poisson random variable with mean $\tilde{\mu}_0(0,t)\int_C \rho(s)\mathrm{d}s$, we get

$$\begin{aligned}
\mathbb{P}(A_1 \cap A_2 \cap A_3 \cap A_4) &= \mathbb{E}\left[\prod_{i=1}^4 \mathbb{P}(A_i|\tilde{\mu}_0)\right] \\
&= \int_a^{(a+b)/2} \rho(s)\mathrm{d}s\,\mathbb{E}\left[\tilde{\mu}_0(0,t)e^{-\tilde{\mu}_0(0,t)\int_\delta^\infty \rho(s)\mathrm{d}s}\mathbb{P}(\tilde{\mu}_t(\delta) < c|\tilde{\mu}_0)\right]
\end{aligned}$$

$$\geq \int_a^{(a+b)/2} \rho(s)\mathrm{d}s\, \mathbb{E}\left[\tilde{\mu}_0(0,t)e^{-\tilde{\mu}_0(0,t)\int_\delta^\infty \rho(s)\mathrm{d}s}\left(1-\frac{\mathbb{E}[\tilde{\mu}_t(\delta)|\tilde{\mu}_0]}{c}\right)\right]$$
$$= \int_a^{(a+b)/2} \rho(s)\mathrm{d}s\, \mathbb{E}\left[\tilde{\mu}_0(0,t)e^{-\tilde{\mu}_0(0,t)\int_\delta^\infty \rho(s)\mathrm{d}s} \times \left(1-\frac{\tilde{\mu}_0(0,t)}{c}\int_0^\delta s\,\rho(s)\mathrm{d}s\right)\right] \tag{50}$$

where we have applied the Markov inequality and the last equality follows again from the Campbell Theorem. Now define

$$u(\delta) := \int_\delta^\infty \rho(s)\mathrm{d}s, \tag{51}$$

simple calculations show that for any $A \subset [0,T]$ and $w > 0$ one has

$$\begin{aligned}\mathbb{E}e^{-w\tilde{\mu}_0(A)}\tilde{\mu}_0(A) &= -\frac{\mathrm{d}}{\mathrm{d}w}\mathbb{E}e^{-w\tilde{\mu}_0(A)} \\ &= c_0P_0(A)\int_0^\infty e^{-sw}s\rho_0(s)\mathrm{d}s\exp\left\{-c_0P_0(A)\int_0^\infty (1-e^{-sw})\rho_0(s)\mathrm{d}s\right\}\end{aligned} \tag{52}$$

and analogously

$$\begin{aligned}\mathbb{E}e^{-w\tilde{\mu}_0(A)}\tilde{\mu}_0^2(A) &= \frac{\mathrm{d}^2}{\mathrm{d}w^2}\mathbb{E}e^{-w\tilde{\mu}_0(A)} \\ &= \left\{c_0P_0(A)\Big(\int_0^\infty e^{-sw}s\rho_0(s)\mathrm{d}s\Big)^2 + \int_0^\infty s^2e^{-sw}\rho_0(s)\mathrm{d}s\right\}. \\ &\quad \times c_0P_0(A)\exp\left\{-c_0P_0(A)\int_0^\infty (1-e^{-sw})\rho_0(s)\mathrm{d}s\right\}\end{aligned} \tag{53}$$

We use equations (52)–(53), with $u(\delta)$ in place of w, to evaluate the expected value in (50):

$$\begin{aligned}&\mathbb{P}(A_1\cap A_2\cap A_3\cap A_4) \\ &\geq \int_a^{(a+b)/2}\rho(s)\mathrm{d}s\cdot\exp\left\{-c_0P_0(0,t)\int_0^\infty(1-e^{-su(\delta)})\rho_0(s)\mathrm{d}s\right\} \\ &\times c_0P_0(0,t)\left\{\int_0^\infty se^{-su(\delta)}\rho_0(s)\mathrm{d}s - \frac{1}{c}\int_0^\delta s\rho(s)\mathrm{d}s\right. \\ &\quad\left.\times\left[c_0P_0(0,t)\left(\int_0^\infty se^{-u(\delta)s}\rho_0(s)\mathrm{d}s\right)^2 + \int_0^\infty s^2e^{-su(\delta)}\rho_0(s)\mathrm{d}s\right]\right\}.\end{aligned}$$

In order to simplify the notation we define

$$\begin{aligned}C(\delta) &:= \exp\left\{-c_0P_0(0,t)\int_0^\infty(1-e^{-su(\delta)})\rho_0(s)\mathrm{d}s\right\}c_0P_0(0,t)\int_a^{(a+b)/2}\rho(s)\mathrm{d}s \\ I_1(\delta) &:= \int_0^\infty se^{-u(\delta)s}\rho_0(s)\mathrm{d}s \\ I_2(\delta) &:= \int_0^\infty s^2e^{-su(\delta)}\rho_0(s)\mathrm{d}s\end{aligned}$$

it is also useful to observe that

$$I_1(\delta) \geq \int_0^1 s e^{-u(\delta)s} \rho_0(s) \mathrm{d}s \geq e^{-u(\delta)} \int_0^1 s \rho_0(s) \mathrm{d}s > 0 \tag{54}$$

and besides

$$\begin{aligned} I_2(\delta) &= \int_0^1 s^2 e^{-su(\delta)} \rho_0(s) \mathrm{d}s + \int_1^\infty s^2 e^{-su(\delta)} \rho_0(s) \mathrm{d}s \\ &\leq \int_0^\infty s e^{-su(\delta)} \rho_0(s) \mathrm{d}s + e^{-u(\delta)} \int_1^\infty s^2 \rho_0(s) \mathrm{d}s \\ &= I_1(\delta) + e^{-u(\delta)} \int_1^\infty s^2 \rho_0(s) \mathrm{d}s. \end{aligned} \tag{55}$$

We now use (54)–(55) to lower bound the probability of interest:

$$\begin{aligned} \mathbb{P}(A_1 \cap A_2 \cap A_3 \cap A_4) &\geq C(\delta) \left\{ I_1(\delta) - \frac{1}{c} \int_0^\delta s \rho(s) \mathrm{d}s (I_1^2(\delta) c_0 P_0(0,t) + I_2(\delta)) \right\} \\ &\geq C(\delta) \Bigg\{ I_1(\delta) - \frac{1}{c} \int_0^\delta s \rho(s) \mathrm{d}s \\ &\qquad \times \left(I_1^2(\delta) c_0 P_0(0,t) + I_1(\delta) + e^{-u(\delta)} \int_1^\infty s^2 \rho_0(s) \mathrm{d}s \right) \Bigg\} \\ &= C(\delta) \Bigg\{ I_1(\delta) \Big(1 - \frac{1}{c} \int_0^\delta s \rho(s) \mathrm{d}s \, (I_1(\delta) c_0 P_0(0,t) + 1) \Big) \\ &\qquad - \frac{e^{-u(\delta)}}{c} \int_1^\infty s^2 \rho_0(s) \mathrm{d}s \int_0^\delta s \rho(s) \mathrm{d}s \Bigg\}. \end{aligned}$$

Since $u(\delta) \nearrow +\infty$ as $\delta \searrow 0$ then $I_1(\delta)$ is a decreasing function of δ, as well as the integral $\int_0^\delta s \rho(s) \mathrm{d}s$, and both these terms go to zero as δ goes to 0, as a consequence we can choose δ sufficiently small in such a way that

$$\Big(1 - \frac{1}{c} \int_0^\delta s \rho(s) \mathrm{d}s \, (I_1(\delta) c_0 P_0(0,t) + 1) \Big) > 0.$$

We now use the lower bound (54) to get

$$\begin{aligned} \mathbb{P}(A_1 \cap A_2 \cap A_3 \cap A_4) \geq C(\delta) e^{-u(\delta)} \Bigg\{ &\int_0^1 s \rho_0(s) \mathrm{d}s \Big(1 - \frac{1}{c} \int_0^\delta s \rho(s) \mathrm{d}s \, (I_1(\delta) c_0 P_0(0,t) + 1) \Big) \\ &- \frac{1}{c} \int_1^\infty s^2 \rho_0(s) \mathrm{d}s \int_0^\delta s \rho(s) \mathrm{d}s \Bigg\} \end{aligned}$$

which is strictly positive for a sufficiently small $\delta > 0$, indeed the negative term on the r.h.s. of the previous inequality can be made as small as required by reducing the value of δ, because it goes to 0 as $\delta \to 0$. □

A.2.4 Proof of Theorem 4

In the case of generalized gamma process and the Dykstra–Laud kernel the assumptions of Theorem 3 are met, therefore to prove consistency we need to check (19). If $h_\ell^{(0)}$ is non decreasing and

$h_\ell^{(0)}(0) = 0$, it defines a measure $\mu_\ell^{(0)}$ on $\mathbb{R}^+$ indeed $\mu_\ell^{(0)}((0,t]) = h_\ell^{(0)}(t)$. As a consequence condition (19) is equivalent to the following

$$\mathbb{P}\left(\sup_{s\le t} |\alpha\tilde{\mu}_\ell(0,s] - \mu_\ell^{(0)}(0,s]| < \eta\right) > 0$$

for any $t > 0$, $\eta > 0$. This can be proved along the same lines as [10, Lemma 11], using Proposition 3 instead of [15, Proposition 1] (concerning the support of a hierarchical CRM) and the independence properties of the hierarchical CRM $\tilde{\mu}_\ell$. □

A.3 Proofs of Section 5

A.3.1 Proof of Theorem 5

First of all, recall that (see for example [6, 22])

$$(-1)^r \frac{\mathrm{d}^r}{\mathrm{d}u^r} e^{-m\psi^{(\ell)}(u)} = e^{-m\psi^{(\ell)}(u)} \sum_{i=1}^r m^i \xi_{r,i}^{(\ell)}(u) \tag{56}$$

where

$$\xi_{r,i}^{(\ell)}(u) = \sum_{(*)} \frac{1}{i!}\binom{r}{q_1,\cdots,q_i} \tau_{q_1}^{(\ell)}(u)\cdots\tau_{q_i}^{(\ell)}(u)$$

and the sum runs over all the vectors $(q_1,\cdots,q_i)$ of positive integers such that $\sum_{j=1}^i q_i = r$. The joint distribution of $(\boldsymbol{X},\boldsymbol{Y})$ and of the partition Ψ_N is determined taking the expected value with respect to the distribution of the vector of CRMs:

$$\pi(\boldsymbol{X},\boldsymbol{Y}^*,\Psi_N) = \mathbb{E}[\mathscr{L}(\tilde{\boldsymbol{\mu}};\boldsymbol{X},\boldsymbol{Y})] = \mathbb{E}\Big[\prod_{\ell=1}^d \exp\Big(-\int_{\mathbb{Y}} K_\ell(y)\tilde{\mu}_\ell(\mathrm{d}y)\Big)\prod_{i=1}^{n_\ell} k(X_{\ell,i};Y_{\ell,i})\tilde{\mu}_\ell(\mathrm{d}Y_{\ell,i})\Big]$$

setting $Q(\boldsymbol{X},\boldsymbol{Y}) := \prod_{\ell=1}^d\prod_{i=1}^{n_\ell} k(X_{i,\ell};Y_{i,\ell})$, we get

$$\pi(\boldsymbol{X},\boldsymbol{Y}^*,\Psi_N) = Q(\boldsymbol{X},\boldsymbol{Y})\mathbb{E}\Big[\prod_{\ell=1}^d \exp\Big(-\int_{\mathbb{Y}} K_\ell(y)\tilde{\mu}_\ell(\mathrm{d}y)\Big)\prod_{i=1}^{n_\ell}\tilde{\mu}_\ell(\mathrm{d}Y_{\ell,i})\Big]$$

We can exploit the hierarchical structure of the model in order to get out the expectation:

$$\begin{aligned}\pi(\boldsymbol{X},\boldsymbol{Y}^*,\Psi_N) &= Q(\boldsymbol{X},\boldsymbol{Y})\mathbb{E}\Big[\mathbb{E}\Big[\prod_{\ell=1}^d \exp\Big(-\int_{\mathbb{Y}} K_\ell(y)\tilde{\mu}_\ell(\mathrm{d}y)\Big)\prod_{i=1}^{n_\ell}\tilde{\mu}_\ell(\mathrm{d}Y_{\ell,i})\Big|\tilde{\mu}_0\Big]\Big] \\ &= Q(\boldsymbol{X},\boldsymbol{Y})\mathbb{E}\Big[\prod_{\ell=1}^d \mathbb{E}\Big[\exp\Big(-\int_{\mathbb{Y}} K_\ell(y)\tilde{\mu}_\ell(\mathrm{d}y)\Big)\prod_{i=1}^{n_\ell}\tilde{\mu}_\ell(\mathrm{d}Y_{\ell,i})\Big|\tilde{\mu}_0\Big]\Big]\end{aligned}$$

conditional on $\tilde{\mu}_0$, $\tilde{\mu}_\ell$ is a completely random measure with known intensity, so we can compute the Laplace functional, in fact, denoting by $\mathbb{Y}^* := \mathbb{Y}\setminus\{Y_1^*,\cdots,Y_k^*\}$, we obtain:

$$\pi(\boldsymbol{X},\boldsymbol{Y}^*,\Psi_N) = Q(\boldsymbol{X},\boldsymbol{Y})\mathbb{E}\Big[\prod_{\ell=1}^d \exp\Big(-\int_{\mathbb{Y}^*}\psi^{(\ell)}(K_\ell(y))\tilde{\mu}_0(\mathrm{d}y)\Big)$$

$$
\times \mathbb{E}\Big[\prod_{j=1}^{k} e^{-K_\ell(Y_j^*)\tilde{\mu}_\ell(\mathrm{d}Y_j^*)}\tilde{\mu}_\ell^{n_{\ell,j}}(\mathrm{d}Y_j^*)\Big|\tilde{\mu}_0\Big]\Big]
$$

$$
= Q(\boldsymbol{X},\boldsymbol{Y})\mathbb{E}\Big[\prod_{\ell=1}^{d}\exp\Big(-\int_{\mathbb{Y}^*}\psi^{(\ell)}(K_\ell(y))\tilde{\mu}_0(\mathrm{d}y)\Big)
$$

$$
\times \prod_{j=1}^{k}\frac{\mathrm{d}^{n_{\ell,j}}}{\mathrm{d}u^{n_{\ell,j}}}(-1)^{n_{\ell,j}}e^{-\psi^{(\ell)}(u)\tilde{\mu}_0(\mathrm{d}Y_j^*)}\Big|_{u=K_\ell(Y_j^*)}\Big]
$$

$$
\overset{(56)}{=} Q(\boldsymbol{X},\boldsymbol{Y})\mathbb{E}\Big[\prod_{\ell=1}^{d}\exp\Big(-\int_{\mathbb{Y}^*}\psi^{(\ell)}(K_\ell(y))\tilde{\mu}_0(\mathrm{d}y)\Big)
$$

$$
\times \sum_{\boldsymbol{i}}\prod_{j=1}^{k}\xi^{(\ell)}_{n_{\ell,j},i_{\ell,j}}(K_\ell(Y_j^*))\tilde{\mu}_0^{i_{\ell,j}}(\mathrm{d}Y_j^*)\Big]
$$

where $i_{\ell,j}\in\{1,\cdots,n_{\ell,j}\}$, with $n_{\ell,j}\geq 0$ and the convention $\sum_{i_{\ell,j}=1}^{0}\equiv 1$. Introducing the notation $i_{\bullet j}:=\sum_{\ell=1}^{d} i_{\ell,j}$, a standard argument shows that:

$$
\pi(\boldsymbol{X},\boldsymbol{Y}^*,\Psi_N) = Q(\boldsymbol{X},\boldsymbol{Y})\sum_{\boldsymbol{i}}\prod_{\ell=1}^{d}\prod_{j=1}^{k}\xi^{(\ell)}_{n_{\ell,j},i_{\ell,j}}(K_\ell(Y_j^*))
$$

$$
\times \mathbb{E}\Big[\exp\Big(-\int_{\mathbb{Y}^*}\sum_{\ell=1}^{d}\psi^{(\ell)}(K_\ell(y))\tilde{\mu}_0(\mathrm{d}y)\Big)\prod_{j=1}^{k}\tilde{\mu}_0^{i_{\bullet j}}(\mathrm{d}Y_j^*)\Big]
$$

$$
= Q(\boldsymbol{X},\boldsymbol{Y})\sum_{\boldsymbol{i}} c_0^k\prod_{j=1}^{k}P_0(\mathrm{d}Y_j^*)\prod_{\ell=1}^{d}\prod_{j=1}^{k}\xi^{(\ell)}_{n_{\ell,j},i_{\ell,j}}(K_\ell(Y_j^*))
$$

$$
\times \exp\Big(-\int_{\mathbb{Y}}\psi^{(0)}\Big(\sum_{\ell=1}^{d}\psi^{(\ell)}(K_\ell(y))\Big)c_0P_0(\mathrm{d}y)\Big)
$$

$$
\times \prod_{j=1}^{k}\tau^{(0)}_{i_{\bullet j}}\Big(\sum_{\ell=1}^{d}\psi^{(\ell)}(K_\ell(Y_j^*))\Big) + o\Big(\prod_{j=1}^{k}P_0(\mathrm{d}Y_j^*)\Big)
$$

$$
= Q(\boldsymbol{X},\boldsymbol{Y})c_0^k\prod_{j=1}^{k}P_0(\mathrm{d}Y_j^*)\exp\Big(-\int_{\mathbb{Y}}\psi^{(0)}\Big(\sum_{\ell=1}^{d}\psi^{(\ell)}(K_\ell(y))\Big)c_0P_0(\mathrm{d}y)\Big)
$$

$$
\times \sum_{\boldsymbol{i}}\prod_{\ell=1}^{d}\prod_{j=1}^{k}\xi^{(\ell)}_{n_{\ell,j},i_{\ell,j}}(K_\ell(Y_j^*))\prod_{j=1}^{k}\tau^{(0)}_{i_{\bullet j}}\Big(\sum_{\ell=1}^{d}\psi^{(\ell)}(K_\ell(Y_j^*))\Big)
$$

$$
+ o\Big(\prod_{j=1}^{k}P_0(\mathrm{d}Y_j^*)\Big).
$$

Putting the expression of $\xi^{(\ell)}_{n_{\ell,j},i_{\ell,j}}$ in the previous equation, and ignoring the higher order terms, we finally obtain:

$$
\pi(\boldsymbol{X},\boldsymbol{Y}^*,\Psi_N) = Q(\boldsymbol{X},\boldsymbol{Y})c_0^k\prod_{j=1}^{k}P_0(\mathrm{d}Y_j^*)\exp\Big(-\int_{\mathbb{Y}}\psi^{(0)}\Big(\sum_{\ell=1}^{d}\psi^{(\ell)}(K_\ell(y))\Big)c_0P_0(\mathrm{d}y)\Big)
$$

$$\times \sum_{\boldsymbol{i}} \sum_{\boldsymbol{q}} \prod_{j=1}^{k} \tau^{(0)}_{i_{\bullet j}} \Big(\sum_{\ell=1}^{d} \psi^{(\ell)}(K_\ell(Y_j^*)) \Big)$$
$$\times \prod_{\ell=1}^{d} \prod_{j=1}^{k} \frac{1}{i_{\ell,j}!} \binom{n_{\ell,j}}{q_{\ell,j,1}, \cdots, q_{\ell,j,i_{\ell,j}}} \prod_{\ell=1}^{d} \prod_{j=1}^{k} \prod_{t=1}^{i_{\ell,j}} \tau^{(\ell)}_{q_{\ell,j,t}}(K_\ell(Y_j^*)).$$

where the sum with respect to $\boldsymbol{q}$ runs over all the vectors of positive integers such that $\sum_{t=1}^{i_{\ell,j}} q_{\ell,j,t} = n_{\ell,j}$. □

A.3.2 Proof of Theorem 6

In order to prove the theorem, one has to verify the validity of (28). The l.h.s. of (28), that is to say the conditional distribution of $(\boldsymbol{Y}^*, \boldsymbol{T}, \Psi_N)$ given $\boldsymbol{X}$, can be easily derived from (26) and it satisfies

$$\pi(\boldsymbol{Y}^*, \boldsymbol{T}, \Psi_N \mid \boldsymbol{X}) \propto c_0^k \prod_{\ell=1}^{d} \prod_{i=1}^{n_\ell} k(X_{\ell,i}; Y_{\ell,i}) \prod_{j=1}^{k} \tau^{(0)}_{i_{\bullet j}} \Big(\sum_{\ell=1}^{d} \psi^{(\ell)}(K_\ell(Y_j^*)) \Big) P_0(\mathrm{d}Y_j^*)$$
$$\times \prod_{\ell=1}^{d} \prod_{j=1}^{k} \prod_{t=1}^{i_{\ell,j}} \tau^{(\ell)}_{q_{\ell,j,t}}(K_\ell(Y_j^*)) P_0(\mathrm{d}T^*_{\ell,j,t}). \tag{57}$$

Marginalizing out m_0 and $\boldsymbol{m}$ in (29), one immediately realizes that the r.h.s. of (28) is proportional to (57) and this concludes the proof. □

A.3.3 Proof of Theorem 7

The posterior distribution of $(\tilde{\mu}_1, \cdots, \tilde{\mu}_d)$ is characterized by the Laplace functional, i.e.

$$\mathbb{E}[e^{-\tilde{\mu}_1(f_1) - \cdots - \tilde{\mu}_d(f_d)} | \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{T}] = \frac{\mathbb{E}[e^{-\tilde{\mu}_1(f_1) - \cdots - \tilde{\mu}_d(f_d)} \mathscr{L}(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{T})]}{\mathbb{E}[\mathscr{L}(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{T})]}$$

where $f_\ell : \mathbb{Y} \to \mathbb{R}^+$ and we have used the notation $\tilde{\mu}_\ell(f_\ell) = \int_{\mathbb{Y}} f_\ell(y) \tilde{\mu}_\ell(\mathrm{d}y)$. The denominator is the likelihood in (26), the numerator can be determined in a similar fashion, so the Laplace functional equals

$$\mathbb{E}[e^{-\tilde{\mu}_1(f_1) - \cdots - \tilde{\mu}_d(f_d)} | \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{T}] = \exp\Big(\int_{\mathbb{Y}} \psi^{(0)} \Big(\sum_{\ell=1}^{d} \psi^{(\ell)}(K_\ell(y)) \Big) c_0 P_0(\mathrm{d}y)$$
$$- \int_{\mathbb{Y}} \psi^{(0)} \Big(\sum_{\ell=1}^{d} \psi^{(\ell)}(K_\ell(y) + f_\ell(y)) \Big) c_0 P_0(\mathrm{d}y) \Big)$$
$$\times \frac{\prod_{j=1}^{k} \tau^{(0)}_{i_{\bullet j}} \Big(\sum_{\ell=1}^{d} \psi^{(\ell)}(K_\ell(Y_j^*) + f_\ell(Y_j^*)) \Big)}{\prod_{j=1}^{k} \tau^{(0)}_{i_{\bullet j}} \Big(\sum_{\ell=1}^{d} \psi^{(\ell)}(K_\ell(Y_j^*)) \Big)}$$
$$\times \frac{\prod_{\ell=1}^{d} \prod_{j=1}^{k} \prod_{t=1}^{i_{\ell,j}} \tau^{(\ell)}_{q_{\ell,j,t}}(K_\ell(Y_j^*) + f_\ell(Y_j^*))}{\prod_{\ell=1}^{d} \prod_{j=1}^{k} \prod_{t=1}^{i_{\ell,j}} \tau^{(\ell)}_{q_{\ell,j,t}}(K_\ell(Y_j^*))} \tag{58}$$

Consider separately the three parts contained in (58):

(1) the exponential part refers to η_0^*, in fact a simple rearrangement of the terms leads to

$$\exp\Big(\int_{\mathbb{Y}}\psi^{(0)}\Big(\sum_{\ell=1}^d\psi^{(\ell)}(K_\ell(y))\Big)c_0P_0(\mathrm{d}y)-\int_{\mathbb{Y}}\psi^{(0)}\Big(\sum_{\ell=1}^d\psi^{(\ell)}(K_\ell(y)+f_\ell(y))\Big)c_0P_0(\mathrm{d}y)\Big)$$
$$=\exp\Big(-\int_{\mathbb{Y}}\psi_*^{(0)}\Big(\sum_{\ell=1}^d\psi_*^{(\ell)}(f_\ell(y))\Big)c_0P_0(\mathrm{d}y)\Big),$$

where we have set:

$$\psi_*^{(0)}(u)=\int_0^\infty(1-e^{-su})e^{-s\sum_{\ell=1}^d\psi^{(\ell)}(K_\ell(y))}\rho_0(s)\mathrm{d}s$$
$$\psi_*^{(\ell)}(u)=\int_0^\infty(1-e^{-su})e^{-sK_\ell(y)}\rho_\ell(s)\mathrm{d}s\quad \ell=1,\ldots,d.$$

$\exp\Big(-\int_{\mathbb{Y}}\psi_*^{(0)}\Big(\sum_{\ell=1}^d\psi_*^{(\ell)}(f_\ell(y))\Big)c_0P_0(\mathrm{d}y)\Big)$ is the Laplace functional of the CRM η_0^* (compare with (7)).

(2) Concentrate on the following term in (58)

$$\frac{\prod_{j=1}^k\tau_{i_{\bullet j}}^{(0)}\Big(\sum_{\ell=1}^d\psi^{(\ell)}(K_\ell(Y_j^*)+f_\ell(Y_j^*))\Big)}{\prod_{j=1}^k\tau_{i_{\bullet j}}^{(0)}\Big(\sum_{\ell=1}^d\psi^{(\ell)}(K_\ell(Y_j^*))\Big)}$$

the denominator can be considered as a normalizing term, so we deal with the numerator:

$$\prod_{j=1}^k\tau_{i_{\bullet j}}^{(0)}\Big(\sum_{\ell=1}^d\psi^{(\ell)}(K_\ell(Y_j^*)+f_\ell(Y_j^*))\Big)=\prod_{j=1}^k\int_0^\infty e^{-s\sum_{\ell=1}^d\psi_*^{(\ell)}(f_\ell(Y_j^*))}s^{i_{\bullet j}}e^{-s\sum_{\ell=1}^d\psi^{(\ell)}(K_\ell(Y_j^*))}\rho_0(s)\mathrm{d}s$$

therefore:

$$\frac{\prod_{j=1}^k\tau_{i_{\bullet j}}^{(0)}\Big(\sum_{\ell=1}^d\psi^{(\ell)}(K_\ell(Y_j^*)+f_\ell(Y_j^*))\Big)}{\prod_{j=1}^k\tau_{i_{\bullet j}}^{(0)}\Big(\sum_{\ell=1}^d\psi^{(\ell)}(K_\ell(Y_j^*))\Big)}=\prod_{j=1}^k\int_0^\infty e^{-s\sum_{\ell=1}^d\psi_*^{(\ell)}(f_\ell(Y_j^*))}f_j(s)\mathrm{d}s,$$

which is the joint Laplace functional of d hierarchical completely random measures η_ℓ^*, $\ell=1,\ldots,d$, of the type:

$$\eta_\ell^*|I_1,\cdots,I_k\sim\mathrm{CRM}(e^{-sK_\ell(y)}\rho_\ell(s)\mathrm{d}s\sum_{j=1}^kI_j\delta_{Y_j^*}(\mathrm{d}y))$$

furthermore these random measures are independent given the weights $I_1,\ldots,I_k$.

(3) Consider the last term in (58):

$$\frac{\prod_{\ell=1}^d\prod_{j=1}^k\prod_{t=1}^{i_{\ell,j}}\tau_{q_{\ell,j,t}}^{(\ell)}(K_\ell(Y_j^*)+f_\ell(Y_j^*))}{\prod_{\ell=1}^d\prod_{j=1}^k\prod_{t=1}^{i_{\ell,j}}\tau_{q_{\ell,j,t}}^{(\ell)}(K_\ell(Y_j^*))}$$

as before the denominator is a normalizing constant, and this is the Laplace functional of $\Big(\sum_{j=1}^{k_1}\sum_{t=1}^{i_{1,j}}J_{1,j,t}\delta_{Y_{1,j}^*},\cdots,\sum_{j=1}^{k_d}\sum_{t=1}^{i_{d,j}}J_{d,j,t}\delta_{Y_{d,j}^*}\Big)$.

Items (1) and (2) give (i), instead (ii) follows from (3) and this concludes the proof. □

A.3.4 Proof of Corollary 3

First observe that:

$$\begin{aligned}\mathbb{E}[\tilde{S}_\ell(t)|\boldsymbol{X},\boldsymbol{Y},\boldsymbol{T}] &= \mathbb{E}[e^{-\int_0^t \tilde{h}_\ell(s)\mathrm{d}s}|\boldsymbol{X},\boldsymbol{Y},\boldsymbol{T}] \\ &= \mathbb{E}[e^{-\int_0^t\int_{\mathbb{Y}} k(s;y)\tilde{\mu}_\ell(\mathrm{d}y)\mathrm{d}s}|\boldsymbol{X},\boldsymbol{Y},\boldsymbol{T}] = \mathbb{E}[e^{-\int_{\mathbb{Y}} K_t^{(\ell)}(y)\tilde{\mu}_\ell(\mathrm{d}y)}|\boldsymbol{X},\boldsymbol{Y},\boldsymbol{T}].\end{aligned}$$

Then the proposition is a simple consequence of Theorem 7, from which we can deduce the posterior distribution of $\tilde{\mu}_\ell$:

$$\mathbb{E}[\tilde{S}_\ell(t)|\boldsymbol{X},\boldsymbol{Y},\boldsymbol{T}] = \mathbb{E}[e^{-\int_{\mathbb{Y}} K_t^{(\ell)}(y)\tilde{\mu}_\ell^*(\mathrm{d}y)}]\mathbb{E}\Big[e^{-\int_{\mathbb{Y}} K_t^{(\ell)}(y)\sum_{j=1}^{k_\ell}\sum_{t=1}^{i_{\ell,j}} J_{\ell,j,t}\delta_{Y^*_{\ell,j}}(\mathrm{d}y)}\Big] \tag{59}$$

The two expectations can be evaluated using their characterization, given in Theorem 7:

$$\begin{aligned}\mathbb{E}[e^{-\int_{\mathbb{Y}} K_t^{(\ell)}(y)\tilde{\mu}_\ell^*(\mathrm{d}y)}] &= \mathbb{E}[\mathbb{E}[e^{-\int_{\mathbb{Y}} K_t^{(\ell)}(y)\tilde{\mu}_\ell^*(\mathrm{d}y)}|\tilde{\mu}_0^*]] \\ &= \mathbb{E}[\exp\Big(-\int_{\mathbb{Y}}\int_0^\infty (1-e^{-sK_t^{(\ell)}(y)})e^{-sK_\ell(y)}\rho_\ell(s)\mathrm{d}s\tilde{\mu}_0^*(\mathrm{d}y)\Big)] \\ &= \mathbb{E}[e^{-\int_{\mathbb{Y}} \psi_*^{(\ell)}(K_\ell(y))\tilde{\mu}_0^*(\mathrm{d}y)}] \\ &= \exp\Big(-\int_{\mathbb{Y}}\int_0^\infty (1-e^{-s\psi_*^{(\ell)}(K_t^{(\ell)}(y))})e^{-s\sum_{\ell=1}^d \psi^{(\ell)}(K_\ell(y))}\rho_0(s)\mathrm{d}s c_0 P_0(\mathrm{d}y)\Big) \\ &\qquad \times \mathbb{E}\Big[\exp\Big(-\int_{\mathbb{Y}}\psi_*^{(\ell)}(K_t^{(\ell)}(y))\sum_{j=1}^k I_j\delta_{Y_j^*}(\mathrm{d}y)\Big)\Big] \\ &= \exp\Big(-\int_{\mathbb{Y}}\psi_*^{(0)}(\psi_*^{(\ell)}(K_t^{(\ell)}(y)))c_0P_0(\mathrm{d}y)\Big)\prod_{j=1}^k \mathbb{E}[e^{-\psi_*^{(\ell)}(K_t^{(\ell)}(Y_j^*))I_j}] \\ &= \exp\Big(-\int_{\mathbb{Y}}\psi_*^{(0)}(\psi_*^{(\ell)}(K_t^{(\ell)}(y)))c_0P_0(\mathrm{d}y)\Big) \\ &\qquad \times \prod_{j=1}^k \frac{\int_0^\infty e^{-s(\psi_*^{(\ell)}(K_t^{(\ell)}(Y_j^*))+\sum_{\ell=1}^d\psi^{(\ell)}(K_\ell(Y_j^*)))}s^{i_{\bullet j}}\rho_0(s)\mathrm{d}s}{\int_0^\infty e^{-s\sum_{\ell=1}^d\psi^{(\ell)}(K_\ell(Y_j^*)))}s^{i_{\bullet j}}\rho_0(s)}.\end{aligned}$$

The second expectation in (59) is equal to:

$$\begin{aligned}\mathbb{E}\Big[e^{-\int_{\mathbb{Y}} K_t^{(\ell)}(y)\left(\sum_{j=1}^{k_\ell}\sum_{t=1}^{i_{\ell,j}} J_{\ell,j,t}\delta_{Y^*_{\ell,j}}\right)(\mathrm{d}y)}\Big] &= \prod_{j=1}^{k_\ell}\prod_{t=1}^{i_{\ell,j}}\mathbb{E}[e^{-K_t^{(\ell)}(Y^*_{\ell,j})J_{\ell,j,t}}] \\ &= \prod_{j=1}^{k_\ell}\prod_{t=1}^{i_{\ell,j}}\frac{\int_0^\infty e^{-s(K_t^{(\ell)}(Y^*_{\ell,j})+K_\ell(Y^*_{\ell,j}))}s^{q_{\ell,j,t}}\rho_\ell(s)\mathrm{d}s}{\int_0^\infty e^{-sK_\ell(Y^*_{\ell,j})}s^{q_{\ell,j,t}}\rho_\ell(s)\mathrm{d}s}.\end{aligned}$$

Putting the previous expressions in (59), the thesis follows. □

A.3.5 Proof of Corollary 4

Thanks to Theorem 7, we can write

$$\begin{aligned}\mathbb{E}[\tilde{h}_\ell(t)|\boldsymbol{X},\boldsymbol{Y},\boldsymbol{T}] &= \mathbb{E}\Big[\int_{\mathbb{Y}} k(t;y)\tilde{\mu}_\ell(\mathrm{d}y)\Big|\boldsymbol{X},\boldsymbol{Y},\boldsymbol{T}\Big] \\ &= \mathbb{E}\Big[\int_{\mathbb{Y}} k(t;y)\tilde{\mu}_\ell^*(\mathrm{d}y)\Big] + \sum_{j=1}^{k_\ell}\sum_{t=1}^{i_{\ell,j}} k(t;Y^*_{\ell,j})\mathbb{E}[J_{\ell,j,t}].\end{aligned} \tag{60}$$

We now focus on the first term of (60)

$$
\begin{aligned}
\mathbb{E}\Big[\int_{\mathbb{Y}} k(t;y)\tilde{\mu}_\ell^*(\mathrm{d}y)\Big] &= \mathbb{E}\Big[\mathbb{E}\Big[\int_{\mathbb{Y}} k(t;y)\tilde{\mu}_\ell^*(\mathrm{d}y)\Big|\tilde{\mu}_0^*\Big]\Big] \\
&= -\frac{\mathrm{d}}{\mathrm{d}u}\Big(\mathbb{E}\Big[\mathbb{E}\Big[e^{-u\int_{\mathbb{Y}} k(t;y)\tilde{\mu}_\ell^*(\mathrm{d}y)}\Big|\tilde{\mu}_0^*\Big]\Big]\Big)\Big|_{u=0} \\
&= -\frac{\mathrm{d}}{\mathrm{d}u}\Big(\mathbb{E}\Big[\exp\Big\{-\int_{\mathbb{Y}}\psi_*^{(\ell)}\left(uk(t;y)\right)\tilde{\mu}_0^*(\mathrm{d}y)\Big\}\Big]\Big)\Big|_{u=0} \\
&= -\frac{\mathrm{d}}{\mathrm{d}u}\Big(\mathbb{E}\Big[\exp\Big\{-\int_{\mathbb{Y}}\psi_*^{(\ell)}\left(uk(t;y)\right)\eta_0^*(\mathrm{d}y)\Big\}\Big] \\
&\quad\times\prod_{j=1}^{k}\mathbb{E}\Big[\exp\Big\{\psi_*^{(\ell)}\left(uk(t;Y_j^*)\right)I_j\Big\}\Big]\Big)\Big|_{u=0},
\end{aligned}
$$

where we have used the posterior representation of $\tilde{\mu}_0^*$ given in Theorem 7. Straightforward calculations of the derivative in the previous expression lead to

$$
\begin{aligned}
\mathbb{E}\Big[\int_{\mathbb{Y}} k(t;y)\tilde{\mu}_\ell^*(\mathrm{d}y)\Big] &= \int_{\mathbb{Y}} k(t;y)\int_0^\infty s e^{-sK_\ell(y)}\rho_\ell(s)\mathrm{d}s \\
&\quad\times\int_0^\infty w e^{-w\sum_{h=1}^d K_h(y)}\rho_0(w)\mathrm{d}w c_0 P_0(\mathrm{d}y) \\
&\quad+\sum_{j=1}^{k} k(t;Y_j^*)\int_0^\infty s e^{-sK_\ell(Y_j^*)}\rho_\ell(s)\mathrm{d}s\int_0^\infty w f_j(w)\mathrm{d}w
\end{aligned}
\tag{61}
$$

being f_j the density function of the jump I_j, for any $j=1,\dots,k$. As for the second term in (60), we have:

$$
\sum_{j=1}^{k_\ell}\sum_{t=1}^{i_{\ell,j}} k(t;Y_{\ell,j}^*)\mathbb{E}[J_{\ell,j,t}] = \sum_{j=1}^{k_\ell}\sum_{t=1}^{i_{\ell,j}} k(t;Y_{\ell,j}^*)\int_0^\infty w f_{\ell,j,t}(w)\mathrm{d}w. \tag{62}
$$

Hence, putting the expressions (61)–(62) in (60), the thesis follows. □

A.4 Proofs of Section 6.1

A.4.1 Proof of Corollary 5

The posterior expected value of the survival function $\mathbb{E}[\tilde{S}_\ell(t)\,|\,\boldsymbol{X},\boldsymbol{Y},\boldsymbol{T}]$ can be computed resorting to (35) and specializing $\psi_*^{(\ell)},\psi_*^{(0)}$ in the case of generalized gamma CRM. With our assumptions, it is easy to see that:

$$
\begin{aligned}
\psi_*^{(\ell)}(u) &= \frac{(u+1+K_\ell(y))^\sigma-(1+K_\ell(y))^\sigma}{\sigma} \qquad \ell=1,\dots,d \\
\psi_*^{(0)}(u) &= \frac{1}{\sigma_0}\Bigg\{\Bigg[u+\sum_{\ell=1}^{d}\frac{(K_\ell(y)+1)^\sigma-1}{\sigma}+1\Bigg]^{\sigma_0}-\Bigg[\sum_{\ell=1}^{d}\frac{(K_\ell(y)+1)^\sigma-1}{\sigma}+1\Bigg]^{\sigma_0}\Bigg\},
\end{aligned}
$$

for any $q\in\mathbb{N}$ and $u>0$, we also have that

$$
\int_0^\infty e^{-su}s^q\rho_\ell(s)\mathrm{d}s = \frac{(1-\sigma)_{q-1}}{(u+1)^{q-\sigma}} \qquad \ell=1,\dots,d.
$$

Substituting these expressions in (35), a straightforward calculation gives the result. □

A.5 A marginal MCMC sampler for hierarchies of generalized Gamma processes

Here we describe the MCMC sampler (see Section 6.1) that allows to marginalize out the latent variables, generating samples from their posterior distribution given the observed survival times $\boldsymbol{X}$. The full conditional distributions may be derived resorting to (26), which in the specific case of hierarchies of generalized Gamma processes becomes

$$\pi(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{T}) \propto \Big\{\prod_{j=1}^{k} c_0 P_0(\mathrm{d}Y_j^*)\Big\}\Big\{\prod_{\ell=1}^{d}\prod_{i=1}^{n_\ell} k(X_{\ell,i};Y_{\ell,i})\Big\}$$
$$\times \exp\left\{-\frac{c_0}{\sigma_0}\int_{\mathbb{Y}}\left[\left(1+\sum_{\ell=1}^{d}\frac{(1+K_\ell(y))^\sigma-1}{\sigma}\right)^{\sigma_0}-1\right]P_0(\mathrm{d}y)\right\}$$
$$\times\prod_{j=1}^{k}\frac{(1-\sigma_0)_{i_{\bullet j}-1}}{\left(1+\sum_{\ell=1}^{d}\frac{(1+K_\ell(Y_j^*))^\sigma-1}{\sigma}\right)^{i_{\bullet j}-\sigma_0}}$$
$$\times\prod_{\ell=1}^{d}\prod_{j=1}^{k}\prod_{t=1}^{i_{\ell,j}}\frac{(1-\sigma)_{q_{\ell,j,t}-1}}{(1+K_\ell(Y_j^*))^{q_{\ell,j,t}-\sigma}}P_0(\mathrm{d}T^*_{\ell,j,t}),$$

where, with some abuse of notation, we agree that $\Gamma(0)\equiv 1$ and $\prod_{p=1}^{0}\equiv 1$. To simplify the notation we will set $\boldsymbol{\Phi} := ((\boldsymbol{Y}_\ell,\boldsymbol{T}_\ell)_\ell,\alpha,c_0,\sigma,\sigma_0)$ and we denote by $\boldsymbol{\Phi}_{-V}$ the vector $\boldsymbol{\Phi}$ with the component V removed. The steps of the MCMC algorithms may be read in terms of the Chinese restaurant franchise metaphor, as well as the marginal sampler in [6], however the full conditional distributions differ from those of [6] due to the different modeling assumptions. All the steps of the MCMC procedure are described below.

Update the couple $(Y_{\ell,r},T_{\ell,r})$ for $r=1,\dots,n_\ell$ and $\ell=1,\dots,d$.

For the sake of clarity we explicitly deal with the case $\ell=1$, since everything can be replicated in a straightforward way for general ℓ. The full conditional distributions are a convex linear combination of three cases

$$\mathbb{P}(Y_{1,r}\in\mathrm{d}y,T_{1,r}\in\mathrm{d}t\mid\boldsymbol{\Phi}_{-(Y_{1,r},T_{1,r})},\boldsymbol{X}) = w_{1,0}G_0(\mathrm{d}y)P_0(\mathrm{d}t)$$
$$+\sum_{h=1}^{k^{-r}} w_{1,h}\delta_{Y_h^{*,-r}}(\mathrm{d}y)P_0(\mathrm{d}t)+\sum_{h=1}^{k_1^{-r}}\sum_{t=1}^{i_{1,h}^{-r}} w_{1,h,t}\delta_{Y_{1,h}^{*,-r}}(\mathrm{d}y)\delta_{T_{1,h,t}^{*,-r}}(\mathrm{d}t)$$

where

$$G_0(\mathrm{d}y)\propto\frac{k(X_{1,r};y)P_0(\mathrm{d}y)}{\left(1+\sum_{\ell=1}^{d}\frac{(1+K_\ell(y))^\sigma-1}{\sigma}\right)^{1-\sigma_0}(1+K_\ell(y))^{1-\sigma}}$$

and the three summands appearing in the expression above can be given straightforward interpretations as follows:

(1) the r–th costumer eats a new dish seating at a new table in the first restaurant with probability

$$w_{1,0}\propto\int_{\mathbb{Y}}\frac{c_0 k(X_{1,r};y)}{\left(1+\sum_{\ell=1}^{d}\frac{(1+K_\ell(y))^\sigma-1}{\sigma}\right)^{1-\sigma_0}(1+K_\ell(y))^{1-\sigma}}P_0(\mathrm{d}y),$$

and, conditional on the fact that a new dish and a new table is chosen, their labels are selected from $G_0 \times P_0$,

(2) the r–th costumer eats the h–th dish already served in the franchise in a new table with probability

$$w_{1,h} \propto \frac{k(X_{1,r}; Y_h^{*,-r})(i_{\bullet h}^{-r} - \sigma_0)}{(1 + \sum_{\ell=1}^d \frac{(1+K_\ell(Y_h^{*,-r}))^\sigma - 1}{\sigma})(1 + K_1(Y_h^{*,-r}))^{1-\sigma}},$$

for each $h = 1, \ldots, k_1^{-r}$, and the new table's label is drawn from P_0;

(3) the r–th costumer eats the h–th dish already served in the franchise while seating at the t–th existing table of the first restaurant, where the dish is actually served, with probability

$$w_{1,h,t} \propto \frac{k(X_{1,r}; Y_{1,h}^{*,-r})(q_{1,h,t}^{-r} - \sigma)}{1 + K_1(Y_{1,h}^{*,-r})},$$

for any $h = 1, \ldots, k_1^{-r}$.

The integral that defines $w_{1,0}$, in (1), can be evaluated through a Monte Carlo method, by generating i.i.d. realizations from P_0, which will be assumed to be a uniform distribution over $[0, T]$, being T large enough so that all the survival times are contained in the interval $[0, T]$.

Sample a new latent variable and acceleration step

A Metropolis–Hastings algorithm has been developed in order to sample a new latent variable from G_0, when P_0 is the probability distribution of a uniform distribution in $[0, T]$. The procedure can be obviously adapted to any alternative choice of a diffuse P_0.

As suggested in [24], we add an acceleration step, which consists in resampling all the distinct values of $\boldsymbol{Y}$ at the end of every iteration. $\{C_{\ell,j}\}_{j=1,\ldots,k_\ell}$ will denote the (finer) random partition associated to $\{Y_{\ell,1}^*, \ldots, Y_{\ell,k_\ell}^*\}$, for every $\ell = 1, \ldots, d$. Setting $\bar{X}_j := \min_{i \in C_{1,j}} X_{1,i} \wedge \ldots \wedge \min_{i \in C_{d,j}} X_{d,i}$, the full conditional distribution for the distinct values is

$$\pi(Y_j^* \mid \boldsymbol{\Phi}_{-Y_j^*}, \boldsymbol{X}) \propto \left(1 + \sum_{h=1}^d \frac{(1 + K_h(Y_j^*))^\sigma - 1}{\sigma}\right)^{-(i_{\bullet j} - \sigma_0)} \times \prod_{\ell=1}^d \frac{1}{(1 + K_\ell(Y_j^*))^{n_{\ell,j} - i_{\ell,j}\sigma}} \mathbb{1}_{(0, \bar{X}_j]}(Y_j^*) P_0(\mathrm{d}Y_j^*).$$

The issue of sampling from the previous distribution can be easily tackled by means of a Metropolis–Hastings step with a change of scale. Indeed, since the full conditional of Y_j^* is concentrated on $(0, \bar{X}_j]$, the transformation $Z_j = g(Y_j^*) := \tan(\pi(Y_j^*/\bar{X}_j - 1/2))$ is invertible and Z_j is, thus, updated through a Metropolis–Hastings step. This is implemented by choosing a proposal that equals a normal distribution centered in the previous value of Z_j and with fixed variance (in our case 1). Finally, one updates Y_j^* by setting $Y_j^* = g^{-1}(Z_j)$.

Update α and c_0

In view of (37), we let κ_α denote that Gamma (a, b) prior for α. The full conditional distribution for α is then

$$\pi(\alpha \,|\, \mathbf{\Phi}_{-\alpha}, \boldsymbol{X}) \propto \kappa_\alpha(\alpha)\, \alpha^{n_1+n_2} \exp\left\{ -\frac{c_0}{\sigma_0} \int_{\mathbb{Y}} \left[\left(1 + \sum_{\ell=1}^{d} \frac{(1+K_\ell(y))^\sigma - 1}{\sigma} \right)^{\sigma_0} - 1 \right] P_0(\mathrm{d}y) \right\}$$
$$\times \prod_{j=1}^{k} \frac{\prod_{\ell=1}^{d} \Big(1 + K_\ell(Y_j^*)\Big)^{-(n_{\ell,j} - i_{\ell,j}\sigma)}}{\Big(1 + \sum_{h=1}^{d} [(1+K_h(Y_j^*))^\sigma - 1]/\sigma\Big)^{i_{\bullet j} - \sigma_0}}.$$

The parameter α is updated through a random walk Metropolis–Hastings step. More precisely we generate $z = \log(\alpha)$ using, as a proposal, a normal distribution centered around the previously accepted value of z with fixed variance 0.5 and, then, determine α by inverting the transformation. Finally, since c_0 is set as a gamma random variable with parameters (a_0, b_0), the full conditional $\pi(c_0 | \mathbf{\Phi}_{-c_0}, \boldsymbol{X})$ is available in closed form and equals

$$\mathrm{Ga}\Big(k + a_0, b_0 + \frac{1}{\sigma_0} \int_{\mathbb{Y}} \left[\left(1 + \sum_{\ell=1}^{d} \frac{(1+K_\ell(y))^\sigma - 1}{\sigma} \right)^{\sigma_0} - 1 \right] P_0(\mathrm{d}y)\Big).$$

which can be directly sampled from, once the integral defining the second parameter of the gamma distribution is approximately evaluated through a plain Monte Carlo procedure.

A.6 Proof and algorithmic details of the conditional sampler

A.6.1 Derivation of the approximation (38)

In the present section we discuss how we can estimate the whole posterior distribution of the survival functions once we have obtained samples of the trajectories of the CRMs $(\tilde{\mu}_1, \ldots, \tilde{\mu}_k)$ *a posteriori*; in particular we discuss how formula (38) has been obtained. Using the same notations as in the Section 6.1, we will denote by $\mathbf{\Phi} := ((\boldsymbol{Y}_\ell, \boldsymbol{T}_\ell)_\ell, \alpha, c_0)$ the vector containing all the latent variables of the model. We first notice that the posterior c.d.f. of the survival function $\tilde{S}_\ell(t)$ may be rewritten as

$$\mathbb{P}(\tilde{S}_\ell(t) \le s | \boldsymbol{X}) = \int_\Theta \mathbb{P}(\tilde{S}_\ell(t) \le s | \mathbf{\Phi}, \boldsymbol{X}) \mathbb{P}(\mathrm{d}\mathbf{\Phi} | \boldsymbol{X}),$$

for any $s \ge 0$ and $\ell = 1, \ldots, d$, where Θ denotes the state space of the whole vector of latent variables Φ. Assume that N vectors of latent variables $\mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_N$ have been sampled through an MCMC procedure, it is then possible to approximate the posterior c.d.f. as follows

$$\mathbb{P}(\tilde{S}_\ell(t) \le s | \boldsymbol{X}) \approx \frac{1}{N} \sum_{i=1}^{N} \mathbb{P}(\tilde{S}_\ell(t) \le s | \mathbf{\Phi}_i, \boldsymbol{X}). \tag{63}$$

An approximation of the probability $\mathbb{P}(\tilde{S}_\ell(t) \le s | \mathbf{\Phi}_i, \boldsymbol{X})$ appearing in (63) may be easily obtained through its empirical c.d.f.. In fact, having generated M trajectories of $\tilde{\mu}_\ell$, conditionally given $\mathbf{\Phi}_i, \boldsymbol{X}$, and denoting them as $\tilde{\mu}_\ell^{(i,1)}, \ldots, \tilde{\mu}_\ell^{(i,M)}$, the posterior c.d.f. appearing in (63) is approximated by

$$\mathbb{P}(\tilde{S}_\ell(t) \le s | \boldsymbol{X}) \approx \frac{1}{NM} \sum_{i=1}^{N} \sum_{k=1}^{M} \mathbb{1}_{\left\{\tilde{S}_\ell^{(i,k)}(t) \le s\right\}}, \tag{64}$$

where

$$\tilde{S}_\ell^{(i,k)}(t) := \exp\Big(- \int_0^t \int_{\mathbb{Y}} k(s;y)\tilde{\mu}_\ell^{(i,k)}(\mathrm{d}y)\mathrm{d}s\Big).$$

□

A.6.2 Proof of Proposition 2

We prove the proposition for $\ell = 0$, the general case proceeds in a similar fashion. Remind that the CRM η_0^* may be represented as

$$\eta_0^* = \sum_{h\geq 1} J_h^{(0)} \delta_{y_h^{(0)}}$$

where the $J_h^{(0)}$'s arise from the Wolpert and Ickstadt [44] representation. We further recall that the jumps in the previous representation of η_0^* satisfy the equation

$$S_h^{(0)} = \int_{J_h^{(0)}}^{\infty} \eta(s, y_h^{(0)})\mathrm{d}s$$

being

$$\eta(s, y_h^{(0)}) = c_0 \mathrm{e}^{-s\sum_{\ell=1}^d \psi_\ell(K_\ell(y_h^{(0)}))}\rho_0(s) \leq \rho_0(s)c_0, \tag{65}$$

where the very last inequality holds true because the functions $\psi^{(\ell)}(\,\cdot\,)$'s are positive. As a consequence, thanks to (65), we get

$$S_h^{(0)} \leq c_0 \int_{J_h^{(0)}}^{\infty} \rho_0(s)\mathrm{d}s. \tag{66}$$

Now recall that $\tilde{J}_h^{(0)}$ is implicitly defined by the equation

$$S_h^{(0)} = c_0 \int_{\tilde{J}_h^{(0)}}^{\infty} \rho_0(s)\mathrm{d}s, \tag{67}$$

whose solution uniquely exists, indeed $S_h^{(0)} > 0$ and $\rho_0(s)$ is such that $\int_0^\infty \rho_0(s)\mathrm{d}s = +\infty$. Therefore we are now provided with two sequences of jumps $\{J_h^{(0)}\}_{h\geq 1}$ and $\{\tilde{J}_h^{(0)}\}_{h\geq 1}$, which satisfy the properties in the statement of the Proposition:

(a) $\{\tilde{J}_h^{(0)}\}$ is monotonically decreasing.
Indeed, since $S_h^{(0)}$ are the points of a Poisson process, if $h_1 < h_2$ then $S_{h_1}^{(0)} \leq S_{h_2}^{(0)}$, which implies that (see (67))

$$\int_{\tilde{J}_{h_1}^{(0)}}^{\infty} \rho_0(s)\mathrm{d}s \leq \int_{\tilde{J}_{h_2}^{(0)}}^{\infty} \rho_0(s)\mathrm{d}s$$

which finally gets $\tilde{J}_{h_1}^{(0)} \geq \tilde{J}_{h_2}^{(0)}$.

(b) $J_h^{(0)} \leq \tilde{J}_h^{(0)}$ for any $h \geq 1$.
Putting equality (67) in (66), we obtain

$$\int_{\tilde{J}_h^{(0)}}^{\infty} \rho_0(s)\mathrm{d}s \leq \int_{J_h^{(0)}}^{\infty} \rho_0(s)\mathrm{d}s,$$

which is true iff $\tilde{J}_h^{(0)} \geq J_h^{(0)}$, for every $h \geq 1$.

□

A.6.3 Conditional sampler for hierarchies of generalized gamma processes

We will specialize the general algorithm developed in Section 6.2 when each $\tilde{\mu}_\ell$ is a generalized gamma CRM, i.e. $\rho_\ell(s) = (\Gamma(1-\sigma))^{-1}e^{-s}/s^{1+\sigma}$ for $\ell = 1,\dots,d$, and $\rho_0(s) = (\Gamma(1-\sigma_0))^{-1}e^{-s}/s^{1+\sigma_0}$. In the sequel we will denote by

$$\Gamma_a(x) := \frac{1}{\Gamma(a)}\int_x^\infty t^{a-1}e^{-t}\mathrm{d}t$$

the incomplete gamma function for any $a > 0$ and $x > 0$. The steps of the algorithm are described below.

(1) Generate $\tilde{\mu}_0$ from its posterior distribution, which is described right before Theorem 7, namely:

(1.a) generate the random jumps I_j, as $j = 1,\dots,k$, from a gamma distribution

$$I_j \sim \mathrm{Gam}\Big(i_{\bullet j} - \sigma_0, 1 + \sigma^{-1}\sum_{\ell=1}^d [(1+K_\ell(Y_j^*))^\sigma - 1]\Big)$$

(1.b) generate η_0^* using the algorithm developed by Wolpert and Ickstadt [44], i.e. fix a threshold level $\varepsilon > 0$ and proceed with the following steps:

- generate the atom of the CRM $y_h^{(0)} \sim P_0$;
- generate the waiting times $S_h^{(0)}$ of a standard Poisson process, that is to say $S_h^{(0)} - S_{h-1}^{(0)}$ are independent and identically distributed exponential random variables with unit mean;
- determine the jump $J_h^{(0)}$ by inverting the Lévy intensity, i.e. solve the equation

$$\frac{S_h^{(0)}\Gamma(1-\sigma_0)\sigma_0}{[1+\sum_{\ell=1}^d \psi_\ell(K_\ell(y_h^{(0)}))]^{\sigma_0}c_0} = \frac{\exp\{-J_h^{(0)}[1+\sum_{\ell=1}^d \psi_\ell(K_\ell(y_h^{(0)}))]\}}{[1+\sum_{\ell=1}^d \psi_\ell(K_\ell(y_h^{(0)}))]^{\sigma_0}(J_h^{(0)})^{\sigma_0}} - \Gamma(1-\sigma_0)\Gamma_{1-\sigma_0}(J_h^{(0)}[1+\sum_{\ell=1}^d \psi_\ell(K_\ell(y_h^{(0)}))]), \tag{68}$$

  note that $\psi_\ell(u) = \sigma^{-1}[(u+1)^\sigma - 1]$ for any $\ell = 1,\dots,d$, we also point out that one may find $J_h^{(0)}$ in equation (68) resorting to a suitable numerical method (e.g. bisection method);
- sample an additional jump $\tilde{J}_h^{(0)}$ by inverting the equation

$$\frac{S_h^{(0)}\Gamma(1-\sigma_0)\sigma_0}{c_0} = \frac{e^{-\tilde{J}_h^{(0)}}}{(\tilde{J}_h^{(0)})^{\sigma_0}} - \Gamma(1-\sigma_0)\Gamma_{1-\sigma_0}(\tilde{J}_h^{(0)})$$

  and stop the procedure if $\tilde{J}_h^{(0)} \le \varepsilon$.

(1.c) Evaluate an approximate draw of $\tilde{\mu}_0^*$

$$\tilde{\mu}_0^* = \sum_{h=1}^{H^\varepsilon} J_h^{(0)}\delta_{y_h^{(0)}} + \sum_{j=1}^k I_j\delta_{Y_j^*}.$$

(2) Generate $\tilde{\mu}_\ell$ given $\tilde{\mu}_0^*, \boldsymbol{X}$, more precisely:

(2.a) generate the random jumps $J_{\ell,j,t}$ from a Gamma distribution

$$J_{\ell,j,t} \sim \mathrm{Gam}(q_{\ell,j,t} - \sigma, K_\ell(Y^*_{\ell,j}) + 1).$$

(2.b) generate the CRM $\tilde{\mu}_\ell^*$ using the posterior representation given in Theorem 7 and the algorithm by Wolpert and Ickstadt [44], namely:

- generate the atom of the CRM $y_h^{(\ell)} \sim P_0^*$, where P_0^* is the probability measure obtained from the normalization of the draw of $\tilde{\mu}_0^*$ in (1.c);
- generate the waiting times $S_h^{(\ell)}$ of a standard Poisson process, that is to say $S_h^{(\ell)} - S_{h-1}^{(\ell)}$ are independent and identically distributed exponential random variables with unit mean;
- determine the jump $J_h^{(\ell)}$ by inverting the Lévy intensity (31), i.e. solve the equation

$$\frac{S_h^{(0)}\Gamma(1-\sigma)\sigma}{(1+K_\ell(y_h^{(\ell)}))^\sigma \tilde{c}} = \frac{\exp\{-J_h^{(\ell)}(1+K_\ell(y_h^{(\ell)}))\}}{(1+K_\ell(y_h^{(\ell)}))^\sigma (J_h^{(\ell)})^\sigma} - \Gamma(1-\sigma)\Gamma_{1-\sigma}(J_h^{(\ell)}(1+K_\ell(y_h^{(\ell)}))),$$

  being $\tilde{c} := \sum_{h=1}^{H^\varepsilon} J_h^{(0)} \delta_{y_h^{(0)}} + \sum_{j=1}^{k} I_j \delta_{Y_j^*}$ the total mass of $\tilde{\mu}_0^*$;
- sample an additional jump $\tilde{J}_h^{(\ell)}$ solving the equation

$$\frac{S_h^{(\ell)}\Gamma(1-\sigma)\sigma}{\tilde{c}} = \frac{e^{-\tilde{J}_h^{(\ell)}}}{(\tilde{J}_h^{(\ell)})^\sigma} - \Gamma(1-\sigma)\Gamma_{1-\sigma}(\tilde{J}_h^{(\ell)})$$

  and stop the procedure if $\tilde{J}_h^{(\ell)} \leq \varepsilon$.

(2.c) Evaluate an approximate draw from the posterior of $\tilde{\mu}_\ell$, by putting

$$\tilde{\mu}_\ell \approx \sum_{h=1}^{H_\ell^\varepsilon} J_h^{(\ell)} \delta_{y_h^{(\ell)}} + \sum_{j=1}^{k_\ell} \sum_{t=1}^{i_{\ell,j}} J_{\ell,j,t} \delta_{Y^*_{\ell,j}}. \tag{69}$$

A.7 Covariates dependent and censored data: an illustration

Here we discuss how to adapt the model in presence of covariates and right–censored survival times: the results we have stated for the case where only exact data without covariates are recorded can be easily extended to this more general setting. We then present an application to tumor survival data in Subsection A.7.1

Suppose d partially exchangeable samples in (9) $\boldsymbol{X}_\ell = (X_{\ell,1}, \cdots, X_{\ell,n_\ell})$, for $\ell = 1, \dots, d$, are available together with their individuals' covariates across all d samples $\{\boldsymbol{z}_\ell : \ell = 1, \dots, d\} = \{(\boldsymbol{z}_{\ell,1}, \cdots, \boldsymbol{z}_{\ell,n_\ell}) : \ell = 1, \dots, d\}$, with each $\boldsymbol{z}_{\ell,i}$ taking values in a set $\mathcal{Z} \subset \mathbb{R}^p$. For instance, $\boldsymbol{X}_\ell$ may be a vector of survival times of patients treated at the ℓ–th hospital and affected by different tumors. We include the covariate information via an exponential term in the mixture representation of the hazard (8). More specifically, our proposal is based on a Cox regression model that defines a semiparametric prior for the hazard rate functions of each individual sample as

$$\tilde{h}_\ell(t;\boldsymbol{z}) = e^{\boldsymbol{\beta}_\ell\cdot\boldsymbol{z}}\int_{\mathbb{Y}} k(t;y)\tilde{\mu}_\ell(\mathrm{d}y),$$

where $\boldsymbol{\beta}_\ell\cdot\boldsymbol{z}$ denotes the scalar product between the vector of predictors and the associated covariates. The survival function for each group ℓ is thus given by

$$\tilde{S}_\ell(t,\boldsymbol{z}) = \exp\left\{-\int_{\mathbb{Y}} K_t^{(\ell)}(y,\boldsymbol{z})\tilde{\mu}_\ell(\mathrm{d}y)\right\}$$

where now $K_t^{(\ell)}(y,\boldsymbol{z}) := \mathrm{e}^{\boldsymbol{\beta}_\ell\cdot\boldsymbol{z}}\int_0^t k(s;y)\,\mathrm{d}s$ for each $\ell = 1,\cdots,d$. The likelihood function (22) in presence of covariates is the following

$$\mathscr{L}(\tilde{\boldsymbol{\mu}},\boldsymbol{\beta};\boldsymbol{X}) = \kappa_{\boldsymbol{\beta}}(\boldsymbol{\beta})\,\mathrm{e}^{-\sum_{\ell=1}^d\int_{\mathbb{Y}} K_\ell(y,\boldsymbol{z}_\ell)\tilde{\mu}_\ell(\mathrm{d}y)}\prod_{\ell=1}^d\prod_{i=1}^{n_\ell}\mathrm{e}^{\boldsymbol{\beta}_\ell\cdot\boldsymbol{z}_{\ell,i}}\int_{\mathbb{Y}} k(X_{\ell,i};y)\,\tilde{\mu}_\ell(\mathrm{d}y) \tag{70}$$

where $\boldsymbol{z}_\ell = (\boldsymbol{z}_{\ell,1},\ldots,\boldsymbol{z}_{\ell,n_\ell})$, with each $\boldsymbol{z}_{\ell,i}$ in $\mathcal{Z}$, and

$$K_\ell(y,\boldsymbol{z}_\ell) := \sum_{i=1}^{n_\ell} K_{X_{\ell,i}}^{(\ell)}(y,\boldsymbol{z}_{\ell,i}) = \mathrm{e}^{\boldsymbol{\beta}_\ell\cdot\boldsymbol{z}_{\ell,i}}\sum_{i=1}^{n_\ell}\int_0^{X_{\ell,i}} k(s,y)\,\mathrm{d}s \tag{71}$$

for $\ell = 1,\ldots,d$, and we have denoted by $\kappa_{\boldsymbol{\beta}}$ the prior distribution of the $\boldsymbol{\beta}$ coefficients. The likelihood (70) differs from (22) for the presence of $K_\ell(y,\boldsymbol{z}_\ell)$ in place of $K_\ell(y)$ and for the presence of the exponential term $\mathrm{e}^{\boldsymbol{\beta}_\ell\cdot\boldsymbol{z}_{\ell,i}}$, which however depends only on the predictor $\boldsymbol{\beta}_\ell$ and not on the other variables. Then it is not difficult to adapt all the results in our paper to the case of covariates dependent data.

We now discuss a further extension to accommodate for censored observations. First of all, we have to define the censoring time $C_{\ell,i}$ corresponding to the observation $X_{\ell,i}$, and the indicator variable $\Delta_{\ell,i} = \mathbb{1}_{(0,C_{\ell,i}]}(X_{\ell,i})$ which is $\Delta_{\ell,i} = 0$ if $X_{\ell,i}$ is censored, $\Delta_{\ell,i} = 1$ otherwise. Here the observed survival time turns out to be $X'_{\ell,i} = \min\{X_{\ell,i},C_{\ell,i}\}$ and the actual observations, then, are summarized through $\boldsymbol{D}_\ell = \{(X'_{\ell,j},\Delta_{\ell,j}) : j = 1,\ldots,n_\ell\}$ for any $\ell = 1,\ldots,d$. For the sake of simplifying notation we define $\boldsymbol{D} := \{\boldsymbol{D}_\ell : \ell = 1,\ldots,d\}$. Taking into account the possible presence of right–censored observations, the expression of the likelihood (70) modifies as follows

$$\begin{aligned}\mathscr{L}(\tilde{\boldsymbol{\mu}},\boldsymbol{\beta};\boldsymbol{D}) = \kappa_{\boldsymbol{\beta}}(\boldsymbol{\beta})\,\exp\Big(-\sum_{\ell=1}^d\int_{\mathbb{Y}} K_\ell(y,\boldsymbol{z}_\ell)\tilde{\mu}_\ell(\mathrm{d}y)\Big)\\ \times\prod_{\ell=1}^d\prod_{\{i:\,\Delta_{\ell,i}=1\}} e^{\boldsymbol{\beta}_\ell\cdot\boldsymbol{z}_{\ell,i}}\int_{\mathbb{Y}} k(X'_{\ell,i};y)\,\tilde{\mu}_\ell(\mathrm{d}y).\end{aligned} \tag{72}$$

From (72) it is apparent that the censored data affect the likelihood only through K_ℓ in the exponential term, whereas the product component is unchanged. Therefore the results of the paper can be easily adapted to deal with this more general setting.

We now consider an application of our model in presence of right–censored survival times. The results are base on 50,000 iterations of the MCMC algorithm with a burn–in period of 20,000 sweeps: the number of iterations allow the convergence of all the MCMC procedures.

Table 2: The estimated values of β_ℓ for the different groups.

	Squamous	small	adeno	large
Standard	-0.9906	-0.6499	-0.8003	-1.1296
Test	-2.2800	-0.6359	-0.7226	-1.3120

A.7.1 Tumor survival data

In this section we refer to a real dataset discussed in [40], who studied 137 advanced lung cancer patients as collected by the Veterans Administration Lung Cancer Study Group. Patients are randomized according to one of two chemotherapeutic agents (standard and test). Tumors are further classified into one of four broad groups (squamous, small, adeno and large). Hence we assume to be provided with $d = 8$ groups of survival times, each one is characterized by the tumor type and the type of treatment. Data include censored observations and a set of covariates, here we consider only the medical status on the normalized scale $[0, 1]$, the others not need to be included in the model as proved in [40]. Hence we have that $d = 8$, $\mathcal{Z} = [0, 1]$ and $\beta_\ell \in \mathbb{R}$, for each $\ell = 1, \dots, 8$. We have set $a = a_0 = 1$, $b = b_0 = 1$ to obtain a prior distribution for each S_i assigning positive probability to survival times in the interval $[0, 1000]$, which contains the data. The base measure P_0 is a uniform distribution over $[0, T]$, being $T = 1000$; finally σ and σ_0 are fixed and equal 0.25. We consider independent priors for each β_ℓ, in particular $\beta_\ell \sim \mathrm{N}(0, 1)$, and we have also modified the MCMC algorithm to update the β_ℓ's from their full conditional distributions.
The estimated survival functions are shown in Figure 4 with the corresponding 95% credible intervals estimated using the conditional algorithm of Section 6.2, and the value of the covariate equals $z = 0.5$. The survival functions have been evaluated using a suitable adaptation of Corollary 5 to this more general framework. It can be seen that the test treatment is better for squamous, but not for the other types of tumors. The values of the covariates are reported in Table 2. The β_ℓ's are negative, as expected, indeed if the medical status increases, then the survival function increases. Moreover the acceptance rates for α and the β_ℓ's of the Metropolis–Hastings step, embedded within the Gibbs sampler, are 39% and $(60\%, 58\%, 57\%, 59\%, 54\%, 57\%, 55\%, 58\%)$. Finally, in order to show that the MCMC does converge, we report here some trace plots. Figure 5 reports trace plots of the β_ℓ's, whereas Figure 6 depicts the two trace plots for α and c_0. All the figures display good mixing properties.

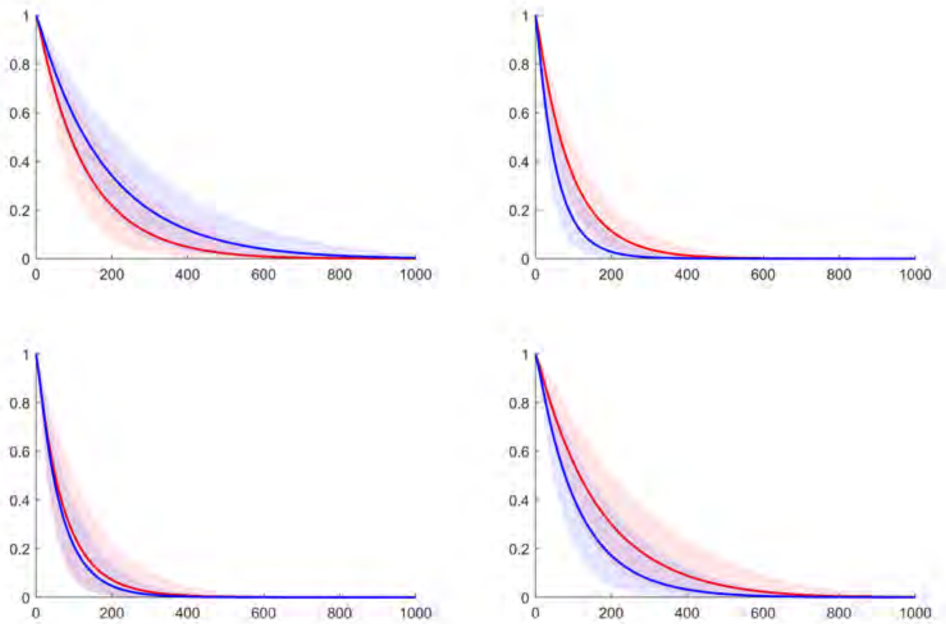


Figure 4: Survival functions for the different types of tumors, from the top on the left we have: squamous, small, adeno and large. In blue the test treatment and in red the standard one when the medical status equals 0.5.

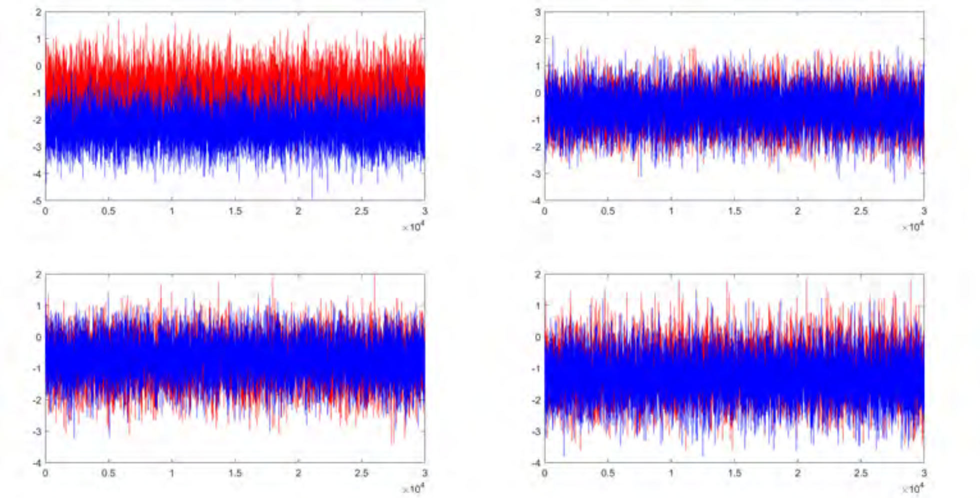
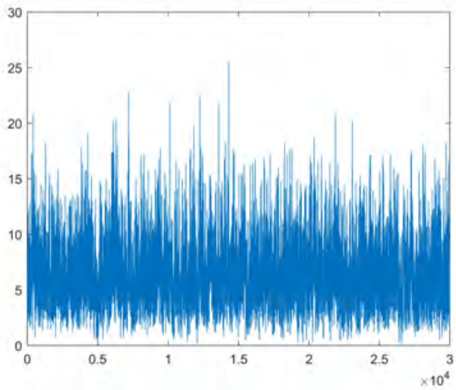
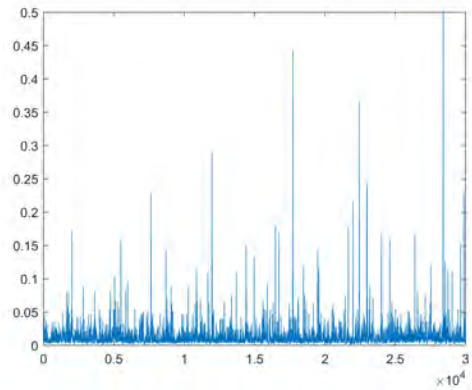


Figure 5: Trace plots of the β_ℓ's, in blue the test treatment and in red the standard one.



(a)



(b)

Figure 6: 6(a): trace plot referring to c_0. 6(b): trace plot referring to α.

References

[1] Barrientos, A.F., Jara, A. and Quintana, F.A. (2012) On the Support of MacEachern's Dependent Dirichlet Processes and Extensions. *Bayesian Anal.* **7**, 277–310.

[2] Bertoin, J. (1999) Subordinators: examples and applications. In *Lectures on probability theory and statistics (Saint-Flour, 1997)*, 1–91, Lecture Notes in Math., 1717, Springer, Berlin.

[3] Billingsley, P. (1995). *Probability and Measure*, 3rd edition. John Wiley & Sons.

[4] Blei, D.M., Ng, A.Y. and Jordan, M.I. (2003) Latent Dirichlet allocation. *J. Mach. Learn. Res.* **3**, 993–1022.

[5] Camerlenghi, F., Lijoi, A. and Prünster, I. (2017) Bayesian prediction with multiple-samples information. *J. Multiv. Anal.* **156**, 18–28.

[6] Camerlenghi, F., Lijoi, A., Orbanz, P. and Prünster, I. (2019) Distribution theory for hierarchical processes. *Ann. Statist.* **47**, 67–92.

[7] Clark, P.K. (1973) A subordinated stochastic process model with finite variance for speculative prices. *Econometrica* **41**, 135–155.

[8] Cont, R. and Tankov, P. (2004) *Financial modelling with jumps.* Chapman & Hall, Boca Raton, FL.

[9] de Finetti, B. (1938) Sur la condition d'equivalence partielle. *Actualités scientiques et industrielles*, **739**, 5–18.

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

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

[12] De Iorio, M., Müller, P., Rosner, G.L. and MacEachern, S.N. (2004) An ANOVA model for dependent random measures. *J. Amer. Statist. Assoc.*, **99**, 205–215.

[13] Donnet, S., Rivoirard, V., Rousseau, J., and Scricciolo, C. (2017) Posterior concentration rates for counting processes with Aalen multiplicative intensities. *Bayesian Anal.* **12**, 53–87.

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

[15] Drăghici, L. and Ramamoorthi, R.V. (2003). Consistency of Dykstra–Laud priors. *Sankhyā* **65** 464–481.

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

[17] Epifani, I. and Lijoi, A. (2010) Nonparametric priors for vectors of survival functions. *Statist. Sinica*, **20**, 1455–1484.

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

[19] Ferguson, T.S. (1974) Prior distributions on spaces of probability measures. *Ann. Statist.* **2**, 615–629.

[20] Ferguson, T.S. and Phadia, E.G. (1979) Bayesian nonparametric estimation based on censored data. *Ann. Statist.*, **7**, 163–186.

[21] Ghosal, S. and van der Vaart. A. (2017) *Fundamentals of nonparametric Bayesian Inference*. Cambridge University Press.

[22] Hardy, M. (2006) Combinatorics of partial derivatives. *Electron. J. Combin.*, **13,** 1–13.

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

[24] 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.

[25] Jacod, J. (1975) Multivariate point processes: predictable projection, Radon-Nikodým derivatives, representation of martingales. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, **31**, 235–253.

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

[27] Kingman, J.F.C. (1993) *Poisson processes.* Oxford University Press, Oxford.

[28] Lijoi, A. and Nipoti, B. (2014) A Class of hazard rate mixtures for combining survival data from different experiments. *J. Amer. Statist. Assoc.*, **109**, 802–814.

[29] Lijoi, A., Nipoti, B. and Prünster, I. (2014) Bayesian inference with dependent normalized completely random measures. *Bernoulli* **20**, 1260–1291.

[30] Lijoi, A. and Prünster, I. (2010) Models beyond the Dirichlet process. In *Bayesian Nonparametrics* (Hjort, N.L., Holmes, C.C, Müller, P., Walker, S.G., eds.), Cambridge Univ. Press, Cambridge, pp. 80–136.

[31] Lo, A.Y. and Weng, C.-S. (1989) On a class of Bayesian nonparametric estimates. II. Hazard rate estimates. *Ann. Inst. Statist. Math.*, **41**, 227–245.

[32] Mandelbrot, B. and Taylor, H.M. (1967) On the distribution of stock price differences. *Oper. Res.*, **15**, 1057–1062.

[33] Müller, P, Quintana, F.A., Jara, A., and Hanson, T. (2015) *Bayesian nonparametric data analysis*. Springer, Berlin.

[34] Müller, P, Quintana, F.A. and Page, G.L. (2018) Nonparametric Bayesian inference in applications. (With discussion). *Statist. Meth. Appl.* **27**, 175–206.

[35] Müller, P., Quintana, F.A., Rosner, G.L. and Maitland, M.L. (2013) Bayesian inference for longitudinal data with non-parametric treatment effects. *Biostatistics*, **15**, 341–352.

[36] Nieto, L.E. and Walker, S.G. (2004) Bayesian nonparametric survival analysis via Lévy driven Markov processes. *Statist. Sinica* **14**, 1127–1146.

[37] Nieto, L.E. and Walker, S.G. (2005) A semi-parametric Bayesian analysis of survival data based on Lévy-driven processes. *Lifetime Data Anal.* **11**, 529–543.

[38] Nieto, L.E. (2014) Bayesian semiparametric analysis of short- and long-term hazard ratios with covariates. *Comput. Statist. Data Anal.* **71**, 477–490.

[39] Peccati, G. and Prünster, I. (2008) Linear and quadratic functionals of random hazard rates: an asymptotic analysis. *Ann. Appl. Probab.* **18**, 1910–1943.

[40] Prentice, R.L. (1973) Exponential Survivals with Censoring and Explanatory Variables. *Biometrika* **60**, 279–288.

[41] Sato, K. (2013) *Lévy processes and infinitely divisible distributions.* Cambridge University Press, Cambridge.

[42] Teh, Y.W., Jordan, M.I., Beal, M.J. and Blei, D.M. (2006) Hierarchical Dirichlet processes. *J. Amer. Statist. Assoc.* **101**, 1566–1581.

[43] Teh, Y.W. and Jordan, M.I. (2010) Hierarchical Bayesian nonparametric models with applications. In *Bayesian Nonparametrics* (Hjort, N.L., Holmes, C.C, Müller, P., Walker, S.G., eds.), Cambridge Univ. Press, Cambridge, pp. 158–207.

[44] Wolpert, R.L. and Ickstadt, K.. Simulation of Lévy random fields. In *Practical nonparametric and semiparametric Bayesian statistics*, pp. 227–242. Springer, New York, 1998.

[45] Zhou, H., Hanson, T., Jara, A. and Zhang, J. (2015) Modeling county level breast cancer survival data using a covariate–adjusted frailty proportional hazards model. *Ann. Appl. Stat.* **9**, 43–68.