Bayesian nonparametric estimation of survival functions with multiple-samples information

Alan Riva Palacio[†] and Fabrizio Leisen[*]

School of Mathematics, Statistics and Actuarial Sciences, University of Kent
Cornwallis Building, Canterbury, Kent CT2 7NF
e-mail: ar515@kent.ac.uk; fabrizio.leisen@gmail.com

Abstract: In many real problems, dependence structures more general than exchangeability are required. For instance, in some settings partial exchangeability is a more reasonable assumption. For this reason, vectors of dependent Bayesian nonparametric priors have recently gained popularity. They provide flexible models which are tractable from a computational and theoretical point of view. In this paper, we focus on their use for estimating multivariate survival functions. Our model extends the work of Epifani and Lijoi (2010) to an arbitrary dimension and allows to model the dependence among survival times of different groups of observations. Theoretical results about the posterior behaviour of the underlying dependent vector of completely random measures are provided. The performance of the model is tested on a simulated dataset arising from a distributional Clayton copula.

MSC 2010 subject classifications: Primary 62F15, 60G57; secondary 60G51.
Keywords and phrases: Bayesian nonparametrics, Survival analysis, Dependent completely random measures.

Contents

1. Introduction

Bayesian nonparametric modelling in *survival analysis* problems often relies on the assumption that the times observed are *exchangeable*, see for example Doksum (1974) and Ishwaran and James (2009). Such assumption fails to hold when we consider events that are pooled from different dependent scenarios. For example, consider patients under the same treatment but in different hospitals. The survival times of patients from the same hospital could be assumed exchangeable. On the other hand, this is not a reasonable assumption when we consider patients from different hospitals since factors specific to each hospital might exert significant influence. In general, we can consider that the data is originated from d different but related studies. Formally, we have d sets of observations where the exchangeability assumption is assumed only within each set. In the above cases, it would be more appropriate to assume a form of dependence called *partial exchangeability* (see Section 2 for a formal account on exchangeability and partial exchangeability). This motivates the extension of Bayesian nonparametric models into a partially exchangeable setting where multiple-samples information could be used.

Applications of Bayesian nonparametrics in survival analysis go back, for example, to Doksum (1974) and Ferguson and Phadia (1979), who used non-decreasing independent increment processes to construct random survival functions. Dykstra and Laud (1981) and Lo and Weng (1989) focused on random hazard rates. More recently, Ishwaran and James (2009) used a general class of random hazard rate-based models, and Nieto-Barajas (2014) used a general short-term and long-term hazard ratios model. There is an ongoing effort in Bayesian nonparametrics to propose flexible dependent random probability measures as set forth with the seminal work of MacEachern (1999). In survival analysis, for example, De Iorio et al. (2009) introduced a model based on a dependent Dirichlet process. In a partial exchangeable setting, survival analysis models have been used, for example, in Epifani and Lijoi (2010) where a dependent two-dimensional extension of the *neutral to the right* (NTR) model was introduced and in Lijoi and Nipoti (2014) where a dependent vector of hazard rates was constructed. Griffin and Leisen (2017) introduced a new class of vectors of dependent completely random measures, called *Compound Random Measures*, where the dependence contribution is modelled with a parametric distribution.

In the seminal work of Doksum (1974), the NTR model for survival functions was introduced. The NTR model can be expressed in terms of a Completely Random Measure (CRM) μ. This means that when μ is evaluated at pairwise disjoint sets it gives rise to mutually independent nonnegative random variables. We say that a positive random variable Y has a NTR distribution given by a CRM μ, denoted $Y \sim \mathrm{NTR}(\mu)$, if

$$S(t) = \mathbb{P}[Y > t \,|\, \mu] = e^{-\mu(0,t]},$$

where μ is such that

$$\lim_{t \to \infty} \mu(0,t] = \infty.$$

NTR distributions have several appealing properties, including the independence of normalized increments and posterior conjugacy for censored to the right data. An extension of the NTR model into a partially exchangeable setting was given by Epifani and Lijoi (2010) for the $2-$dimensional case. In the present work, we follow the approach of Epifani and Lijoi (2010) and focus on models based on a d-dimensional *vector of completely random measures* (VCRM). More precisely, we consider d collections of survival times $\{Y_j^{(1)}\}_{j=1}^{\infty}, \ldots, \{Y_j^{(d)}\}_{j=1}^{\infty}$ such that, for $\boldsymbol{t} = (t_1, \ldots, t_d) \in (\mathbb{R}^+)^d$,

$$S(\boldsymbol{t}) = \mathbb{P}\Big[Y_{i_1}^{(1)} > t_1, \ldots, Y_{i_d}^{(d)} > t_d \,|\, (\mu_1, \ldots, \mu_d)\Big] = e^{-\mu_1(0,t_1] - \cdots - \mu_d(0,t_d]}, \quad (1)$$

with arbitrary $i_1, \ldots, i_d \in \mathbb{N} \setminus \{0\}$. This model is convenient for modeling data where the dependence among the entries of the VCRM $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_d)$ accounts for dependence among the multiple-samples in a partially exchangeable setting. Furthermore, marginally we recover the NTR model, namely

$$Y_1^{(i)}, \ldots, Y_{n_i}^{(i)} \overset{\text{i.i.d.}}{\sim} \mathrm{NTR}(\mu_i) \quad (2)$$

with $i \in \{1, \ldots, d\}$, $n_i \in \mathbb{N} \setminus \{0\}$. In (2) we want to model the dependence of the VCRM $\boldsymbol{\mu}$ in a way that allows us to fix a marginal behavior so to exploit the fact that marginally we recover a NTR model; *Lévy copulas* are a natural framework to model the dependence structure of VCRM's in such way.

In this paper we provide a posterior characterization for the above model, see Theorem 1. Similarly to Epifani and Lijoi (2010) for 2-dimensional setting, we show that the posterior distribution corresponds to a survival function of the type as in (1) leading to a conjugacy property. Extensions of some results in Epifani and Lijoi (2010) are also provided. We would like to stress that the derivation of such results are not trivial when considering an arbitrary dimension. In particular, Proposition 1 gives a general expression for the Laplace exponent when a Lévy copula is considered to set the dependence of the VCRM underlying the $d-$dimensional NTR model; Proposition 3 gives an alternative characterization of the multivariate NTR. Furthermore, other theoretical results are proved in order to facilitate the calculation of posterior means when the inferential exercise is implemented. Finally, we illustrate the methodology on a synthetic dataset.

The paper is organized as follows: Section 2 presents the preliminary notions which are needed in this work. In Section 3 we extend some results in Epifani and Lijoi (2010) to the multivariate setting. In particular, we state the posterior characterization of the model and provide some useful corollaries for implementing the posterior inference. In Section 4, an application with synthetic data is illustrated. All the proofs can be found in the appendix.

2. Preliminaries

In this section, we provide some preliminaries about exchangeability, partial exchangeability and vectors of completely random measures which are the building blocks of our Bayesian nonparametric proposal. Furthermore, we will illustrate the concept of a positive Lévy copula which is useful to model the dependence structure between the components of a vector of completely random measures.

2.1. Exchangeability and partial exchangeability

Let $\mathbb{Z}$ be a complete and separable metric space, with corresponding Borel σ-algebra $\mathcal{Z} = \mathcal{B}(\mathbb{Z})$

Definition 1. A collection of random variables $\{Z_i\}_{i=1}^\infty$ in $\mathbb{Z}$ is exchangeable if for any permutation π of $\{1, \ldots, m\}$ we have that

$$\{Z_1, \ldots, Z_m\} \stackrel{\mathrm{d}}{=} \{Z_{\pi(1)}, \ldots, Z_{\pi(m)}\}.$$

As highlighted in the Introduction, in several problems the exchangeability assumption appears far too restrictive. In particular, we considered d groups of observations where the order in which they are collected within each group is irrelevant. To describe this setting we resorted to the notion of partial exchangeability, as set forth by de Finetti (1938), that formalizes the idea of partitioning the entire set of observations into a certain number of classes, say d, in such a way that exchangeability may be reasonably assumed within each class. For ease of exposition, we confine ourselves to consider the case where d = 2.

Definition 2. The collection of random vectors

$$\left\{\left(Z_i^{(1)}, Z_i^{(2)}\right)\right\}_{i=1}^\infty$$

in $\mathbb{Z}^2$ is partially exchangeable if, for any $m_1, m_2 \geq 1$ and for all permutations π_1 and π_2 of $\{1, \ldots, m_1\}$ and $\{1, \ldots, m_2\}$ respectively, we have that

$$\{Z_1^{(1)}, \ldots, Z_{m_1}^{(1)}, Z_1^{(2)}, \ldots, Z_{m_2}^{(2)}\} \stackrel{\mathrm{d}}{=} \{Z_{\pi_1(1)}^{(1)}, \ldots, Z_{\pi_1(m_1)}^{(1)}, Z_{\pi_2(1)}^{(2)}, \ldots, Z_{\pi_2(m_2)}^{(2)}\}.$$

2.2. Vectors of completely random measures

Given a complete and separable metric space $\mathbb{X}$, with corresponding Borel σ-algebra $\mathcal{X} = \mathcal{B}(\mathbb{X})$, we call a measure μ on $(\mathbb{X}, \mathcal{X})$ boundedly finite if $\mu(A) < \infty$ for any bounded set $A \in \mathcal{X}$. A random measure is a measurable function from a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ onto $(\mathbb{M}_{\mathbb{X}}, \mathcal{M}_{\mathbb{X}})$ which is the measure space formed by $\mathbb{M}_{\mathbb{X}}$, the space of boundedly finite measures on $(\mathbb{X}, \mathcal{X})$, and its corresponding Borel σ-algebra $\mathcal{M}_{\mathbb{X}}$. In particular we will focus on the class of *completely random measures* as introduced in Kingman (1967).

Definition 3. A random measure μ on a complete and separable metric space $\mathbb{X}$ with corresponding Borel σ-algebra $\mathcal{X} = \mathcal{B}(\mathbb{X})$ is called a completely random measure (CRM) if for any collection of disjoint sets $\{A_1, \ldots, A_n\} \subset \mathcal{X}$ the random variables $\mu(A_1), \ldots, \mu(A_n)$ are mutually independent.

A CRM μ has the following representation (Kingman, 1967),

$$\mu = \mu_d + \mu_r + \mu_{fl},$$

where μ_d is a deterministic measure, μ_{fl} is a measure that consists on jumps with possibly random jump heights but fixed jump locations, and

$$\mu_r = \sum_{i=1}^{\infty} W_i \delta_{X_i},$$

where for $i \in \{1, 2, \ldots\}$ $X_i \in \mathbb{X}$ are random jump locations and $W_i \in \mathbb{R}^+$ are random jump heights. The measures μ_d, μ_{fl} and μ_r are mutually independent. In particular, μ_r is again a CRM and is characterized by the following Laplace transform

$$\mathbb{E}\left[e^{-\lambda \mu_r(A)}\right] = e^{-\int_{\mathbb{R}^+ \times A}(1 - e^{-\lambda s})\nu(\mathrm{d}s, \mathrm{d}x)}, \tag{3}$$

where $\lambda > 0$ and ν is a measure on $\mathbb{R}^+ \times \mathbb{X}$ such that

$$\int_{\mathbb{R}^+ \times A} \min\{s, 1\} \nu(\mathrm{d}s, \mathrm{d}x) < \infty,$$

for any bounded set $A \in \mathcal{X}$. The measure ν is usually called the Lévy intensity of μ_r. In the remainder of this work we only consider CRM's μ without fixed jump locations nor deterministic part so we take $\mu = \mu_r$ to be solely determined by (3). In particular we focus on Lévy intensities ν which are *homogeneous*, i.e.

$$\nu(\mathrm{d}s, \mathrm{d}x) = \rho(\mathrm{d}s)\alpha(\mathrm{d}x),$$

where α is a non-atomic measure on $\mathbb{X}$ referring to the jump locations and ρ is a measure on $\mathbb{R}^+$ referring to the jump heights. A popular example of an homogeneous CRM is the σ-stable process given by

$$\nu(\mathrm{d}s, \mathrm{d}x) = \frac{A\sigma s^{-1-\sigma}}{\Gamma(1-\sigma)} \mathrm{d}s\alpha(\mathrm{d}x). \tag{4}$$
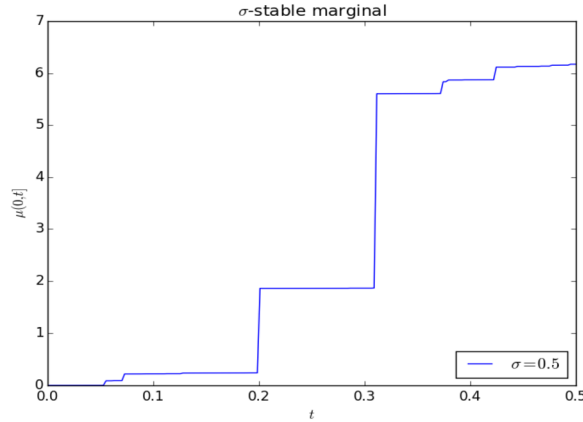
FIG 1. *Plot of $\mu(0, t]$ when a σ-stable process is considered.*

As an illustration, we plot in Figure 1 the associated process $\mu(0, t]$ for the σ-stable process (4) with $\alpha(\mathrm{d}x) = \mathrm{d}x$.

We extend this framework to the multivariate setting by considering vectors $(\mu_1, \ldots, \mu_d)$ where each μ_i is a homogeneous CRM on $(\mathbb{X}, \mathcal{X})$ with respective Lévy intensities $\bar{\nu}_j(\mathrm{d}s, \mathrm{d}x) = \nu_j(\mathrm{d}s)\alpha(\mathrm{d}x)$. Moreover we take the intensity α to be smooth in the sense that $\alpha((0, t]) = \gamma(t)$ with $\gamma : [0, \infty) \to \mathbb{R}^+$ a non-decreasing and differentiable function such that $\gamma(0) = 0$ and $\lim_{t \to \infty} \gamma(t) = \infty$; this last conditions on the limit behaviour will enable us to get, marginally, the associated NTR cumulative distributions in our models. We have that for any $A_1, \ldots, A_n$ in $\mathcal{X}$, with $A_i \cap A_j = \emptyset$ for any $i \neq j$, the random vectors $(\mu_1(A_i), \ldots, \mu_d(A_i))$ and $(\mu_1(A_j), \ldots, \mu_d(A_j))$ are mutually independent; furthermore, one has a multivariate analogue of the Laplace transform (3)

$$\mathbb{E}\left[e^{-\lambda_1 \mu_1(A) - \cdots - \lambda_d \mu_d(A)}\right] = e^{-\int_{(\mathbb{R}^+)^d \times A} (1 - e^{-\lambda_1 s_1 - \cdots - \lambda_d s_d})\rho_d(\mathrm{d}s_1, \ldots, \mathrm{d}s_d)\alpha(\mathrm{d}x)}, \quad (5)$$

where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_d) \in (\mathbb{R}^+)^d$ and ρ_d is a measure on $(\mathbb{R}^+)^d$. In particular, we introduce the notation for the multivariate Laplace transform

$$\mathbb{E}\left[e^{-\lambda_1 \mu_1(0, t] - \cdots - \lambda_d \mu_d(0, t]}\right] = e^{-\psi_t(\boldsymbol{\lambda})}. \quad (6)$$

Henceforth, $\psi_t(\boldsymbol{\lambda})$ is called the Laplace exponent of $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_d)$; in the case at hand, $\psi_t(\boldsymbol{\lambda}) = \gamma(t)\psi(\boldsymbol{\lambda})$ where $\psi(\boldsymbol{\lambda}) = \int_{(\mathbb{R}^+)^d}(1 - e^{-<\boldsymbol{\lambda}, \boldsymbol{s}>})\rho_d(\mathrm{d}\boldsymbol{s})$ and $<\boldsymbol{\lambda}, \boldsymbol{s}> = \sum_{i=1}^{d} \lambda_i s_i$ is the usual inner product in $\mathbb{R}^d$. Marginalizing, we have that

$$\nu_i(A) = \int_A \nu_i(\mathrm{d}s) = \int_{(\mathbb{R}^+)^{d-1}} \rho_d(\mathrm{d}s_1, \ldots, \mathrm{d}s_{i-1}, A, \mathrm{d}s_{i+1}, \ldots, \mathrm{d}s_d).$$

In Section 3, we use this particular kind of homogeneous and additive vector of CRM's to construct priors for survival analysis models.

2.3. Positive Lévy copulas

Although in this work we consider vectors of CRM's with fixed marginal behaviour, it remains to establish the dependence structure. Kallsen and Tankov (2006) introduced the concept of positive Lévy copulas which allows to construct vectors of CRM's with fixed marginals.

Definition 4. A function $\mathcal{C}(\boldsymbol{s} = (s_1, \ldots, s_d)) : [0, \infty)^d \to [0, \infty]$ is a positive Lèvy copula if

(1) $\forall\, B = [s_1, t_1] \times \cdots \times [s_d, t_d] \subset [0, \infty)^d$ such that $s_1 \leq t_1, \ldots, s_d < t_d$ we have that
$$\sum_{\{\boldsymbol{v}\,:\,\boldsymbol{v} \text{ is a vertex of B}\}} \mathrm{sign}(\boldsymbol{v})\mathcal{C}(\boldsymbol{v}) \geq 0,$$
with
$$\mathrm{sign}(\boldsymbol{v}) = \begin{cases} 1, & \text{if } v_k = s_k \text{ for an even number of vertices,} \\ -1, & \text{if } v_k = s_k \text{ for an odd number of vertices.} \end{cases}$$

(2) If $\boldsymbol{s}$ is such that $s_i = 0$ for some $i \in \{1, \ldots, d\}$ then $\mathcal{C}(\boldsymbol{s}) = 0$.
(3) Let $y_1 = \cdots = y_{k-1} = y_{k+1} = \cdots = y_d = \infty$ and
$C_k(s) = \mathcal{C}(y_1, \ldots, y_{k-1}, s_k, y_{k+1}, \ldots, y_d)$ for $k \in \{1, \ldots, d\}$ then $C_k(s) = s$.

For example, a vector of independent CRM's is obtained with
$$\mathcal{C}_{\perp,d}(\boldsymbol{s}) = s_1 \mathbf{1}_{s_2=\infty,\ldots,s_d=\infty} + \cdots + s_d \mathbf{1}_{s_1=\infty,\ldots,s_{d-1}=\infty}.$$

A vector of completely dependent CRM's, in the sense that the jumps of the stochastic vector are in a set S such that whenever $\boldsymbol{v}, \boldsymbol{u} \in S$ then either $v_i < u_i$ or $u_i < v_i$ for all $i \in \{1, \ldots, d\}$, is obtained with
$$\mathcal{C}_{\parallel,d}(\boldsymbol{s}) = \min\{s_1, \ldots, s_d\}.$$

An interesting example of positive Lévy copulas is the Clayton Lévy copula
$$\mathcal{C}_{\theta,d}(\boldsymbol{s}) = (s_1^{-\theta} + \cdots + s_d^{-\theta})^{-\frac{1}{\theta}}. \tag{7}$$

The parameter θ is positive and regulates the level of dependence. The above copulas are special cases of the Clayton Lévy copula, i.e.
$$\lim_{\theta \to 0} \mathcal{C}_{\theta,d}(\boldsymbol{s}) = \mathcal{C}_{\perp,d}(\boldsymbol{s}) \text{ and } \lim_{\theta \to \infty} \mathcal{C}_{\theta,d}(\boldsymbol{s}) = \mathcal{C}_{\parallel,d}(\boldsymbol{s}).$$

We define the tail integral of an univariate Lévy intensity ν to be $U(x) = \int_x^\infty \nu(s)\mathrm{d}s$. In the setting of Section **2.1** we use a Lévy copula $\mathcal{C}_d$ and the marginal tail integrals $U_1, \ldots, U_d$ associated to $\nu_1, \ldots, \nu_d$ to specify an absolutely continuous $\rho_d(\mathrm{d}\boldsymbol{s}) = \rho_d(\boldsymbol{s})\mathrm{d}\boldsymbol{s}$ via

$$\begin{aligned} U(\boldsymbol{x}) &= \int_{x_1}^\infty \cdots \int_{x_d}^\infty \rho_d(\boldsymbol{s})\mathrm{d}\boldsymbol{s} \\ &= \int_{x_1}^\infty \cdots \int_{x_d}^\infty \frac{\partial^d}{\partial u_1 \cdots \partial u_d} \mathcal{C}_d(\boldsymbol{u}) \Big|_{u_1=U_1(s_1),\cdots u_d=U_d(s_d)} \nu_1(s_1) \cdots \nu_d(s_d)\mathrm{d}\boldsymbol{s}. \end{aligned}$$
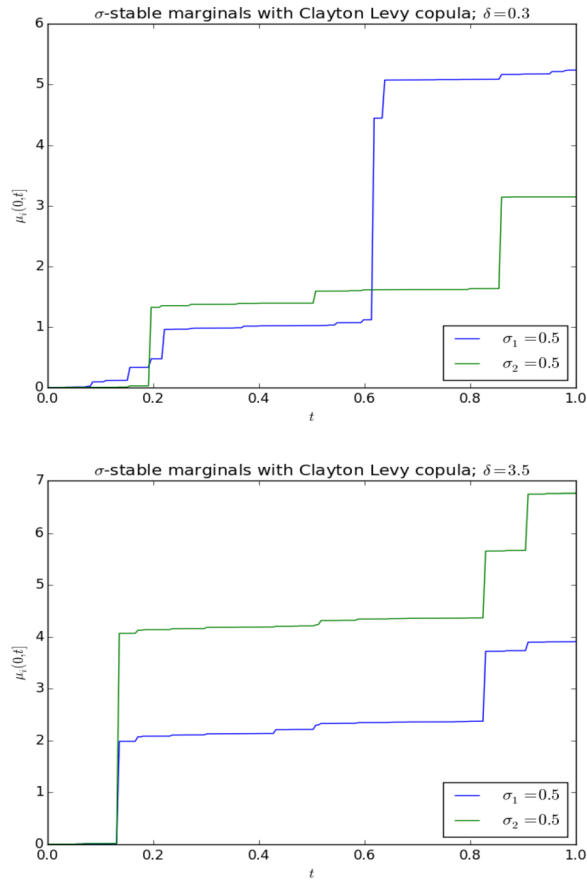
FIG 2. *Plot of dependent σ-stable processes with dependence given by Clayton Lévy copula with parameter $\theta = 0.3$ (top) and $\theta = 3.5$ (bottom).*

Therefore, under suitable regularity conditions, we can recover the multivariate Lévy intensity from the copula and marginal intensities in the following way

$$\rho_d(\boldsymbol{s}) = \frac{\partial^d}{\partial u_1 \cdots \partial u_d} \mathcal{C}_d(\boldsymbol{u}) \Bigg|_{u_1 = U_1(s_1), \cdots, x_d = U_d(s_d)} \nu_1(s_1) \cdots \nu_d(s_d). \qquad (8)$$

For example, consider the Clayton Lévy copula with σ-stable margins, given by (4), and $\alpha(\mathrm{d}x) = \mathrm{d}x$. Figure 2 shows the dependence behaviour when a 2-dimensional Clayton Lévy copula with parameter $\theta = 0.3$ and $\theta = 3.5$ is employed; we plot the associated stochastic processes $\mu_i(0,t]$ with $i \in \{1,2\}$ similarly to Figure 1. As expected, when $\theta = 0.3$, at each jumping time, the processes have one jump weight big and one small since we are close to the independence case (where the processes almost surely share no jumping times).

On the other hand, when θ is increased to 3.5, we can appreciate the higher dependence induced by a larger value of the copula parameter. We simulated the trajectories in Figure 2 by using Algorithm 6.15 in Cont and Tankov (2004), where a full treatment of the dependence structure of Lévy intensities is also given. Leisen and Lijoi (2011), Leisen, Lijoi and Spano (2013) and Zhu and Leisen (2015) used a Lévy copula approach for building vectors of dependent completely random measures.

2.3.1. Working example

If we consider the Lévy intensity arising from (8) when considering the d-dimensional Clayton Lévy copula, (7), with parameter θ and σ-stable marginals, (4), with parameters A, σ, we obtain

$$\rho_{d,\theta,A,\sigma}(\boldsymbol{s}) = \frac{A(1+\theta)(1+2\theta)\cdots(1+(d-1)\theta)\sigma^d \left(s_1 s_2 \cdots s_d\right)^{\sigma\theta-1}}{\Gamma(1-\sigma)\left(s_1^{\sigma\theta} + \cdots + s_d^{\sigma\theta}\right)^{\frac{1}{\theta}+d}}.$$

Furthermore, if we take $\theta = 1/\sigma$ we obtain the simplified Lévy intensity

$$\rho_{d,A,\sigma}(\boldsymbol{s}) = \frac{A(\sigma+1)(\sigma+2)\cdots(\sigma+d-1)\sigma}{\Gamma(1-\sigma)\left(s_1 + \cdots + s_d\right)^{\sigma+d}}. \tag{9}$$

Such intensity corresponds to a particular family of vectors of completely random measures known as *Compound Random Measures* (CoRM's) and introduced in Griffin and Leisen (2017); the previous Lévy intensity arises when taking $\phi = 1$ in equation (4.4) of the aforementioned paper. A convenient feature of this Lévy intensity is that, as shown in Proposition 3.1 of Zhu and Leisen (2015), we can explicitly get the corresponding Laplace exponent

$$\psi_{d,A,\sigma}(\boldsymbol{\lambda}) = \sum_{i=1}^{d} \frac{\lambda_i^{\sigma+d-1}}{\prod_{j=1,\,j\neq i}^{d}(\lambda_i - \lambda_j)}; \quad \lambda_i \neq \lambda_j \text{ for } j \neq i, \tag{10}$$

where we take the appropriate limits when $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_d)$ is such that $\lambda_i = \lambda_j$ for distinct $i, j \in \{1, \ldots, d\}$. As indicated in the remark at the end of Section 3, evaluation of the Laplace exponent is necessary for the explicit calculation of the posterior mean of the survival function given censored data.

3. Main results

Let $d \in \mathbb{N} \setminus \{0\}$, and suppose we have d collections of random variables

$$\{\{Y_j^{(i)}\}_{j=1}^{\infty}\}_{i=1}^{d}. \tag{11}$$

We characterize the probability distribution of these random variables in terms of a vector of CRM's $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_d)$. For $\boldsymbol{t} = (t_1, \ldots, t_d) \in (\mathbb{R}^+)^d$, let

$$\mathbb{P}\Big[Y_1^{(1)} > t_{1,1}, \ldots, Y_{n_1}^{(1)} > t_{1,n_1}, \ldots, Y_1^{(d)} > t_{d,1}, \ldots, Y_{n_d}^{(d)} > t_{d,n_d} \,|(\mu_1, \ldots, \mu_d)\Big]$$
$$= \prod_{i=1}^{d} \prod_{j=1}^{n_i} e^{-\mu_i(0,t_{i,j}]}. \tag{12}$$

We observe that under such model the random variables (12) are partially exchangeable and marginally follow a NTR process. The dependence structure in this model can be given through the Lévy copula associated to the CRM $\boldsymbol{\mu}$. This model extends the one in Epifani and Lijoi (2010) to an arbitrary dimension d.

The family of Clayton Lévy copulas is of interest because it has both the independence and complete dependence cases as limit behaviour. In the next result, we work towards finding expressions for the Laplace exponent associated to the Clayton family in such a way that the dependence structure is decoupled across dimensions. This result could be useful since, as we will see, an explicit calculation of ψ is of key importance to implement the Bayesian inference in our survival analysis model.

Let $\rho_d(\boldsymbol{s}; \theta)$ be the Lévy intensity associated via (8) to the Clayton Lévy copula $\mathcal{C}_{\theta,d}$ and fixed marginal Lévy intensities $\nu_1, \ldots, \nu_d$ with corresponding Laplace transforms $\psi_1, \ldots, \psi_d$. We denote the vector of tail integrals corresponding to the marginal Lévy intensities as $\boldsymbol{U}_d(\boldsymbol{x}) = (U_1(x_1), \ldots, U_d(x_d))$ and fix the notation

$$\kappa(\theta; \boldsymbol{\lambda}, \boldsymbol{i}) = \lambda_{i_1} \cdots \lambda_{i_m} \int_{(\mathbb{R}^+)^m} \mathrm{e}^{-\lambda_{i_1} s_1 - \cdots - \lambda_{i_m} s_m} C_{\theta,m}(U_{i_1}(s_1), \ldots, U_{i_m}(s_m)) \mathrm{d}\boldsymbol{s},$$

where $d \in \mathbb{N} \setminus \{0\}$, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_d) \in (\mathbb{R}^+)^d$, $m \in \{1, \ldots, d\}$, and $\boldsymbol{i} = (i_1, i_2, \ldots, i_m) \in \{1, \ldots, d\}^m$ is such that $i_1 < \cdots < i_m$.

Proposition 1. *Suppose that $d \in \{2, 3, \ldots\}$ and*

$$\int_{\|\boldsymbol{s}\| \leq 1} \|\boldsymbol{s}\| \rho_d(\boldsymbol{s}; \theta) \mathrm{d}\boldsymbol{s} < \infty, \tag{13}$$

then

$$\psi(\boldsymbol{\lambda}) = \int_{(\mathbb{R}^+)^d} (1 - \mathrm{e}^{-<\boldsymbol{\lambda}, \boldsymbol{s}>}) \frac{\partial^d}{\partial u_d \cdots \partial u_1} C_{\theta,d}(\boldsymbol{u})\big|_{\boldsymbol{u}=\boldsymbol{U}_d(\boldsymbol{s})} \nu_1(s_1) \cdots \nu_d(s_d) \mathrm{d}\boldsymbol{s}$$

$$= \sum_{i=1}^{d} \psi_i(\lambda_i) - \sum_{\substack{\boldsymbol{i}=(i_1,i_2)\in\{1,\ldots,d\}^2 \\ i_1<i_2}} \kappa(\theta; \boldsymbol{\lambda}, \boldsymbol{i}) + \cdots$$

$$\cdots + (-1)^d \sum_{\substack{\boldsymbol{i}=(i_1,\ldots,i_{d-1})\in\{1,\ldots,d\}^{d-1} \\ i_1<\cdots<i_{d-1}}} \kappa(\theta; \boldsymbol{\lambda}, \boldsymbol{i}) + (-1)^{d+1} \kappa(\theta; \boldsymbol{\lambda}, (1, \ldots, d)),$$

where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_d) \in (\mathbb{R}^+)^d$.

We refer to the Appendix A.1 for the proof. We incorporate the Lévy exponent ψ in the multivariate survival analysis setting of (12), in the next result. We introduce the notation

$$\nu_{i_1,\ldots,i_h}(s_{i_1},\ldots,s_{i_h}) = \int_0^\infty \cdots \int_0^\infty \rho_d(\boldsymbol{s}) \prod_{j \notin \{i_1,\ldots,i_h\}} \mathrm{d}s_j$$

for $h \in \{1,\ldots,d\}$ and distinct $i_1,\ldots,i_h \in \{1,\ldots,d\}$; and denote $\psi_{i_1,\cdots,i_h}$ for the respective Laplace exponents.

Proposition 2. *In the context of (12), let $\mathbf{1} = (1,\ldots,1)$. For $t_1 \leq \cdots \leq t_d$ and $i_1,\ldots,i_d \in \{1,\ldots,d\}$ such that $t_{i_1} \leq \cdots \leq t_{i_d}$ then*

$$\mathbb{P}\Big[Y^{(1)} > t_1,\ldots,Y^{(d)} > t_d\Big] = \mathrm{e}^{-\gamma(t_{i_1})\psi(\mathbf{1})}\mathrm{e}^{-[\gamma(t_{i_2})-\gamma(t_{i_1})]\psi_{i_2,\ldots,i_d}(\mathbf{1})} \qquad (14)$$

$$\times \cdots \mathrm{e}^{-[\gamma(t_{i_d})-\gamma(t_{i_{d-1}})]\psi_{i_d}(\mathbf{1})}. \qquad (15)$$

We refer to the Appendix A.2 for the proof. This result showcases the importance of the Laplace exponent ψ for calculating probabilities in the model and the impact of the function $\gamma(t)$, related to the time depending part of the Laplace exponent, in the survival function. In Section 4, we will show that the availability of the Laplace exponent is also of main importance to implement the Bayesian inference for the model. The model we are working on generalizes to arbitrary dimension the classic model of Doksum (1974). We present a multivariate extension of Theorem 3.1 in Doksum (1974), which relates our model with the notion of neutrality to the right. Let F be a d-variate random distribution function on $(\mathbb{R}^+)^d$ and, for a d-variate vector of CRM's $\boldsymbol{\mu} = (\mu_1,\ldots,\mu_d)$, denote $\mu_i(t) = \mu_i((0,t])$ with $i \in \{1,\ldots,d\}$. Then, we have the following multivariate extension to Theorem 3.1 in Doksum (1974) and Proposition 4 in Epifani and Lijoi (2010).

Proposition 3. $F(\boldsymbol{t} = (t_1,\ldots,t_d))$ *has the same distribution as*

$$[1 - \mathrm{e}^{-\mu_1(t_1)}] \cdots [1 - \mathrm{e}^{-\mu_d(t_d)}]$$

for some d-variate CRM $\boldsymbol{\mu} = (\mu_1,\ldots,\mu_d)$ if and only if for $h \in \{1,2,\ldots\}$ and vectors $\boldsymbol{t}_1 = (t_{1,1},\ldots,t_{d,1}),\ldots,\boldsymbol{t}_h = (t_{1,h},\ldots,t_{d,h})$ with $t_{0,i} = 0 < t_{1,i} < \cdots < t_{d,i}$ and $t_{j,0} = 0 < t_{j,1} < \cdots < t_{j,h}$, there exists h independent random vectors $(V_{1,1},\ldots V_{d,1}),\ldots,(V_{1,h},\ldots V_{d,h})$ such that

$$(F(\boldsymbol{t}_1),\ldots,F(\boldsymbol{t}_h)) \overset{d}{=}$$

$$\left(V_{1,1}\cdots V_{d,1}, [1 - \bar{V}_{1,1}\bar{V}_{1,2}]\cdots[1 - \bar{V}_{d,1}\bar{V}_{d,2}],\ldots,[1 - \prod_{j=1}^h \bar{V}_{1,j}]\cdots[1 - \prod_{j=1}^h \bar{V}_{d,j}]\right)$$

$$(16)$$

Where $\bar{V}_{i,j} = 1 - V_{i,j}$ with $i \in \{1,\ldots,d\}$ and $j \in \{1,\ldots,h\}$.

We refer to the Appendix A.3 for the proof. We now establish some notation in order to address the posterior distribution arising from (12) when some survival data is available. Let $\boldsymbol{Y}_{n_i}^{(i)} = \left(Y_1^{(i)}, \ldots, Y_{n_i}^{(i)}\right)$, $i = 1, \ldots, d$, be d groups of observations that come from the distribution given by

$$\mathbb{P}\left[\boldsymbol{Y}_{n_1}^{(1)} > \boldsymbol{t}_{1,n_1}, \ldots, \boldsymbol{Y}_{n_d}^{(d)} > \boldsymbol{t}_{d,n_d} \,|\, (\mu_1, \ldots, \mu_d)\right] = \prod_{i=1}^{d} \prod_{j=1}^{n_i} \mathrm{e}^{-\mu_i(0, t_{i,j})},$$

where $\boldsymbol{t}_{i,n_i} = (t_{i,1}, \ldots, t_{i,n_i})$ and the event $\{\boldsymbol{Y}_{n_i}^{(i)} > \boldsymbol{t}_{i,n_i}\}$ corresponds to the event $\{Y_1^{(i)} > t_{i,1}, \ldots, Y_{n_i}^{(i)} > t_{i,n_i}\}$. Let $c_1^{(1)}, \ldots, c_{n_1}^{(1)}, \ldots, c_1^{(d)}, \ldots, c_{n_d}^{(d)}$ be their respective censoring times; therefore, the set of censored data is the following

$$\boldsymbol{D} = \bigcup_{i=1}^{d} \{(T_j^{(i)}, \delta_j^{(i)})\}_{j=1}^{n_i},$$

where $T_j^{(i)} = \min\{Y_j^{(i)}, c_j^{(i)}\}$ and $\delta_j^{(i)} = \mathbb{1}_{(0, c_j^{(i)}]}\left(Y_j^{(i)}\right)$. The number of exact observations is $n_e = \sum_{i=1}^{d} \sum_{j=1}^{n_i} \delta_j^{(i)}$ and the number of censored observations is $n_c = n_1 + n_2 - n_e$. Taking into account the possible repetition of values among the observations, we consider the order statistics $(T_{(1)}, \ldots, T_{(k)})$ of the distinct observations where k is the number of distinct observed times among all groups.

Let define the set functions

$$m_i^e(A) = \sum_{j=1}^{n_i} \delta_j^{(i)} \mathbb{1}_A(T_j^{(i)}) \quad ; \quad m_i^c(A) = \sum_{j=1}^{n_i} (1 - \delta_j^{(i)}) \mathbb{1}_A(T_j^{(i)})$$

for $i \in \{1, \ldots, d\}$, which denote the number of, respectively, exact and censored marginal observations in A, with respect to group i. We define $N_i^e(x) = m_i^e((x, \infty))$, $N_i^c(x) = m_i^c((x, \infty))$, for $i \in \{1, \ldots, d\}$ and $n_{i,j}^e = m_i^e(\{T_{(j)}\})$, $n_{i,j}^c = m_i^c(\{T_{(j)}\})$, $\bar{n}_{i,j}^e = \sum_{r=j}^{k} n_{i,r}^e$ $\bar{n}_{i,j}^c = \sum_{r=j}^{k} n_{i,r}^c$ for $(i,j) \in \{1, \ldots, d\} \times \{1, \ldots, k\}$; and the corresponding vectors

$$\bar{\boldsymbol{n}}_j^e = (\bar{n}_{1,j}^e, \ldots, \bar{n}_{d,j}^e) \quad ; \quad \bar{\boldsymbol{n}}_j^c = (\bar{n}_{1,j}^c, \ldots, \bar{n}_{d,j}^c)$$

for $j \in \{1, \ldots, k\}$ and $\boldsymbol{N}^e(x) = (N_1^e(x), \ldots, N_d^e(x))$, $\boldsymbol{N}^c(x) = (N_1^c(x), \ldots, N_d^c(x))$.

The next theorem determines the calculation of the posterior distribution for a vector of CRM's given some censored data and it applies to general vectors of CRM's. In particular, the assumption that the respective Lèvy intensity is homogeneous has been dropped.

Theorem 1. *Let $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_d)$ be a d-variate CRM such that its correspond-ing Lèvy intensity $\nu(\boldsymbol{s}, \mathrm{d}t)\mathrm{d}\boldsymbol{s}$ is differentiable with respect to t_0 on $\mathbb{R}^+ \setminus \{0\}$ in the sense that for $\eta_t = \nu(\boldsymbol{s}, (0, t])$ the partial derivative $\eta_{t_0}'(\boldsymbol{s}) = \partial \eta_t(\boldsymbol{s})/\partial t\big|_{t=t_0}$ exists. Moreover we assume that the entries of $\boldsymbol{\mu}$ are not independent. Then*

the posterior distribution of $\boldsymbol{\mu}$ given data $\boldsymbol{D}$ is the distribution of the random measure

$$(\mu_1^\star, \ldots, \mu_d^\star) + \sum_{\{j\,:\,T_{(j)} \text{ is an exact observation}\}} (J_{1,j}\delta_{T_{(j)}}, \ldots, J_{d,j}\delta_{T_{(j)}})$$

where

(1) $\boldsymbol{\mu}^\star = (\mu_1^\star, \ldots, \mu_d^\star)$ is a d-variate CRM with Lévy intensity $\nu^\star$ such that

$$\nu^\star(\mathrm{d}\boldsymbol{s}, \mathrm{d}x)\big|_{\mathrm{d}x \in (T_{(j-1)}, T_{(j)})} = \mathrm{e}^{-\langle \bar{\boldsymbol{n}}_j^c + \bar{\boldsymbol{n}}_j^e, \boldsymbol{s}\rangle} \nu(\mathrm{d}\boldsymbol{s}, \mathrm{d}x)$$

   for $j \in \{1, \ldots, k+1\}$.
(2) The vectors of jumps $\{(J_{1,j}, \ldots, J_{d,j})\}_{j \in J}$, where

$$J = \{j\,:\,T_{(j)} \text{ is an exact observation}\},$$

   are mutually independent and the vector of jumps corresponding to the exact observation $T_{(j)}$ has density

$$f_j(\boldsymbol{s}) \propto \prod_{i=1}^d \left\{ \mathrm{e}^{-(\bar{n}_{i,j}^c + \bar{n}_{i,j+1}^e)s_i}(1 - \mathrm{e}^{-s_i})^{n_{i,j}^e} \right\} \eta'_{T_{(j)}}(\boldsymbol{s}).$$

(3) The random measure $\boldsymbol{\mu}^\star$ is independent of $\{(J_{1,j}, \ldots, J_{d,j})\}_{j \in J}$, with $J = \{j\,:\,T_{(j)} \text{ is an exact observation}\}$.

We refer to the Appendix A.4 for the proof. The previous result showcases that the posterior distribution arising from (12) can be modeled in the same framework via a vector of CRM's by updating the prior vector of CRM's $\boldsymbol{\mu}$ to $\boldsymbol{\mu}^\star$ as above.

This result is enough to provide a scheme for posterior inference. In particular, in the setting of (12) and Theorem 1, we want to estimate the corresponding survival function $\mathbb{P}\big[Y^{(1)} > t_1, \ldots, Y^{(d)} > t_d\,|\,(\mu_1, \ldots, \mu_d)\big]$ when multiple samples information is available.

A natural approach in Bayesian nonparametrics is to marginalize over the infinite dimensional random element which characterizes the probability model. In our case, given censored data $\boldsymbol{D}$, we calculate the mean of the survival function given the data by marginalizing over the vector of CRM's $\boldsymbol{\mu}$. As a result of Theorem 1, we can calculate such quantity. The next results allow us to implement the necessary inferential scheme for performing the estimation of the survival function as a posterior mean. We denote $\boldsymbol{e}_i$ for the canonical basis of $\mathbb{R}^d$, and $S_L(t) = S(t \sum_{l \in L} \boldsymbol{e}_l)$ for $t > 0$, $\emptyset \neq L \subset \{1, \ldots, d\}$. In view of the independent increments of the CRM's, calculation of the posterior mean of S_L is all that is needed for the evaluation of the posterior mean of S. The next corollary shows how to evaluate the posterior mean of S_L.

Corollary 1. *Let $\boldsymbol{\mu}$ be a vector of CRM's with corresponding Lèvy intensity such that $\eta_t(\boldsymbol{s}) = \gamma(t)\nu(\boldsymbol{s})$ with γ a differentiable function satisfying $\gamma'(t) \neq 0$*

for $t > 0$. Moreover we assume that the entries of $\boldsymbol{\mu}$ are not independent. Let $\emptyset \neq L \subset \{1, \ldots, d\}$ and set

$$J_t = \{j \,:\, T_{(j)} \leq t\}$$

where $T_{(k+1)} = \infty$. Then,

$$\hat{S}_L(t) = \mathbb{E}[\mathbb{E}[S_L(t)|\boldsymbol{\mu}]\,|\boldsymbol{D}] = \mathrm{e}^{-\sum_{j=1}^{k+1}\left[\gamma(t \wedge T_{(j)}) - \gamma(T_{j-1})\right]\mathbf{1}_{[T_{(j-1)},\infty)}(t)\psi_j^\star\left(\sum_{l \in L} \boldsymbol{e}_l\right)}$$

$$\times \prod_{j \in J_t} \gamma'(T_{(j)}) \left[\frac{\int_{(\mathbb{R}^+)^d} \prod_{i=1}^d \left\{ \mathrm{e}^{-(\mathbf{1}_{i \in L} + \bar{n}_{i,j}^c + \bar{n}_{i,j+1}^e)s_i}(1 - \mathrm{e}^{-s_i})^{n_{i,j}^e} \right\} \nu(\boldsymbol{s})\mathrm{d}\boldsymbol{s}}{\int_{(\mathbb{R}^+)^d} \prod_{i=1}^d \left\{ \mathrm{e}^{-[\bar{n}_{i,j}^c + \bar{n}_{i,j+1}^e]s_i}(1 - \mathrm{e}^{-s_i})^{n_{i,j}^e} \right\} \nu(\boldsymbol{s})\mathrm{d}\boldsymbol{s}} \right]$$

where $T_{(0)} = 0$ and for $\boldsymbol{\lambda} \in (\mathbb{R}^+)^d$

$$\psi_j^\star(\boldsymbol{\lambda}) = \int_{(\mathbb{R}^+)^d} \left(1 - \mathrm{e}^{-\langle \boldsymbol{\lambda}, \boldsymbol{s}\rangle}\right) \mathrm{e}^{-\langle \bar{\boldsymbol{n}}_j^c + \bar{\boldsymbol{n}}_j^e, \boldsymbol{s}\rangle} \nu(\boldsymbol{s})\mathrm{d}\boldsymbol{s}$$

$$= \psi(\boldsymbol{\lambda} + \bar{\boldsymbol{n}}_j^c + \bar{\boldsymbol{n}}_j^e) - \psi(\bar{\boldsymbol{n}}_j^c + \bar{\boldsymbol{n}}_j^e).$$

We see that we can estimate $S(\boldsymbol{t})$ for arbitrary $\boldsymbol{t} \in (\mathbb{R}^+)^d$ in terms of the estimates defined in the previous corollary. Indeed, let $\boldsymbol{t} = (t_1, \ldots, t_d)$ and π be a permutation of $\{1, \ldots, d\}$ such that $t_{\pi(1)} \leq t_{\pi(2)} \leq \cdots \leq t_{\pi(d)}$. We define, for $i \in \{1, \ldots, d-1\}$, the following sets

$$L_i = \{j \,:\, \pi^{(-1)}(j) \geq i\}.$$

From the independence of increments of CRM's, it follows that the posterior mean of the survival function given censored data $\boldsymbol{D}$ is

$$\hat{S}(\boldsymbol{t}) = \mathbb{E}[\mathbb{E}[S(\boldsymbol{t})|\boldsymbol{\mu}]\,|\boldsymbol{D}] = \hat{S}_{L_1}(t_{\pi(1)}) \prod_{i=1}^{d-1} \frac{\hat{S}_{L_i}(t_{\pi(i+1)})}{\hat{S}_{L_i}(t_{\pi(i)})} \quad \boldsymbol{t} \in (\mathbb{R}^+)^d. \qquad (17)$$

Usually, we deal with Lévy intensities which exhibit some dependences in a vector of hyper-parameters $\boldsymbol{c}$. On the proof of Theorem 1, it is outlined how, given censored data $\boldsymbol{D}$ as before, we could derive the likelihood of the hyper-parameters in the Lévy intensity. This likelihood is necessary for implementing the inferential procedure and it is displayed in the next corollary.

Corollary 2. *Let $\boldsymbol{\mu}$ be a vector of CRM's with corresponding Lèvy intensity such that $\eta_t(\boldsymbol{s}) = \gamma(t)\rho_{d,\boldsymbol{c}}(\boldsymbol{s})$ with γ a differentiable function satisfying $\gamma'(t) \neq 0$ for $t > 0$, and $\boldsymbol{c}$ a vector of hyper-parameters. Given censored data $\boldsymbol{D}$ we get the likelihood on $\boldsymbol{c}$.*

$$l(\boldsymbol{c}; \mathcal{D}) = \mathrm{e}^{-\sum_{j=1}^k \left[\gamma(T_{(j)}) - \gamma(T_{(j-1)})\right]\psi_{\boldsymbol{d},\boldsymbol{c}}(\bar{\boldsymbol{n}}_j^c + \bar{\boldsymbol{n}}_j^e)}$$

$$\times \prod_{j \in J} \gamma'(T_{(j)}) \int_{(\mathbb{R}^+)^d} \prod_{i=1}^d \left\{ \mathrm{e}^{-(\bar{n}_{i,j}^c + \bar{n}_{i,j}^e)s_i}(1 - \mathrm{e}^{-s_i})^{n_{i,j}^e} \rho_{d,\boldsymbol{c}}(\boldsymbol{s})\mathrm{d}\boldsymbol{s} \right\},$$

where $\psi_{d,\boldsymbol{c}}$ is the Laplace exponent associated to $\rho_{d,\boldsymbol{c}}$.

The next lemma provides a useful identity for the computation of the integrals in Corollary 1 and Corollary 2.

Lemma 1. *For $\boldsymbol{q} = (q_1, \ldots, q_d) \in (\mathbb{R}^+)^d$ and $\boldsymbol{n} = (n_1, \ldots, n_d) \in \mathbb{N}^d$*

$$
\int_{(\mathbb{R}^+)^d} \mathrm{e}^{-\langle \boldsymbol{q}, \boldsymbol{x} \rangle} \prod_{i=1}^{d} \left(1 - \mathrm{e}^{-s_i}\right)^{n_i} \nu(\boldsymbol{s}) \mathrm{d}\boldsymbol{s} = \sum_{i=1}^{d} \sum_{k=1}^{n_i} \binom{n_i}{k} (-1)^{k+1} [\psi(k\boldsymbol{e}_i + \boldsymbol{q}) - \psi(\boldsymbol{q})]
$$

$$
+ \sum_{\substack{i_1 \neq i_2 \\ n_{i_1}, n_{i_2} \notin \{0\}}} \sum_{k_1=1}^{n_1} \sum_{k_2=1}^{n_2} \binom{n_1}{k_1} \binom{n_2}{k_2} (-1)^{k_1+k_2+1} [\psi(k_1 \boldsymbol{e}_{i_1} + k_2 \boldsymbol{e}_{i_2} + \boldsymbol{q}) - \psi(\boldsymbol{q})]
$$

$$
+ \ldots
$$

$$
+ \mathbf{1}_{\{n_1 \neq 0, \ldots, n_d \neq 0\}} \sum_{k_1=1}^{n_1} \cdots \sum_{k_d=1}^{n_d} (-1)^{k_1 + \cdots + k_d + 1} [\psi(k_1 \boldsymbol{e}_1 + \cdots + k_d \boldsymbol{e}_d + \boldsymbol{q}) - \psi(\boldsymbol{q})].
$$

We omit the proof as it is just an application of the binomial theorem in the same line as the proof of Lemma 3 in the appendix.

Remark. *The previous results highlights that the implementation of the inferential procedure depends on whether we can perform evaluations of the Laplace exponent or not.*

4. Application

In this section we perform the fitting of a multivariate survival function given censored to the right data in the framework of (12). As mentioned in the previous remark, the evaluation of the Laplace exponent of $\boldsymbol{\mu}$ in (12) is necessary to evaluate the posterior mean in Corollary 1 and the likelihood in Corollary 2; with this in mind, we choose the random measure $\boldsymbol{\mu}$ given by the Lévy intensity showcased in (9), so that the corresponding Laplace exponent is readily given by (10). For illustration purposes, we use 4-dimensional data arising from a distributional copula with fixed marginal distributions, see Nelsen (2013) for an overview of distributional copulas. More precisely, we generate simulated data $\boldsymbol{Y} = (Y_1, \ldots, Y_4)$ with probability distribution $F_{\theta, \lambda}$ given by a distributional Clayton copula with parameter θ and exponential marginals with parameter λ. Then, we perform right-censoring by considering censoring time variables $\boldsymbol{c}$ consisting of independent exponential random variables with parameter λ_c, and define

$$
\begin{aligned}
\boldsymbol{\delta} &= (\mathbf{1}_{Y_1 < c_1}, \ldots, \mathbf{1}_{Y_4 < c_4}), \\
\boldsymbol{T} &= (\min\{Y_1, c_1\}, \ldots, \min\{Y_4, c_4\}).
\end{aligned} \tag{18}
$$

For fitting the data, we use the 4-dimensional Lévy intensity given by (9) and assign priors for the hyper-parameters in (9), σ and A. We choose a log-normal prior for the parameter A and a Beta prior for the parameter σ. We use the

Metropolis within Gibbs algorithm to draw samples from the posterior distributions of A and σ by making use of the likelihood showed in Corollary 2. We present a Monte Carlo approximation of the estimator (17), where we have averaged over the posterior draws of A and σ. A more in depth description of the simulation algorithm is given in Appendix A.5. In Figures 3 and 4 we show the fit for 150 possibly right censored observations as in (18). The simulated synthetic observations are such that

$$
\begin{aligned}
\boldsymbol{Y}_j &\sim F_{\theta=0.3,\lambda=1.}, & j &= 1,\ldots,150 \\
c_{i,j} &\sim \mathrm{Exp}(\lambda_c = 3.7), & i &= 1,...,4; \quad j = 1,\ldots,150 \\
T_{i,j} &= \min\{Y_{i,j}, c_{i,j}\}, & i &= 1,...,4; \quad j = 1,\ldots,150.
\end{aligned}
$$

We chose $\lambda_c = 3.7$ so we have at least 75% of exact observations for $\boldsymbol{T}$ in each dimension. The construction of $F_{\theta,\lambda}$ through a distributional Clayton allows us to calculate explicitly the associated survival function as showcased in Appendix A.6. We use the true survival function for comparison with the fitted survival functions. The estimated survival function are given by the posterior mean

$$
\hat{S}(t_1, t_2, t_3, t_4) = \mathbb{E}[\mathbb{E}[S(t_1, t_2, t_3, t_4)|\boldsymbol{\mu}]\,|\boldsymbol{D}]\,,
$$

as in (17). The prior distributions of the hyperparameters are

$$
\begin{aligned}
\sigma &\sim \ \mathrm{Beta}(\mu = 0.4, \sigma^2 = 0.1) \\
A &\sim \ \mathrm{Log\text{-}Norm}(\mu = \log(0.88), \sigma^2 = 3.5).
\end{aligned}
$$

We ran 1000 iterations for the associated Metropolis within Gibbs sampler. Figure 3 and Figure 4 show that the estimated survival functions approximate well the true functions. For comparison purposes, we presented a Kaplan-Meier estimator for the true survival function, see for example Aalen et al. (2008). As there is no multivariate Kaplan-Meier, we use the next estimator for a multivariate survival function:

$$
\begin{aligned}
\hat{S}_{\mathrm{KM}}(t_1,\ldots,t_d) &= \\
S_{\mathrm{KM}}(t_1|T_2 > t_2,\ldots,T_d > t_d)&S_{\mathrm{KM}}(t_2|T_3 > t_3,\cdots,T_d > t_d)\ldots S_{\mathrm{KM}}(t_d),
\end{aligned}
$$

where each S_{KM} estimator is treated as a univariate Kaplan-Meier estimator restricted to the corresponding set of observations. In Figure 3 and Figure 4, we could appreciate that in the last subplots of each column the Kaplan-Meier can fit poorly as there are less observations on the conditioned Kaplan-Meier functions, as presented in the formula above.

Appendix

A.1. Proof of Proposition 1

Given $d \in \{2, 3, \ldots\}$, we use the notation $\nu_{-i}(\boldsymbol{s}) = \prod_{j=i+1}^{d} \nu_j(s_j)$ and $\boldsymbol{U}_{k:d}(\boldsymbol{s}) = (U_k(s_1), \ldots, U_d(s_{d-k+1}))$ for $\boldsymbol{s} \in (\mathbb{R}^+)^d$. Furthermore we define integrals

$$
a_{0,m}(\boldsymbol{\lambda}) = \int_{(\mathbb{R}^+)^m} (1 - \mathrm{e}^{-<\boldsymbol{\lambda},\boldsymbol{s}>}) \frac{\partial^d}{\partial u_d \cdots \partial u_1} C_{\theta,m}(\boldsymbol{u})\big|_{\boldsymbol{u}=\boldsymbol{U}_{d-m+1:d}(\boldsymbol{s})} \nu_{-0}(\boldsymbol{s})\mathrm{d}\boldsymbol{s}
$$
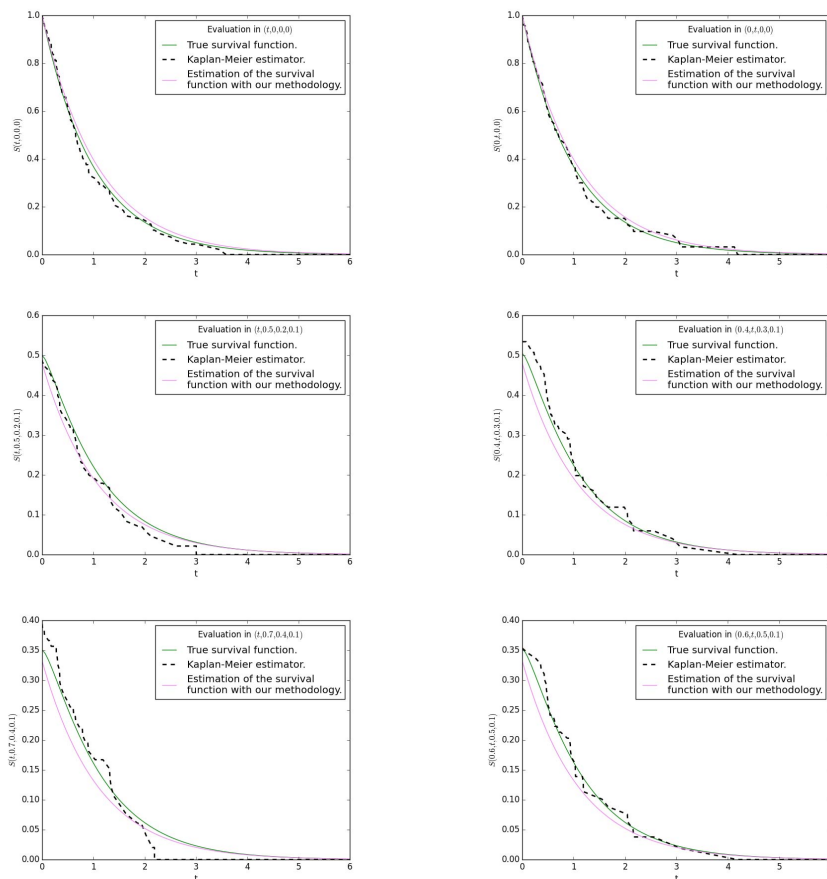
(a) Fits with first dimension not fixed.   (b) Fits with second dimension not fixed.

FIG 3. *Plot of our methodology fits (violet lines), compared with Kaplan-Meier fits (dashed lines) and the true survival function associated to the distributions $F_{\theta=0.3, \lambda=1}$. (green lines). The first column shows fits of the survival function with fixed values in all dimensions except the first one; the second column has fixed values in all dimensions except the second one.*

and

$$
\begin{aligned}
& a_{k,m}(\boldsymbol{\lambda}) \\
& = (-1)^{k+1} \int_{(\mathbb{R}^+)^m} \lambda_1 \cdots \lambda_k \mathrm{e}^{-<\boldsymbol{\lambda},\boldsymbol{s}>} \frac{\partial^{d-k}}{\partial u_d \cdots \partial u_{k+1}} C_{\theta,m}(\boldsymbol{U}_{d-m+1:d}(\boldsymbol{s}))\nu_{-k}(\boldsymbol{s})\mathrm{d}\boldsymbol{s}
\end{aligned}
$$

where $k \in \{1, \dots, d\}$, $m \in \{0, 1, \dots, d\}$ and $\boldsymbol{\lambda} \in (\mathbb{R}^+)^d$ such that $a_{0,d}(\boldsymbol{\lambda}) < \infty$; we also define $\prod_{j=k}^{l} a_j = 1$ when $k > l$, and denote $\boldsymbol{x}_{-i}$ for the vector $\boldsymbol{x}$ without its i-th entry.
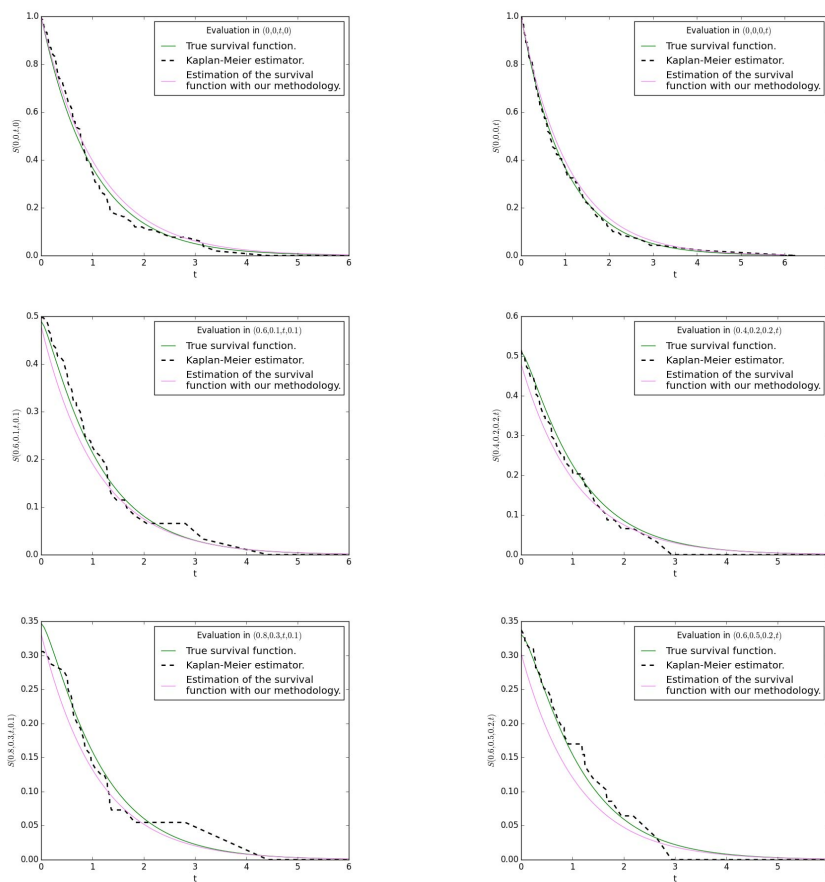
(a) Fits with third dimension not fixed.　　(b) Fits with fourth dimension not fixed.

FIG 4. *Plot of our methodology fits (violet lines), compared with Kaplan-Meier fits (dashed lines) and the true survival function associated to the distributions $F_{\theta=0.3,\lambda=1}$. (green lines). The first column shows fits of the survival function with fixed values in all dimensions except the third one; the second column has fixed values in all dimensions except the fourth one.*

An integration by parts shows that

$$
\begin{aligned}
a_{0,d} = & -\int_{(\mathbb{R}^+)^{d-1}} (1 - \mathrm{e}^{-<\boldsymbol{\lambda},\boldsymbol{s}>}) \frac{\partial^{d-1}}{\partial u_d \cdots \partial u_2} C_{\theta,d}(\boldsymbol{u})\big|_{\boldsymbol{u}=\boldsymbol{U}_d(\boldsymbol{s})} \nu_{-1}(\boldsymbol{s})\bigg|_{s_1=0}^{s_1=\infty} \mathrm{d}\boldsymbol{s}_{-1} \\
& + \int_{(\mathbb{R}^+)^d} \lambda_1 \mathrm{e}^{-<\boldsymbol{\lambda},\boldsymbol{s}>} \frac{\partial^{d-1}}{\partial u_d \cdots \partial u_2} C_{\theta,d}(\boldsymbol{u})\big|_{\boldsymbol{u}=\boldsymbol{U}_d(\boldsymbol{s})} \nu_{-1}(\boldsymbol{s})\mathrm{d}\boldsymbol{s} \\
= & \; a_{0,d-1}(\boldsymbol{\lambda}_{-1}) + a_{1,d}(\boldsymbol{\lambda})
\end{aligned}
$$

and in general for $r \in \{1, \ldots, d\}$ we get the recursion formula

$$
a_{r,d}(\boldsymbol{\lambda}) = a_{r,d-1}(\boldsymbol{\lambda}_{-(r+1)}) + a_{r+1,d}(\boldsymbol{\lambda}) \tag{A.19}
$$

We prove the next technical lemma

Lemma 2. *If $a_{0,d}(\boldsymbol{\lambda}) < \infty$ then the next $d+1$ identities hold*

$$a_{0,d}(\boldsymbol{\lambda}) = \sum_{i=1}^{d} \psi_i(\lambda_i) - \sum_{\substack{\boldsymbol{i}=(i_1,i_2)\in\{1,\ldots,d\}^2 \\ i_1<i_2}} \kappa(\theta;\boldsymbol{\lambda},\boldsymbol{i}) + \cdots$$

$$\cdots + (-1)^d \sum_{\substack{\boldsymbol{i}=(i_1,\ldots,i_{d-1})\in\{1,\ldots,d\}^{d-1} \\ i_1<\cdots<i_{d-1}}} \kappa\left(\theta;\boldsymbol{\lambda},(i_1,\ldots,i_{d-1})\right)$$

$$+ (-1)^{d+1}\kappa(\theta;\boldsymbol{\lambda},(1,\ldots,d))$$

$$a_{1,d}(\boldsymbol{\lambda}) = \psi_1(\lambda_1) - \sum_{i=2}^{d} \kappa\left(\theta;\boldsymbol{\lambda},(1,i)\right) + \sum_{\substack{i_1,i_2\in\{2,\ldots,d\} \\ i_1<i_2}} \kappa\left(\theta;\boldsymbol{\lambda},(1,i_1,i_2)\right) + \cdots$$

$$\cdots + (-1)^d \sum_{\substack{i_1,\ldots,i_{d-2}\in\{2,\ldots,d\} \\ i_1<\cdots<i_{d-2}}} \kappa((\theta;\boldsymbol{\lambda},(1,i_1,\ldots,i_{d-2}))$$

$$+ (-1)^{d+1}\kappa\left(\theta;\boldsymbol{\lambda},(1,\ldots,d)\right)$$

$$\vdots$$

$$a_{d-1,d}(\boldsymbol{\lambda}) = (-1)^d \kappa\left(\theta;\boldsymbol{\lambda},(1,\ldots,d-1)\right) + (-1)^{d+1}\kappa\left(\theta;\boldsymbol{\lambda},(1,\ldots,d)\right)$$

$$a_{d,d}(\boldsymbol{\lambda}) = (-1)^{d+1}\kappa\left(\theta;\boldsymbol{\lambda},(1,\ldots,d)\right) \tag{A.20}$$

Proof. We proceed by mathematical induction over the dimension d. We observe that from the definition of κ we always have

$$a_{d,d}(\boldsymbol{\lambda}) = (-1)^{d+1}\kappa(\theta;\boldsymbol{\lambda},(1,\ldots,d))$$

For the case $d=2$ we have from Proposition 1 in Epifani and Lijoi (2010) that

$$a_{0,2}(\lambda_1,\lambda_2) = \psi_1(\lambda_1) + \psi_2(\lambda_2) - \kappa(\theta;(\lambda_1,\lambda_2),(1,2))$$

And integrating by parts we obtain

$$a_{1,2}(\lambda_1,\lambda_2) =$$
$$\int_{\mathbb{R}^+} \lambda_1 e^{-\lambda_1 s_1} U_1(x_1)\mathrm{d}s_1 - \lambda_1 \quad \lambda_2 \int_{(\mathbb{R}^+)^2} e^{-\lambda_1 x_2 - \lambda_2 s_2} C_\theta(U_1(s_1), U_2(s_2))\mathrm{d}s_1\mathrm{d}s_2$$
$$= \psi_1(\lambda_1) - \kappa\left(\theta;\boldsymbol{\lambda},(1,2)\right)$$

Therefore, we get the validity of the equations in (A.20) for the case $d=2$. Now, suppose that (A.20) is true for $d=m-1$, we must show the validity for $d=m$. From the recursion formula (A.19) we get for $r \in \{0,1,\cdots,d\}$

$$a_{r,m}(\boldsymbol{\lambda}) = a_{r,m-1}(\boldsymbol{\lambda}_{-(r+1)}) + a_{r+1,m-1}(\boldsymbol{\lambda}_{-(r+2)})$$
$$+ \cdots + a_{m-1,m-1}(\boldsymbol{\lambda}_{-m}) + a_{m,m}(\boldsymbol{\lambda})$$

The validity of (A.20) for $d = m$ follows from the validity for $d = m - 1$ and a combinatorial argument. $\qquad\square$

Proposition 1 follows by considering the first equation in the Lemma statement and the definition of $a_{0,d}$.

A.2. Proof of Proposition 2

Proof. Using the independent increments property of CRM's we get that

$$\mathbb{P}\Big[Y^{(1)} > t_1, \ldots, Y^{(d)} > t_d\Big] = \mathbb{E}\Big[e^{-\mu_1(0,t_1] - \cdots - \mu_d(0,t_d]}\Big]$$

$$= \mathbb{E}\Big[e^{-\mu_{i_1}(0,t_{i_1}] - \cdots - \mu_{i_d}(0,t_{i_1}]}\Big]$$

$$\times \mathbb{E}\Big[e^{-\mu_{i_2}(t_{i_1},t_{i_2}] - \cdots - \mu_{i_d}(t_{i_1},t_{i_2}]}\Big] \cdots \mathbb{E}\Big[e^{-\mu_{i_d}(t_{i_{d-1}},t_{i_d}]}\Big]$$

$$= e^{-\gamma(t_{i_1})\psi(\mathbf{1})} e^{-[\gamma(t_{i_2}) - \gamma(t_{i_1})]\psi_{i_2,\ldots,i_d}(\mathbf{1})} \cdots e^{-[\gamma(t_{i_d}) - \gamma(t_{i_{d-1}})]\psi_{i_d}(\mathbf{1})} \qquad\square$$

A.3. Proof of Proposition 3

For notation purposes, in this proof we use the shorthand $\mu(t) = \mu\left((0,t]\right)$ for a measure μ and positive real number t.

Proof. For the only if part we define $V_{i,j} = 1 - e^{-[\mu_i(t_{i,j}) - \mu_i(t_{i,j-1})]}$ for $i \in \{1, \ldots, d\}$ and $j \in \{1, \ldots, h\}$ so by supposing $(F_1(t_1), \ldots, F_d(t_d)) \overset{d}{=} (1 - e^{-\mu_1(t_1)}, \ldots, 1 - e^{-\mu_d(t_d)})$ we have

$$F(t_{1,1}, \ldots, t_{d,1}) \overset{d}{=} [1 - e^{-\mu_1(t_{1,1})}] \cdots [1 - e^{-\mu_d(t_{d,1})}]$$

$$= [1 - e^{-[\mu_1(t_{1,1}) - \mu_1(t_{1,0})]}] \cdots [1 - e^{-[\mu_d(t_{d,1}) - \mu_d(t_{d,0})]}]$$

$$= V_{1,1} \cdots V_{d,1}$$

We observe that for $i \in \{2, \ldots, h\}$ and $r \in \{1, \ldots, d\}$

$$1 - \prod_{j=1}^{i} \bar{V}_{r,j} = 1 - \prod_{j=1}^{i}(1 - V_{r,j}) = 1 - \prod_{j=1}^{i} e^{-[\mu_r(t_{r,j}) - \mu_r(t_{r,j-1})]} = 1 - e^{-\mu_r(t_{r,i})}$$

So for $i \in \{2, \ldots d\}$

$$F(t_{1,i}, \ldots, t_{d,i}) \overset{d}{=} [1 - e^{-\mu_1(t_{1,i})}] \cdots [1 - e^{-\mu_d(t_{d,i})}]$$

$$= [1 - \prod_{j=1}^{i} \bar{V}_{1,j}] \cdots [1 - \prod_{j=1}^{i} \bar{V}_{d,j}].$$

Concluding the only if part.
For the if part we define $\mu_i(t) = -\log(1 - F_i(t))$ for $i \in \{1, \ldots, d\}$ and suppose for $h \in \{1, 2, \ldots\}$, $\boldsymbol{t}_1 = (t_{1,1}, \ldots, t_{d,1}), \ldots, \boldsymbol{t}_h = (t_{1,h}, \ldots, t_{d,h})$ with $t_{0,i} = 0 <$

$t_{1,i} < \cdots < t_{d,i}$ and $t_{j,0} = 0 < t_{j,1} < \cdots < t_{j,h}$ the existence of independent random vectors $(V_{1,1}, \ldots V_{d,1}), \ldots, (V_{1,h}, \ldots V_{d,h})$ such that we have (16). Marginalizing in (16), we can apply Theorem 3.1 of Doksum (1974) to each F_i so we obtain that $F_i \sim \text{NTR}(\mu_i)$ for some CRM μ_i that is stochastically continuous, almost surely non-decreasing and has the appropriate limit behaviour. We observe that

$$(\mu_1(t_j) - \mu_1(t_{j-1}), \ldots, \mu_d(t_j) - \mu_d(t_{j-1})) \overset{d}{=} \left(\log(\frac{1}{1 - V_{1,j}}), \ldots, \log(\frac{1}{1 - V_{d,j}}) \right)$$

Hence $(\mu_1, \ldots, \mu_d)$ defines a vector of CRM's. $\qquad \square$

A.4. Proof of Theorem 1

This proof is not only restricted to the homogeneous Lévy intensity case; in this general setting, we recall that the Laplace exponent has the form (6). In order to prove the theorem we use the next technical lemma.

Lemma 3. *Let $(\mu_1, \ldots, \mu_d)$ be a d-variate CRM such that $\mu_1, \ldots, \mu_d$ are not independent and let the Lévy intensity $\nu(\boldsymbol{s}, \mathrm{d}t)\mathrm{d}\boldsymbol{s}$ of $(\mu_1, \ldots, \mu_d)$ be such that $\eta_t = \nu(\boldsymbol{x}, (0, t])$ is differentiable with respect to $t \in \mathbb{R}^+$ at some $t_0 \neq 0$ and denote $\eta'_{t_0}(\boldsymbol{s}) = \partial \eta_t(\boldsymbol{s})/\partial t \big|_{t=t_0}$. If $\boldsymbol{q} = (q_1, \ldots, q_d) \in \mathbb{N}^d$ are such that $\max\{q_1, \ldots, q_d\} \geq 1$ and $\boldsymbol{r} = (r_1, \ldots, r_d) \in (\mathbb{R}^+)^d$ are such that $\min\{r_1, \ldots, r_d\} \geq 1$, then*

$$\mathbb{E}\left[e^{-r_1\mu_1(A_\epsilon) - \cdots - r_d\mu_d(A_\epsilon)} \left(1 - e^{-\mu_1(A_\epsilon)}\right)^{q_1} \cdots \left(1 - e^{-\mu_d(A_\epsilon)}\right)^{q_d} \right]$$

$$= \epsilon \int_{(\mathbb{R}^+)^d} e^{-\langle \boldsymbol{r}, \boldsymbol{s} \rangle} (1 - e^{-s_1})^{q_1} \cdots (1 - e^{-s_d})^{q_d} \eta'_{t_0}(\boldsymbol{s}) \mathrm{d}\boldsymbol{s} + o(\epsilon)$$

as $0 < \epsilon \to 0$, with $A_\epsilon = (t_0 - \epsilon, t_0]$ for some $t_0 \in \mathbb{R}^+ \setminus \{0\}$.

Proof. We denote $\triangle_{s_1}^{s_2} f_t(\boldsymbol{r}) = f_{s_2}(\boldsymbol{r}) - f_{s_1}(\boldsymbol{r})$ for a function f where $s_1, s_2 \in \mathbb{R}^+$ and $\boldsymbol{r} \in \mathbb{R}^d$. We use the binomial theorem and apply expectation to write the left hand side in the equation above as

$$\sum_{j_1=0}^{q_1} \cdots \sum_{j_d=0}^{q_d} \binom{q_1}{j_1} \cdots \binom{q_d}{j_d} (-1)^{\langle \boldsymbol{1}, \boldsymbol{j} \rangle} e^{-[\psi_{t_0}(r_1+j_1, \ldots, r_d+j_d) - \psi_{t_0-\epsilon}(r_1+j_1, \ldots, r_d+j_d)]}$$

$$= e^{-\triangle_{t_0-\epsilon}^{t_0} \psi_t(\boldsymbol{r})} + e^{-\triangle_{t_0-\epsilon}^{t_0} \psi_t(\boldsymbol{r})} \left\{ \sum_{i=1}^{d} \sum_{j=1}^{q_i} \binom{q_i}{j} (-1)^j e^{-\triangle_{t_0-\epsilon}^{t_0}[\psi_t(\boldsymbol{r}+j\boldsymbol{e}_i) - \psi_t(\boldsymbol{r})]} \right.$$

$$+ \sum_{\substack{i_1, i_2 \in \{1, \ldots, d\} \\ i_1 < i_2}} \sum_{j_1=1}^{q_{i_1}} \sum_{j_2=1}^{q_{i_2}} \binom{q_{i_1}}{j_1} \binom{q_{i_2}}{j_2} (-1)^{j_1+j_2} e^{-\triangle_{t_0-\epsilon}^{t_0}[\psi_t(\boldsymbol{r}+j_1\boldsymbol{e}_{i_1}+j_2\boldsymbol{e}_{i_2}) - \psi_t(\boldsymbol{r})]}$$

$$\left. + \cdots + \sum_{j_1=1}^{q_1} \cdots \sum_{j_d=1}^{q_d} \binom{q_1}{j_1} \cdots \binom{q_d}{j_d} (-1)^{\langle \boldsymbol{1}, \boldsymbol{j} \rangle} e^{-\triangle_{t_0-\epsilon}^{t_0}[\psi_t(\boldsymbol{r}+\boldsymbol{j}) - \psi_t(\boldsymbol{r})]} \right\} \quad (\text{D.21})$$

We note that for $j_i \in \{0, \ldots, x_i\}$, $i \in \{1, \ldots, d\}$, $\boldsymbol{j} = (j_1, \ldots, j_d)$, a Taylor expansion yields

$$\mathrm{e}^{-\triangle_{t_0-\epsilon}^{t_0}[\psi_t(\boldsymbol{r}+\boldsymbol{j})-\psi_t(\boldsymbol{r})]} = \mathrm{e}^{-\int_{(\mathbb{R}+)^d} \mathrm{e}^{-\langle \boldsymbol{r},\boldsymbol{x}\rangle}(1-\mathrm{e}^{-\langle \boldsymbol{j},\boldsymbol{s}\rangle})\triangle_{t_0-\epsilon}^{t_0}\eta_t(\boldsymbol{s})\mathrm{d}\boldsymbol{s}}$$

$$= 1 - \epsilon \int_{(\mathbb{R}+)^d} \mathrm{e}^{-\langle \boldsymbol{r},\boldsymbol{s}\rangle}(1 - \mathrm{e}^{-\langle \boldsymbol{j},\boldsymbol{s}\rangle})\eta'_{t_0}(\boldsymbol{s})\mathrm{d}\boldsymbol{s} + o(\epsilon) \tag{D.22}$$

Furthermore by the binomial theorem we get the next d identities

$$(1) \sum_{i=1}^{d}\sum_{j=1}^{q}\binom{q}{j}(-1)^j(1-\mathrm{e}^{-js}) = -\sum_{i=1}^{d}(1-\mathrm{e}^{-s})^q$$

$$(2) \sum_{\substack{i_1,i_2\in\{1,\ldots,d\}\\i_1<i_2}} \sum_{j_1=1}^{q_{i_1}}\sum_{j_2=1}^{q_{i_2}}\binom{q_{i_1}}{j_1}\binom{q_{i_2}}{j_2}(-1)^{j_1+j_2}(1-\mathrm{e}^{-j_1s_{i_1}-j_2s_{i_2}})$$

$$= \sum_{\substack{i_1,i_2\in\{1,\ldots,d\}\\i_1<i_2}}\left\{(1-\mathrm{e}^{-s_{i_1}})^{q_{i_1}} + (1-\mathrm{e}^{-s_{i_2}})^{q_{i_2}} - (1-\mathrm{e}^{-s_{i_1}})^{q_{i_1}}(1-\mathrm{e}^{-s_{i_2}})^{q_{i_2}}\right\}$$

$$\vdots$$

$$(d-1) \sum_{\substack{i_1,\ldots,i_{d-1}\in\{1,\ldots,d\}\\i_1<\cdots<i_{d-1}}} \sum_{j_1=1}^{q_{i_1}}\cdots\sum_{j_{d-1}=1}^{q_{i_{d-1}}}\binom{q_{i_1}}{j_1}\cdots\binom{q_{i_{d-1}}}{j_{d-1}}(-1)^{j_1+\cdots+j_{d-1}}$$

$$\times(1-\mathrm{e}^{-j_1s_{i_1}-\cdots-j_{d-1}s_{i_{d-1}}})$$

$$= \sum_{\substack{i_1,\ldots,i_{d-1}\in\{1,\ldots,d\}\\i_1<\cdots<i_{d-1}}}\left\{(-1)^{d-1}\sum_{j=1}^{d-1}(1-\mathrm{e}^{-s_{ij}})^{q_{ij}}\right.$$

$$+ (-1)^{d-2}\sum_{\substack{j_1,j_2\in\{i_1,\ldots,i_{d-1}\}\\j_1<j_2}}(1-\mathrm{e}^{-s_{j_1}})^{q_{j_1}}(1-\mathrm{e}^{-s_{j_2}})^{q_{j_2}}$$

$$\left.+\cdots-(1-\mathrm{e}^{-s_{i_1}})^{q_{i_1}}\cdots(1-\mathrm{e}^{-s_{i_{d-1}}})^{q_{i_{d-1}}}\right\}$$

$$(d) \sum_{j_1=1}^{q_1}\cdots\sum_{j_d=1}^{q_d}\binom{q_1}{j_1}\cdots\binom{q_d}{j_d}(-1)^{\langle \boldsymbol{1},\boldsymbol{j}\rangle}(1-\mathrm{e}^{-\langle \boldsymbol{j},\boldsymbol{s}\rangle}) = (-1)^d\sum_{j=1}^{d}(1-\mathrm{e}^{-s_j})^{q_j}$$

$$+ (-1)^{d-1}\sum_{\substack{j_1,j_2\in\{1,\ldots,d\}\\j_1<j_2}}(1-\mathrm{e}^{-s_{j_1}})^{q_{j_1}}(1-\mathrm{e}^{-s_{j_2}})^{q_{j_2}}$$

$$+\cdots-(1-\mathrm{e}^{-s_{i_1}})^{q_{i_1}}\cdots(1-\mathrm{e}^{-s_{i_d}})^{q_{i_d}}$$

So we have that (D.21) becomes

$$e^{-\triangle_{t_0-\epsilon}^{t_0}\psi_t(\boldsymbol{r})}\left\{1+\sum_{i=1}^{d}\sum_{j=1}^{q_i}\binom{q_i}{j}(-1)^j\right.$$

$$-\epsilon\int_{(\mathbb{R}^+)^d}e^{-\langle\boldsymbol{r},\boldsymbol{s}\rangle}\sum_{i=1}^{d}\sum_{j=1}^{q_i}\binom{q_i}{j}(-1)^j(1-e^{-j_1s_1})\eta'_{t_0}(\boldsymbol{s})\mathrm{d}\boldsymbol{s}$$

$$+\sum_{\substack{i_1,i_2\in\{1,\ldots,d\}\\i_1<i_2}}\sum_{j_1=1}^{q_{i_1}}\sum_{j_2=1}^{q_{i_2}}\binom{q_{i_1}}{j_1}\binom{q_{i_2}}{j_2}(-1)^{j_1+j_2}$$

$$-\epsilon\int_{(\mathbb{R}^+)^d}e^{-\langle\boldsymbol{r},\boldsymbol{s}\rangle}\sum_{\substack{i_1,i_2\in\{1,\ldots,d\}\\i_1<i_2}}\sum_{j_1=1}^{q_{i_1}}\sum_{j_2=1}^{q_{i_2}}\binom{q_{i_1}}{j_1}\binom{q_{i_2}}{j_2}(-1)^{j_1+j_2}$$

$$\times(1-e^{-j_1s_{i_1}-j_2s_{i_2}})\eta'_{t_0}(\boldsymbol{s})\mathrm{d}\boldsymbol{s}+\cdots+\sum_{j_1=1}^{q_1}\cdots\sum_{j_d=1}^{q_d}\binom{q_1}{j_1}\cdots\binom{q_d}{j_d}(-1)^{\langle\boldsymbol{1},\boldsymbol{j}\rangle}$$

$$-\epsilon\int_{(\mathbb{R}^+)^d}e^{-\langle\boldsymbol{r},\boldsymbol{s}\rangle}\sum_{j_1=1}^{q_1}\cdots\sum_{j_d=1}^{q_d}\binom{q_1}{j_1}\cdots\binom{q_d}{j_d}(-1)^{\langle\boldsymbol{1},\boldsymbol{j}\rangle}(1-e^{-\langle\boldsymbol{j},\boldsymbol{s}\rangle})\eta'_{t_0}(\boldsymbol{s})\mathrm{d}\boldsymbol{s}$$

$$\left.+o(\epsilon)\right\}$$

$$=e^{-\triangle_{t_0-\epsilon}^{t_0}\psi_t(\boldsymbol{r})}\left\{\epsilon\int_{(\mathbb{R}^+)^d}e^{-\langle\boldsymbol{r},\boldsymbol{s}\rangle}(1-e^{-s_1})^{q_1}\cdots(1-e^{-s_d})^{q_d}\eta'_{t_0}(\boldsymbol{s})\mathrm{d}\boldsymbol{s}+o(\epsilon)\right\}$$

$$=\{1+o(1)\}\left\{\epsilon\int_{(\mathbb{R}^+)^d}e^{-\langle\boldsymbol{r},\boldsymbol{s}\rangle}(1-e^{-s_1})^{q_1}\cdots(1-e^{-s_d})^{q_d}\eta'_{t_0}(\boldsymbol{s})\mathrm{d}\boldsymbol{s}+o(\epsilon)\right\}$$

$$=\left\{\epsilon\int_{(\mathbb{R}^+)^d}e^{-\langle\boldsymbol{r},\boldsymbol{s}\rangle}(1-e^{-s_1})^{q_1}\cdots(1-e^{-s_d})^{q_d}\eta'_{t_0}(\boldsymbol{s})\mathrm{d}\boldsymbol{s}+o(\epsilon)\right\} \qquad \square$$

Define

$$\Gamma_{\boldsymbol{D},\epsilon}=\bigcap_{i=1}^{d}\bigcap_{j=1}^{k}\left\{((t_1^{(i)},\delta_1^{(i)},\ldots,t_{n_1}^{(i)},\delta_{n_1}^{(i)})\right.$$

$$\left.:m_i^c\left(\{T_{(j)}\}\right)=n_{i,j}^c,m_i^e\left((T_{(j)}-\epsilon,T_{(j)}]\right)=n_{i,j}^e\right\}$$

so that

$$\mathbb{E}\left[e^{-\lambda_1\mu_1(0,t]-\cdots\lambda_d\mu_d(0,t]}|\boldsymbol{D}\right]=\lim_{\epsilon\to0}\frac{\mathbb{E}\left[e^{-\lambda_1\mu_1(0,t]-\cdots-\lambda_d\mu_d(0,t]}\mathbb{1}_{\Gamma_{\boldsymbol{D},\epsilon}}(\boldsymbol{D})\right]}{\mathbb{P}[\boldsymbol{D}\in\Gamma_{\boldsymbol{D},\epsilon}]}$$

We observe that defining $T_{(0)}=0$, $\bar{n}_{i,k+1}^e=0$ for $i\in\{1,\ldots,d\}$ and selecting ϵ sufficiently small such that $t\notin(T_{(j)}-\epsilon,T_{(j)})$ for all $j\in\{1,\ldots,k\}$

$$\mathbb{E}\left[\mathrm{e}^{-\lambda_1\mu_1(0,t]-\cdots-\lambda_d\mu_d(0,t]}\mathbb{1}_{\Gamma_{\boldsymbol{D},\epsilon}}(\boldsymbol{D})\,|(\mu_1,\ldots,\mu_d)\right]$$

$$=\prod_{i=1}^{d}\mathrm{e}^{-\lambda_i\mu_i(0,t]}\prod_{j=1}^{k}\mathrm{e}^{-n_{i,j}^c\mu_i(0,T_{(j)}]-n_{i,j}^e\mu_i(0,T_{(j)}-\epsilon]}\left(1-\mathrm{e}^{-\mu_i(T_{(j)}-\epsilon,T_{(j)}]}\right)^{n_{i,j}^e}$$

$$=\prod_{i=1}^{d}\mathrm{e}^{-\lambda_i\mathbb{1}_{(0,t]}(T_{(k)})\mu_i(T_{(k)},t]}\prod_{j=1}^{k}\left\{\mathrm{e}^{-\lambda_i\mathbb{1}_{(0,t)}(T_{(j-1)})\mu_i(T_{(j-1)},\min\{t,T_{(j)}-\epsilon\}]}\right.$$

$$\times\,\mathrm{e}^{-\lambda_i\mathbb{1}_{(0,t]}(T_{(j)})\mu_i(T_{(j)}-\epsilon,T_{(j)}]-n_{i,j}^c\sum_{r=1}^{j}\left(\mu_i(T_{(r)}-\epsilon,T_{(r)}]+\mu_i(T_{(r-1)},T_{(r)}-\epsilon]\right)}$$

$$\times\,\mathrm{e}^{-n_{i,j}^e\sum_{r=1}^{j}\mu_i(T_{(r-1)},T_{(r)}-\epsilon]-n_{i,j}^e\sum_{r=1}^{j-1}\mu_i(T_{(r)}-\epsilon,T_{(r)}]}$$

$$\times\,\left.\left(1-\mathrm{e}^{-\mu_i(T_{(j)}-\epsilon,T_{(j)}]}\right)^{n_{i,j}^e}\right\}$$

$$=\prod_{i=1}^{d}\mathrm{e}^{-\lambda_i\mathbb{1}_{(0,t]}(T_{(k)})\mu_i(T_{(k)},t]-\sum_{j=1}^{k}n_{i,j}^c\sum_{r=1}^{j}\left(\mu_i(T_{(r)}-\epsilon,T_{(r)}]+\mu_i(T_{(r-1)},T_{(r)}-\epsilon]\right)}$$

$$\times\,\mathrm{e}^{-\sum_{j=1}^{k}n_{i,j}^e\sum_{r=1}^{j}\mu_i(T_{(r-1)},T_{(r)}-\epsilon]-\sum_{j=1}^{k}n_{i,j}^e\sum_{r=1}^{j-1}\mu_i(T_{(r)}-\epsilon,T_{(r)}]}$$

$$\times\,\prod_{j=1}^{k}\left\{\mathrm{e}^{-\lambda_i\mathbb{1}_{(0,t)}(T_{(j-1)})\mu_i(T_{(j-1)},\min\{t,T_{(j)}-\epsilon\}]-\lambda_i\mathbb{1}_{(0,t]}(T_{(j)})\mu_i(T_{(j)}-\epsilon,T_{(j)}]}\right.$$

$$\left.\left(1-\mathrm{e}^{-\mu_i(T_{(j)}-\epsilon,T_{(j)}]}\right)^{n_{i,j}^e}\right\}$$

$$=\prod_{i=1}^{d}\prod_{j=1}^{k}\left\{\mathrm{e}^{-\left[\lambda_i\mathbb{1}_{(0,t]}(T_{(j)})+\bar{n}_{i,j}^c+\bar{n}_{i,j+1}^e\right]\mu_i(T_{(j)}-\epsilon,T_{(j)}]}\left(1-\mathrm{e}^{-\mu_i(T_{(j)}-\epsilon,T_{(j)}]}\right)^{n_{i,j}^e}\right\}$$

$$\times\,e^{-\lambda_i\mathbb{1}_{(0,t]}(T_{(k)})\mu_i(T_{(k)},t]}$$

$$\times\,\prod_{j=1}^{k}\left\{\mathrm{e}^{-\lambda_i\mathbb{1}_{(0,t)}(T_{(j-1)})\mu_i(T_{(j-1)},\min\{t,T_{(j)}-\epsilon\}]}\right.$$

$$\times\,\left.\mathrm{e}^{-\bar{n}_{i,j}^c\mu_i(T_{(j-1)},T_{(j)}-\epsilon]-\bar{n}_{i,j}^e\mu_i(T_{(j-1)},T_{(j)}-\epsilon]}\right\}$$

So defining

$$I_{1,\epsilon}=\prod_{j=1}^{k}\prod_{i=1}^{d}\left\{\mathrm{e}^{-\left[\lambda_i\mathbb{1}_{(0,t]}(T_{(j)})+\bar{n}_{i,j}^c+\bar{n}_{i,j+1}^e\right]\mu_i(T_{(j)}-\epsilon,T_{(j)}]}\left(1-\mathrm{e}^{-\mu_i(T_{(j)}-\epsilon,T_{(j)}]}\right)^{n_{i,j}^e}\right\}$$

$$I_{2,\epsilon}=\prod_{i=1}^{d}\mathrm{e}^{-\lambda_i\mathbb{1}_{(0,t]}(T_{(k)})\mu_i(T_{(k)},t]}\prod_{j=1}^{k}\left\{\mathrm{e}^{-\lambda_i\mathbb{1}_{(0,t)}(T_{(j-1)})\mu_i(T_{(j-1)},\min\{t,T_{(j)}-\epsilon\}]}\right.$$

$$\left.\times e^{-(\bar{n}_{i,j}^c+\bar{n}_{i,j}^e)\mu_i(T_{(j-1)},T_{(j)}-\epsilon]}\right\}$$

We get from the independence property of CRM's that

$$\mathbb{E}\left[\mathrm{e}^{-\lambda_1\mu_1(0,t]-\cdots-\lambda_d\mu_d(0,t]}\mathbb{1}_{\Gamma_{\boldsymbol{D},\epsilon}}(\boldsymbol{D})\right]=\mathbb{E}[I_{1,\epsilon}]\,\mathbb{E}[I_{2,\epsilon}]\qquad\text{(D.23)}$$

We observe that for $r_i = \lambda_i \mathbb{1}_{(0,t]}(T_{(j)}) + \bar{n}_{i,j}^c + \bar{n}_{i,j+1}^e$, $i \in \{1, \ldots, d\}$ we have that $\min\{r_1, \ldots, r_d\} \geq 1$ and for $j \in \{1, \ldots, k\}$ such that $T_{(j)}$ is an exact observation we have that $\max\{n_{1,j}, \ldots, n_{d,j}\} \geq 1$ so Lemma 2 can be applied yielding

$$
\mathbb{E}\left[\prod_{i=1}^{d}\left\{e^{-\left[\lambda_i \mathbb{1}_{(0,t]}(T_{(j)}) + \bar{n}_{i,j}^c + \bar{n}_{i,j+1}^e\right]\mu_i(T_{(j)} - \epsilon, T_{(j)}]}\left(1 - e^{-\mu_i(T_{(j)} - \epsilon, T_{(j)}]}\right)^{n_{i,j}^e}\right\}\right]
$$

$$
= \epsilon \int_{(\mathbb{R}^+)^d} \prod_{i=1}^{d}\left\{e^{-[\lambda_i \mathbb{1}_{(0,t]}(T_{(j)}) + \bar{n}_{i,j}^c + \bar{n}_{i,j+1}^e]s_i}(1 - e^{-s_i})^{n_{i,j}^e}\right\} \eta'_t T_{(j)}(\boldsymbol{s})\mathrm{d}\boldsymbol{s} + o(\epsilon)
$$
(D.24)

On the other hand, for $j \notin \mathcal{J} = \{j : T_{(j)}$ is an exact observation$\}$ we have $n_{i,j}^e = 0$ so by the continuity of $\eta_t(\boldsymbol{s})$ in t we have

$$
\lim_{\epsilon \to 0}\mathbb{E}\left[\prod_{i=1}^{d}\left\{e^{-\left[\lambda_i \mathbb{1}_{(0,t]}(T_{(j)}) + \bar{n}_{i,j}^c + \bar{n}_{i,j+1}^e\right]\mu_i(T_{(j)} - \epsilon, T_{(j)}]}\left(1 - e^{-\mu_i(T_{(j)} - \epsilon, T_{(j)}]}\right)^{n_{i,j}^e}\right\}\right]
$$

$$
= \lim_{\epsilon \to 0}\mathbb{E}\left[\prod_{i=1}^{d}\left\{e^{-\left[\lambda_i \mathbb{1}_{(0,t]}(T_{(j)}) + + \bar{n}_{i,j}^c + \bar{n}_{i,j+1}^e\right]\mu_i(T_{(j)} - \epsilon, T_{(j)}]}\right\}\right] = 1 \qquad \text{(D.25)}
$$

From (D.24), (D.25) and the independence property of CRM's we obtain

$$
\lim_{\epsilon \to 0}\mathbb{E}[I_{1,\epsilon}] = \lim_{\epsilon \to 0}\prod_{j \in \mathcal{J}}\left\{\epsilon \int_{(\mathbb{R}^+)^d}\prod_{i=1}^{d}\left\{e^{-[\lambda_i \mathbb{1}_{(0,t]}(T_{(j)}) + \bar{n}_{i,j}^c + \bar{n}_{i,j+1}^e]s_i}\right.\right.
$$

$$
\left.\left. \times \ (1 - e^{-s_i})^{n_{i,j}^e}\right\} \eta'_{T_{(j)}}(\boldsymbol{s})\mathrm{d}\boldsymbol{s} + o(\epsilon)\right\}
$$

Also by continuity and independence, defining $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_d)$, we get

$$
\lim_{\epsilon \to 0}\mathbb{E}[I_{2,\epsilon}] = e^{-[\psi_t(\mathbb{1}_{(0,t]}(T_{(k)})\boldsymbol{\lambda}) - \psi_{T_{(k)}}(\mathbb{1}_{(0,t]}(T_{(k)})\boldsymbol{\lambda})]} \times
$$

$$
\times \prod_{j=1}^{k}\left\{e^{-[\psi_{t \wedge T_{(j)}}(\mathbb{1}_{(0,t]}(T_{(j-1)})\boldsymbol{\lambda} + \bar{\boldsymbol{n}}_j^c + \bar{\boldsymbol{n}}_j^e) - \psi_{T_{(j-1)}}(\mathbb{1}_{(0,t]}(T_{(j-1)})\boldsymbol{\lambda} + \bar{\boldsymbol{n}}_j^c + \bar{\boldsymbol{n}}_j^e)]}\right.
$$

$$
\left. \times e^{-[\psi_{T_{(j)}}(\bar{\boldsymbol{n}}_j^c + \bar{\boldsymbol{n}}_j^e) - \psi_{t \wedge T_{(j)}}(\bar{\boldsymbol{n}}_j^c + \bar{\boldsymbol{n}}_j^e)]}\right\}
$$

So by (D.23), (D.25) and (D.24) we get that

$$
\lim_{\epsilon \to 0}\mathbb{E}\left[e^{-\lambda_1\mu_1(0,t] - \cdots - \lambda_d\mu_d(0,t]}\mathbb{1}_{\Gamma_{\boldsymbol{D},\epsilon}}(\boldsymbol{D})\right] = e^{-\triangle_{T_{(k)}}^{t}\psi_t(\mathbb{1}_{(0,t]}(T_{(k)})\boldsymbol{\lambda})}
$$

$$
\times e^{-\sum_{j=1}^{k}\triangle_{T_{(j-1)}}^{t \wedge T_{(j)}}\psi_t(\mathbb{1}_{(0,t]}(T_{(j-1)})\boldsymbol{\lambda} + \bar{\boldsymbol{n}}_j^c + \bar{\boldsymbol{n}}_j^e)}
$$

$$
\times \prod_{j \in \mathcal{J}}\lim_{\epsilon \to 0}\left\{\epsilon \int_{(\mathbb{R}^+)^d}\prod_{i=1}^{d}\left\{e^{-[\lambda_i \mathbb{1}_{(0,t]}(T_{(j)}) + \bar{n}_{i,j}^c + \bar{n}_{i,j+1}^e]s_i}\right.\right.
$$

$$\times\ (1 - \mathrm{e}^{-s_i})^{n_{i,j}^e}\Big\}\ \eta'_{T_{(j)}}(\boldsymbol{s})\mathrm{d}\boldsymbol{s} + o(\epsilon)\Big\}$$

$$\times\ \mathrm{e}^{-\sum_{j=1}^k \triangle_{t\wedge T_{(j)}}^{T_{(j)}}\psi_t\big(\bar{\boldsymbol{n}}_j^c + \bar{\boldsymbol{n}}_j^e\big)}$$

And similarly

$$\lim_{\epsilon\to0}\mathbb{P}[\boldsymbol{D}\in\Gamma_{\boldsymbol{D},\epsilon}] = \mathrm{e}^{-\sum_{j=1}^k \triangle_{T_{(j-1)}}^{T_{(j)}}\psi_t\big(\bar{\boldsymbol{n}}_j^c + \bar{\boldsymbol{n}}_j^e\big)}$$

$$\times\ \prod_{j\in\mathcal{J}}\lim_{\epsilon\to0}\left\{\epsilon\int_{(\mathbb{R}^+)^d}\prod_{i=1}^d\left\{\mathrm{e}^{-(\bar{n}_{i,j}^c + \bar{n}_{i,j+1}^e)s_i}(1 - \mathrm{e}^{-s_i})^{n_{i,j}^e}\right\}\eta'_t T_{(j)}(\boldsymbol{s})\mathrm{d}\boldsymbol{s} + o(\epsilon)\right\}$$

We set $T_{(k+1)} = \infty$ so we conclude

$$\mathbb{E}\Big[\mathrm{e}^{-\lambda_1\mu_1(0,t]-\cdots-\lambda_d\mu_d(0,t]}|\boldsymbol{D}\Big] = \lim_{\epsilon\to0}\frac{\mathbb{E}\big[\mathrm{e}^{-\lambda_1\mu_1(0,t]-\cdots-\lambda_d\mu_d(0,t]}\mathbb{1}_{\Gamma_{\boldsymbol{D},\epsilon}}(\boldsymbol{D})\big]}{\mathbb{P}[\boldsymbol{D}\in\Gamma_{\boldsymbol{D},\epsilon}]}$$

$$= \mathrm{e}^{-\sum_{j=1}^{k+1}\triangle_{T_{(j-1)}}^{t\wedge T_{(j)}}\big[\psi_t\big(\mathbb{1}_{(0,t]}(T_{(j-1)})\boldsymbol{\lambda} + \bar{\boldsymbol{n}}_j^c + \bar{\boldsymbol{n}}_j^e\big) - \psi_t\big(\bar{\boldsymbol{n}}_j^c + \bar{\boldsymbol{n}}_j^e\big)\big]}\prod_{j\in\mathcal{J}}\lim_{\epsilon\to0}\Bigg\{$$

$$\frac{\epsilon\int_{(\mathbb{R}^+)^d}\prod_{i=1}^d\left\{\mathrm{e}^{-[\lambda_i\mathbb{1}_{(0,t]}(T_{(j)})+\bar{n}_{i,j}^c+\bar{n}_{i,j+1}^e]s_i}(1 - \mathrm{e}^{-s_i})^{n_{i,j}^e}\right\}\eta'_{T_{(j)}}(\boldsymbol{s})\mathrm{d}\boldsymbol{s} + o(\epsilon)}{\epsilon\int_{(\mathbb{R}^+)^d}\prod_{i=1}^d\left\{\mathrm{e}^{-[\bar{n}_{i,j}^c+\bar{n}_{i,j+1}^e]s_i}(1 - \mathrm{e}^{-s_i})^{n_{i,j}^e}\right\}\eta'_{T_{(j)}}(\boldsymbol{s})\mathrm{d}\boldsymbol{s} + o(\epsilon)}\Bigg\}$$

$$= \mathrm{e}^{-\sum_{j=1}^{k+1}\int_{(\mathbb{R}^+)^d\times(T_{(j-1)},t\wedge T_{(j)}]}\mathbb{1}_{(0,t]}(T_{(j-1)})\big(1-\mathrm{e}^{-\langle\boldsymbol{\lambda},\boldsymbol{s}\rangle}\big)\mathrm{e}^{-\langle\bar{\boldsymbol{n}}_j^c+\bar{\boldsymbol{n}}_j^e,\boldsymbol{s}\rangle}\nu(\mathrm{d}\boldsymbol{s},\mathrm{d}u)}$$

$$\times\ \prod_{j\in\mathcal{J}}\left\{\frac{\int_{(\mathbb{R}^+)^d}\prod_{i=1}^d\left\{\mathrm{e}^{-[\lambda_i\mathbb{1}_{(0,t]}(T_{(j)})+\bar{n}_{i,j}^c+\bar{n}_{i,j+1}^e]s_i}(1 - \mathrm{e}^{-s_i})^{n_{i,j}^e}\right\}\eta'_{T_{(j)}}(\boldsymbol{s})\mathrm{d}\boldsymbol{s}}{\int_{(\mathbb{R}^+)^d}\prod_{i=1}^d\left\{\mathrm{e}^{-[\bar{n}_{i,j}^c+\bar{n}_{i,j+1}^e]s_i}(1 - \mathrm{e}^{-s_i})^{n_{i,j}^e}\right\}\eta'_{T_{(j)}}(\boldsymbol{s})\mathrm{d}\boldsymbol{s}}\right\}$$

A.5. Simulation algorithm

We use a Metropolis within Gibbs sampler to draw simulations from $\sigma|\boldsymbol{D}$ and $A|\boldsymbol{D}$ as in Section 4. We recall that Corollary 2 gives the likelihood $l(\sigma, A; \boldsymbol{D})$ and we denote p_σ, p_A for the prior distributions of σ and A as in Section 4. Given initial values $\sigma^{(0)}$, $A^{0)}$, the algorithm is as follows

(1) Draw $A^{(i+1)}$ from a Metropolis-Hastings sampler with proposal distribution $g(x'|x) \sim \text{Log-Norm}(\log(x), 1)$ and target distribution

$$l(\sigma^{(i)}, x; \boldsymbol{D})p_A(x).$$

(2) Draw $\sigma^{(i+1)}$ from a Metropolis-Hastings sampler with Uniform proposal distribution and target distribution

$$l(x, A^{(i+1)}; \boldsymbol{D})p_\sigma(x).$$

For the fits in Section 4 we used 100 iterations for each inner Metropolis-Hasting sampler and 1000 iterations for the overall Gibbs sampler.

A.6. Survival function of $F_{\theta,\lambda}$.

Let $\mathcal{C}_{\theta,d}$ be a d-dimensional distributional Clayton copula and $\tilde{F}_i$, $i = 1, \ldots, d$, a collection of marginal cumulative distribution functions; then the survival function associated to the Clayton distributional copula and marginals is given by

$$S(x_1, \ldots, x_d) = 1 - \sum_{i=1}^{d} \tilde{F}_i(x_i) + \sum_{j=2}^{d} (-1)^j \sum_{\substack{i_1, \ldots, i_j \in \{1, \ldots, d\} \\ i_1 < \ldots < i_j}} \mathcal{C}_{\theta,j}(x_{i_1}, \ldots, x_{i_j}),$$

see Section 2.6 in Nelsen (2013).

Acknowledgments

References

Aalen, O., Borgan, O., and Gjessing, H. (2008). *Survival and event history analysis: a process point of view.* Springer Science & Business Media. MR2449233

Cont, R. and Tankov, P. (2004). *Financial modelling with jump processes.* Chapman & Hall. MR2042661

De Finetti, B. (1938). 'Sur la condition de "equivalence partielle"', Colloque consacré à la théorie des probabilités, Vol. VI, Université de Genève, Hermann et C. ie, Paris. MR0000909

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

Doksum, K. (1974). Tailfree and neutral random probabilities and their posterior distributions. *The Annals of Probability*, **2**, 183–201. MR0373081

Dykstra, R. L. and Laud, P. (1981). A Bayesian nonparametric approach to reliability. *The Annals of Statistics*, **9**, 356–367. MR0606619

Epifani, I. and Lijoi. A. (2010). Nonparametric priors for vectors of survival functions. *Statistica Sinica*, **20**, 1455–1484. MR2777332

Ferguson, T. S., and Phadia, E. G. (1979). Bayesian nonparametric estimation based on censored data. *The Annals of Statistics*, **7**, 163–186. MR0515691

Griffin, J. and Leisen, F. (2017). Compound random measures and their use in Bayesian nonparametrics. *Journal of the Royal Statistical Society - Series B*, **79**, 525–545. MR3611758

Ishwaran, H., and James, L. F. (2004). Computational methods for multiplicative intensity models using weighted gamma processes: proportional hazards, marked point processes, and panel count data. *Journal of the American Statistical Association*, **99**, 175–190. MR2054297

Kallsen, J. and Tankov, P. (2006). Characterization of dependence of multidimensional Lèvy processes using Lèvy copulas. *Journal of Multivariate Analysis*, **97**, 1551–1572. MR2275419

Kingman, J. (1967). Completely random measures. *Pacific Journal of Mathematics*, **21**, 59–78. MR0210185

Leisen, F., and Lijoi, A. (2011) Vectors of two-parameter Poisson–Dirichlet processes. *Journal of Multivariate Analysis*, **102**, 482–495. MR2755010

Leisen F., Lijoi A. and Spano D. (2013). A Vector of Dirichlet processes. *Electronic Journal of Statistics*, **7**, 62–90. MR3020414

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

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

MacEachern S. N. (1999). Dependent nonparametric processes. In *ASA Proceedings of the Section on Bayesian Statistical Science*, Alexandria, VA: American Statistical Association.

Nelsen, Roger B. (2013). *An introduction to copulas.* Springer Science & Business Media, **139**. MR1653203

Nieto-Barajas, L. E. (2014). Bayesian semiparametric analysis of short-and long-term hazard ratios with covariates. *Computational Statistics and Data Analysis*, **71**, 477–490. MR3131984

Zhu, W., and Leisen, F. (2015). "A multivariate extension of a vector of two-parameter Poisson–Dirichlet processes." *Journal of Nonparametric Statistics*, **27**, 89–105. MR3304361