Bayesian Nonparametric Methods for Heterogeneous Data

Alan Riva Palacio Cohen

School of Mathematics Statistics and Actuarial Sciences

University of Kent

Abstract: Various types of data heterogeneity can be modelled in a Bayesian nonparametric setting by using vectors of random measures. Special interest is placed on the class of compound random measures, for which novel results are presented. Regression models allowing for multiple-sample information in survival analysis with heterogeneous data are studied.

This dissertation is submitted for the degree of

Doctor of Philosophy

125 pages / 38923 words

June 2019

I would like to dedicate this thesis to Roberto Bolaño and my grandmother, whom would have gotten along great.

Acknowledgements

I would like to acknowledge all the help and support from my main supervisor Dr Fabrizio Leisen; it was a great experience to work with him and learn from his thoughtful and empathic advices. I also want to thank the support of my second supervisor Professor Jim Griffin, and thank Dr Ramses Mena who encouraged me to pursue my PhD studies. I want to thank all my friends in SMSAS who helped me have a better experience in the UK and my family that has always supported me. I am grateful for the work from the people in the school, specially Derek Baldwin and Claire Carter. I thank all my friends who travelled long distances to visit me, specially Berenice who was there for me for the most part of my PhD.

Table of contents

List of figures

Nomenclature

Symbols

$\mathscr{B}(A)$ Borel sigma-algebra associated to the set A

$\delta_x(\cdot)$ Dirac measure at x

$\mathrm{Leb}(\cdot)$ Lebesgue measure

$\mathscr{A} \otimes \mathscr{B}$ sigma-algebra generated by the cross product $\mathscr{A} \times \mathscr{B}$ of two sigma-algebras

$\langle \boldsymbol{u}, \boldsymbol{v} \rangle$ Inner product between two vectors in $\mathbb{R}^d$

$\{\boldsymbol{e}_i\}_{i=1}^d$ Canonical basis in $\mathbb{R}^d$

$\mathbf{1} = \sum_{i=1}^d \boldsymbol{e}_i$ Vector of number one's in $\mathbb{R}^d$

$I^{(d)}$ $d \times d$ identity matrix

$\triangle_{s_1}^{s_2} f_t = f_{s_2} - f_{s_1}$ Difference operator for a function f_t with $t \in \mathbb{R}$

Acronyms and abbreviations

CRM Completely random measure, page 2

PRM Poisson random measure, page 4

i.i.d. Independent, identically distributed, page 5

r.v. Random variable, page 5

VCRM Vector of completely random measures, page 11

NTR Neutral to the right, page 13

MCMC Markov Chain Monte Carlo, page 16

CoRM Compound random measures, page 31

Chapter 1

Introduction and preliminaries

1.1 Thesis introduction and overview

The present work focuses on the use of random measures for the modelling of complex data which presents heterogeneity in various aspects. We will mainly focus on two kinds of heterogeneity.

1) Heterogeneity due to belonging to different populations or samples.

2) Heterogeneity due to having different covariate values.

In the present chapter we present the necessary preliminaries for the rest of the thesis, including the notion of completely random measures, vectors of completely random measures and neutral to the right distributions. Vectors of completely random measure are the fundamental mathematical tool for the work presented throughout the thesis; while the neutral to the right distribution are of key importance for the models introduced in chapter 4 and 5. Chapter 2 revolves around the concept of completely random measures which are a flexible and manageable class of vectors of completely random measures that has been previously introduced and used for Bayesian analysis of heterogeneous data in the literature. In this chapter we present recently published results regarding the integrability conditions of compound random measure and present a variety of results, including the introduction of a new compound random measure, related to the LogNormal distribution, and a new formula for the associated Laplace exponent which allows for an approximation via Monte-Carlo methods. In Chapter 3 we focus on the link between compound random measures and Lévy copulas, which are a framework for the modelling of vectors of completely random measures that has been very popular in the literature. Of main importance is the introduction of a new class of Lévy copulas which is a generalization of the widely used Clayton Lévy copula. Many models for heterogeneous data in the framework of Lévy copulas have been proposed and can be used with the specific choice of our extended Clayton Lévy copula proposal. Chapter 4 presents a recently published model which generalizes neutral to the right distributions for multiple-sample information. This model addresses heterogeneity as in point 1) above, where different populations can exhibit different behaviours which can be dependent among each other. Finally, in Chapter 5 we introduce a new model for regression in survival analysis where the two heterogeneity points above, 1)-2), are modelled. We argue that such model can even be used for the identification of different populations which keep the symmetry of exchangeability and are determined by the covariate values of the individuals.

1.2 Completely random measures and vectors thereof

Even though the main application of this thesis is in survival analysis where we consider survival times that correspond to random variables supported in the real positive axis $\mathbb{R}^+ = (0,\infty)$, results in Chapter 2 are given for more general spaces than $\mathbb{R}^+$. With that in mind, we present this section's preliminaries in a general setting. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and $\mathbb{X}$ a Polish space with corresponding Borel σ-algebra $\mathscr{X}$. We call a measure μ on $(\mathbb{X}, \mathscr{X})$ boundedly finite if $\mu(A) < \infty$ for any bounded set $A \in \mathscr{X}$; we denote by $\mathbb{M}_{\mathbb{X}}$ the space of boundedly finite measures on the measurable space $(\mathbb{X}, \mathscr{X})$ and by $\mathscr{M}_{\mathbb{X}}$ the associated Borel σ-algebra, see Appendix 2 in Daley and Vere-Jones (2007) for technical details. A random measure is a measurable function from the probability space $(\Omega, \mathscr{F}, \mathbb{P})$ onto $(\mathbb{M}_{\mathbb{X}}, \mathscr{M}_{\mathbb{X}})$. We will focus on the class of *completely random measures* as introduced by Kingman (1967).

Definition 1. A random measure μ on $(\mathbb{X}, \mathscr{X})$ is called a *completely random measure* (CRM) if for any $n > 1$ and disjoint sets $A_1, \dots, A_n \in \mathscr{X}$ the random variables $\mu(A_1), \dots, \mu(A_n)$ are mutually independent.

A CRM μ has the following representation, see 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, i.e.

$$\mu_{fl} = \sum_{i=1}^{\infty} W_i^{(\mathrm{fl})} \delta_{x_i^{(\mathrm{fl})}},$$

with $\{x_i^{(\mathrm{fl})}\}_{i=1}^{\infty} \subset \mathbb{X}$ the fixed jump locations and $\{W_i^{(\mathrm{fl})}\}_{i=1}^{\infty} \subset \mathbb{R}^+$ the random jump heights; and μ_r is a measure made of jumps with random locations and random heights, i.e.

$$\mu_r = \sum_{i=1}^{\infty} W_i^{(\mathrm{r})} \delta_{X_i^{(\mathrm{r})}},$$

where $\{X_i^{(\mathrm{r})}\}_{i=1}^{\infty} \subset \mathbb{X}$ are the random jump locations and $\{W_i^{(\mathrm{r})}\}_{i=1}^{\infty} \subset \mathbb{R}^+$ the random jump heights. The measures μ_d, μ_{fl} and μ_r are mutually independent. In this thesis we will consider CRM's without the deterministic part, μ_d; even more we will treat the part with fixed locations separately. So by a CRM we will refer to a measure which has the form as μ_r above. Such CRM's are characterized by their Laplace transform

$$\mathbb{E}\left[e^{-\mu(f)}\right] = e^{-\int_{\mathbb{R}^+\times\mathbb{X}}(1-e^{-f(x)s})\nu(\mathrm{d}s,\mathrm{d}x)} \tag{1.1}$$

for $f : \mathbb{X} \to \mathbb{R}^+$ such that $\mu(f) < \infty$, where $\mu(f) = \int_{\mathbb{X}} f(x)\mu(\mathrm{d}x)$, and $\nu(\mathrm{d}s,\mathrm{d}x)$ is a measure in $(\mathbb{R}^+ \times \mathbb{X}, \mathscr{B}(\mathbb{R}^+) \otimes \mathscr{X})$ such that

$$\int_{\mathbb{R}^+\times X} \min\{1,s\}\nu(\mathrm{d}s,\mathrm{d}x) < \infty \tag{1.2}$$

for any bounded set $X \in \mathscr{X}$. A measure ν satisfying the condition displayed in equation (1.2) is called the *Lévy intensity* of μ. We say that ν is homogeneous when

$$\nu(\mathrm{d}s,\mathrm{d}x) = \rho(\mathrm{d}s)\kappa(\mathrm{d}x) \tag{1.3}$$

with ρ a measure in $(\mathbb{R}^+, \mathscr{B}(\mathbb{R}^+))$ referring to the jump heights and κ a non-atomic measure in $(\mathbb{X}, \mathscr{X})$ referring to the jump locations. From the integrability condition equation 1.2, we observe that if κ is a boundedly finite measure in $\mathbb{X}$, i.e. for any bounded set $X \in \mathscr{X}$ $\kappa(X) < \infty$, then ρ must satisfy the condition 1.2 by its own. All the homogeneous Lévy intensities in the present work will be such that the associated κ measure is boundedly finite so the associated ρ measure is such that $\rho(\mathrm{d}s)\mathrm{d}x$ is a Lévy intensity. If in the Laplace transform (1.1) we set $f(x) = \lambda \mathbb{1}_{\{x\in A\}}$ with $\lambda \in \mathbb{R}^+$ and A a bounded set in $\mathscr{X}$, we get

$$\mathbb{E}\left[e^{-\lambda\mu(A)}\right] = e^{-\int_{\mathbb{R}^+\times A}(1-e^{-\lambda s})\nu(\mathrm{d}s,\mathrm{d}x)}. \tag{1.4}$$

Such Laplace transform is often sufficient to perform the majority of the calculations in this work. With that in mind we define the Laplace exponent

$$\psi_t(\lambda) = \int_{\mathbb{R}^+ \times (0,t]} (1 - e^{-\lambda s}) \nu(\mathrm{d}s, \mathrm{d}x). \tag{1.5}$$

If the Lévy intensity of interest is homogeneous as in (1.3) then the corresponding Laplace exponent can be written as $\psi_t(\lambda) = \gamma(t)\psi(\lambda)$ where $\gamma(t) = \int_0^t \kappa(\mathrm{d}x)$ and $\psi(\lambda) = \int_0^\infty (1 - e^{-\lambda s})\rho(\mathrm{d}s)$. Sometimes it is useful to write a CRM in terms of a *Poisson random measure* (PRM). Following Sato (1999), we define the latter measure as follows

Definition 2. Let ν be a σ-finite measure in $(\mathbb{X}, \mathscr{X})$. A Poisson random measure in $(\mathbb{X}, \mathscr{X})$ is a completely random measure N such that if $C \in \mathscr{X}$ then $N(C)$ is a Poisson random variable with intensity $\nu(C)$.

Given a CRM μ with Lévy intensity ν we can set a Poisson random measure N in $(\mathbb{R}^+ \times \mathbb{X}, \mathscr{B}(\mathbb{R}^+) \otimes \mathscr{X})$ such that by the Lévy-Ito decomposition, see Sato (1999), we have that

$$\mu(A) = \int_{\mathbb{R}^+ \times A} sN(\mathrm{d}s, \mathrm{d}x). \tag{1.6}$$

An advantage of the above representation is that it allows us to use PRM's results in a CRM setting. A key result for PRM's which will be of use later on follows

Proposition 1. *(Rosiński (2001)) Let M and N be PRM's defined on possibly different probability spaces and taking values in Polish spaces $\mathbb{S}$, $\mathbb{T}_0$, with respective Borel σ-algebras $\mathscr{S}$, $\mathscr{T}_0$ and intensity measures η, ν. We suppose that $\mathbb{T}_0 \subset \mathbb{T}$ for some Polish space $\mathbb{T}$, with Borel σ-algebra $\mathscr{T}$, and that we have a measurable function $h : \mathbb{S} \to \mathbb{T}$ such that*

$$\nu = \eta \circ h^{-1} \text{ on } \mathscr{T}_0$$

then

$$N \overset{d}{=} M \circ h^{-1}$$

In addition, if N is defined in a probability space rich enough to allow the existence of a standard uniform random variable that is independent of N and

$$M = \sum_{i=1}^{\infty} \delta_{S_i}$$

for some $\mathbb{S}$*-valued random elements* S_i*,* $i \geq 1$*; then there exists a sequence* $\{\tilde{S}_i\}_{i=1}^{\infty}$ *of* $\mathbb{S}$*-valued random elements defined in the same probability space as* N *and such that*

$$\{\tilde{S}_i\}_{i=1}^{\infty} \stackrel{d}{=} \{S_i\}_{i=1}^{\infty}$$

and

$$N \stackrel{a.s.}{=} \sum_{i=1}^{\infty} \delta_{h(\tilde{s}_i)}.$$

In the next example we use the above proposition to construct a series representation for a CRM with homogeneous Lévy intensity.

Example. Let there be sequences of independent, identically distributed (i.i.d.) random variables (r.v.'s)

$$\begin{aligned} U_i &\stackrel{\text{i.i.d.}}{\sim} \text{Uniform}(0,1), \quad i \in \{1,2,\ldots\} \\ \Gamma_i &\stackrel{\text{i.i.d.}}{\sim} \text{Exponential}(1), \quad i \in \{1,2,\ldots\} \\ \hat{\Gamma}_i &= \sum_{j=1}^{i} \Gamma_i, \quad i \in \{1,2,\ldots\}, \end{aligned}$$

$\nu(\mathrm{d}s,\mathrm{d}x) = \rho(\mathrm{d}s)\mathrm{d}x$ be a Lévy intensity in $\mathbb{R}^+ \times \mathbb{R}^+$ with ρ a Borel measure in $\mathbb{R}^+$ and denote the Lebesgue measure in $\mathbb{R}^+$ as $\text{Leb}(\mathrm{d}x)$. An homogeneous PRM in $\mathbb{R}^+$ with Lévy intensity $\tilde{\eta}(A) = \text{Leb}(A)$, can be written as $\sum_{i=1}^{\infty} \delta_{\hat{\Gamma}_i}$. It follows that

$$M = \sum_{i=1}^{\infty} \delta_{(\hat{\Gamma}_i, U_i)}$$

has intensity $\text{Leb} \times \text{Leb}$ in $\mathbb{R}^+ \times [0,1]$. If we define

$$h_1(s) = \inf\left\{ u > 0 : \int_{[u.\infty)} \rho(\mathrm{d}y) < s \right\}$$

then for $a < b$ we observe that $h_1^{-1}([a,b]) = \big(\rho([b,\infty)), \rho([a,\infty))\big)$ has Lebesgue measure $\rho\big([a,\infty)\big) - \rho\big([b,\infty)\big) = \rho\big([a,b)\big)$, so extending the measure it follows from Proposition 1 that

$$N = \sum_{i=1}^{\infty} \delta_{\left(h_1(\hat{\Gamma}_i), U_i\right)}$$

is a PRM with intensity $\nu(\mathrm{d}s, \mathrm{d}x) = \mathrm{Leb} \circ h_1^{-1}(\mathrm{d}s) \times \mathrm{Leb} \circ \mathrm{Id}^{-1}(\mathrm{d}x) = \rho(\mathrm{d}s) \times \mathrm{Leb}(\mathrm{d}x)$. We can use equation (1.6) to construct a CRM with Lévy intensity ν from the series representation of the PRM N above.

In the above example the function h_1 can be seen as the generalized inverse function of the so called *tail integral* of a Lévy intensity $\rho(\mathrm{d}s)\mathrm{d}x$:

Definition 3. Let $\rho(\mathrm{d}s)\mathrm{d}x$ be a Lévy intensity defined in

$$\left(\mathbb{R}^+ \times [0,1], \mathscr{B}(\mathbb{R}^+) \otimes \mathscr{B}([0,1])\right).$$

The tail integral of ρ is the function

$$U(u) = \int_{[u,\infty)} \rho(\mathrm{d}s).$$

We observe that the tail integral, by definition a decreasing function, has a decreasing generalized inverse, denoted h_1 in the above example; so the sequence $\{h_1(\hat{\Gamma}_i)\}_{i=1}^{\infty}$ is monotonically decreasing. For non-homogeneous Lévy intensities we can also use Proposition 1 to get a series representation of the associated CRM, as showed in the next example.

Example. Let $\nu(\mathrm{d}s, \mathrm{d}x)$ be a Lévy intensity in $\mathbb{R}^+ \times \mathbb{R}^+$, $\{W_i\}_{i=1}^{\infty} \subset \mathbb{R}^+$ such that $\sum_{i=1}^{\infty} \delta_{W_i}$ is a PRM with intensity $\hat{\rho}(\mathrm{d}s) = \nu(\mathrm{d}s, \mathbb{R}^+)$ and $\{U_i\}_{i=1}^{\infty}$ i.i.d. Uniform$(0,1)$ random variables; then

$$M = \sum_{i=1}^{\infty} \delta_{(W_i, U_i)}$$

has Lévy intensity $\hat{\rho} \times \mathrm{Leb}$ in $\mathbb{R}^+ \times [0,1]$. If we define

$$F_{X|W}(s) = \frac{\nu(\mathrm{d}W, [0,s])}{\hat{\rho}(\mathrm{d}W)}$$
$$F_{X|W}^{\leftarrow}(s) = \inf\left\{u > 0 : F_{X|W}(u) \geq s\right\}$$

then for $a < b$ we observe that

$$\left(F_{X|W}^{\leftarrow}\right)^{-1}([a,b]) = \left(F(a)_{X|W}, F(b)_{X|W}\right)$$

has Lebesgue measure

$$\frac{\nu(\mathrm{d}W, (a,b])}{\hat{\rho}(\mathrm{d}W)}$$

so extending the measure it follows from Proposition 1 that

$$N = \sum_{i=1}^{\infty} \delta_{\left(W_i, F^{\leftarrow}_{X|W_i}(U_i)\right)}$$

is a PRM with intensity $\hat{\rho} \circ \mathrm{Id}^{-1}(\mathrm{d}s) \times \mathrm{Leb} \circ \left(F^{\leftarrow}_{X|W}\right)^{-1}(\mathrm{d}x) = \hat{\rho}(\mathrm{d}s) \times F_{X|s}(\mathrm{d}x) = \nu(\mathrm{d}s, \mathrm{d}x)$.

The above examples, albeit considering $\mathbb{X}$ to be a general Polish space, can be used to motivate the next algorithm for simulation of a CRM μ with Lévy intensity ν as above.

Algorithm 1 Ferguson-Klass

1: Let $K \in \mathbb{N}$ and simulate

$$U_i \overset{\text{i.i.d.}}{\sim} \text{Uniform}(0,1), \quad i \in \{1,2,\ldots,K\};$$
$$\Gamma_i \overset{\text{i.i.d.}}{\sim} \text{Exponential}(1), \quad i \in \{1,2,\ldots,K\};$$
$$\hat{\Gamma}_i = \sum_{n=1}^{i} \Gamma_i, \quad i \in \{1,2,\ldots,K\}.$$

2: Define

$$W_i = \inf\left\{ u > 0 : \int_{[u,\infty)\times\mathbb{X}} \nu(\mathrm{d}s, \mathrm{d}x) < \hat{\Gamma}_i \right\}$$

and

$$X_i = \inf\left\{ s > 0 : \frac{\nu\left(\mathrm{d}W_i, [0,s]\right)}{\nu\left(\mathrm{d}W_i, \mathbb{X}\right)} \geq U_i \right\}.$$

3: Approximate μ by using

$$\mu \approx \sum_{i=1}^{K} W_i \delta_{X_i}$$

Again we observe that the weights W_i are given by the generalized inverse function of the tail integral of $\rho(\mathrm{d}s) = \nu(\mathrm{d}s, \mathbb{X})$ and as such conform a monotonically decreasing sequence. For the truncation parameter K in step 3 of the Ferguson-Klass algorithm above we exploit the monotonicity of the weights and will use

$$K = \max\{i : W_i \geq b\}$$

for some $b \in \mathbb{R}^+$. The above algorithm was proposed in Ferguson and Klass (1972). A variety of algorithms for simulation of CRM's are available; see for instance Rosiński (2001) and Cont and Tankov (2004) for such algorithms in a Lévy process context.

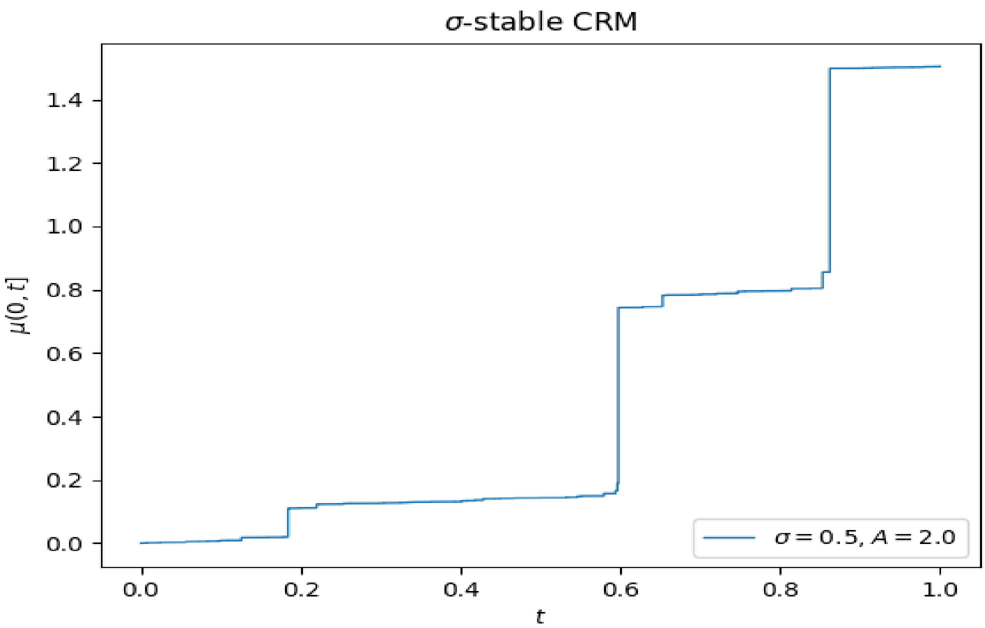


Fig. 1.1 Plot of $\mu(0,t]$ when a σ-stable CRM is considered with $\kappa(\mathrm{d}x) = \mathrm{d}x$. The plot was obtained by using Algorithm 1; truncation level in step 3 of the algorithm is given by the bound $b = 10^{-6}$ as discussed above.

A popular example of an homogeneous CRM is the σ-stable

Example 1. σ-stable CRM

Let $\sigma \in (0,1)$, the Lévy intensity determining a σ-stable CRM is

$$\nu(\mathrm{d}s, \mathrm{d}x) = \frac{A\sigma s^{-1-\sigma}}{\Gamma(1-\sigma)} \mathrm{d}s \kappa(\mathrm{d}x). \tag{1.7}$$

As an illustration, we plot in Figure 1.1 the associated process $\mu(0,t]$ for the σ-stable process in (1.7) with $\kappa(\mathrm{d}x) = \mathrm{d}x$ by making use of Algorithm 1, where

$$U(x) = \frac{A\sigma}{\Gamma(1-\sigma)} \int_x^\infty s^{-1-\sigma} \mathrm{d}s = \frac{A}{\Gamma(1-\sigma) x^\sigma}$$

is the associated tail integral and

$$U^{\leftarrow}(x) = \left(\frac{A}{\Gamma(1-\sigma)x}\right)^{\frac{1}{\sigma}}.$$

is the corresponding generalized inverse used to generate the jump weights W_i in step 2 of Algorithm 1. We choose the truncation level K in step 3 of Algorithm 1 such that the weights $W_i \geq b$ for $i \in \{1,\ldots,K\}$ for $b \in \mathbb{R}^+$.

Another important example of an homogeneous CRM is the Gamma.

Example 2. Gamma CRM
Let $\alpha, \beta \in (0,\infty)$, the Lévy intensity determining a Gamma(α,β) CRM is

$$\nu(\mathrm{d}s,\mathrm{d}x) = \frac{\beta e^{-\alpha s}}{s}\mathrm{d}s\kappa(\mathrm{d}x) \tag{1.8}$$

If for example we take $\kappa(\mathrm{d}x) = \mathrm{d}x$ above then we have the tail integral

$$U(x) = \beta \int_x^\infty \frac{e^{-\alpha s}}{s}\mathrm{d}s$$

Following Wolpert and Ickstadt (1998) we define

$$E(x) = \int_x^\infty \frac{e^{-s}}{s}\mathrm{d}s$$

so $U(x) = \beta E(\alpha x)$, and observe that if χ_d^2 is a Chi-squared distribution with d degrees of freedom then

$$\begin{aligned} E(x) &= \lim_{d\to 0}\Gamma\left(\frac{d}{2}\right)\mathbb{P}\left[\chi_d^2 > 2x\right] \\ &= \left(\lim_{d\to 0}\Gamma\left(\frac{d}{2}+1\right)\right)\left(\lim_{d\to 0}\frac{2}{d}\mathbb{P}\left[\chi_d^2 > 2x\right]\right) \\ &= \lim_{d\to 0}\frac{2}{d}\mathbb{P}\left[\chi_d^2 > 2x\right]. \end{aligned}$$

The generalized inverse of E, denoted $E^{\leftarrow}$, can be written in terms of the quantile function of a χ_d^2-distribution which we denote $Q_{\chi_d^2}$ as follows

$$E^{\leftarrow}(x) = \lim_{d\to 0}\frac{Q_{\chi_d^2}\left(1-\frac{dx}{2}\right)}{2}.$$

So

$$U^{\leftarrow}(x) = \frac{E^{\leftarrow}\left(\frac{x}{\beta}\right)}{\alpha} = \lim_{d\to 0} \frac{Q_{\chi_d^2}\left(1-\frac{dx}{2\beta}\right)}{2\alpha} \tag{1.9}$$

The Laplace exponent of the Gamma CRM is given by

$$\psi(\lambda) = \beta \log(1+\frac{\lambda}{\alpha}), \tag{1.10}$$

see for instance Kyprianou (2006). As another illustration of Algorithm 1, we plot in Figure 1.2 the stochastic process $\mu(0,t]$, with $t \in \mathbb{R}^+$, given by a Gamma(α,β) CRM μ with $\kappa(\mathrm{d}x) = \mathrm{d}x$, so we are in the homogeneous case. To evaluate the generalized inverse of the associated tail integral in step 2 of Algorithm 1 we use the limit approximation in (1.9) by choosing d sufficiently small. We choose the truncation level K in step 3 of Algorithm 1 such that the weights $W_i \geq b$ for $i \in \{1,\dots,K\}$ for $b \in \mathbb{R}^+$.

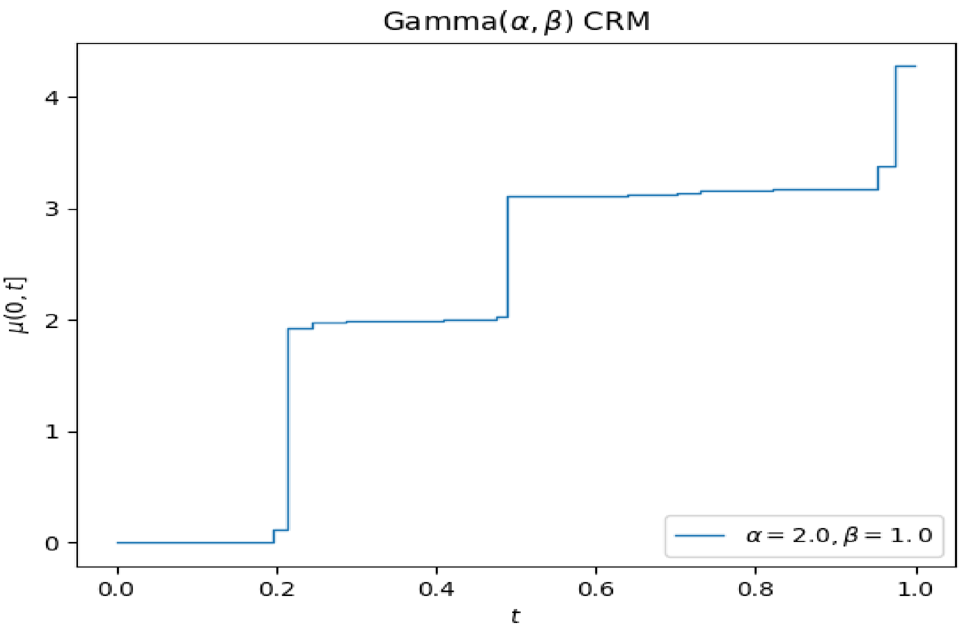


Fig. 1.2 Plot of $\mu(0,t]$ when a *Gamma*$(2,1)$ CRM is considered with $\kappa(\mathrm{d}x) = \mathrm{d}x$. The plot was obtained by using Algorithm 1; in step 2 of the Algorithm the limit approximation (1.9) was used with $d = 10^{-39}$ and in step 3 the truncation level is given by the bound $b = 10^{-19}$ as discussed above.

The notion of a completely random measure can be generalized to higher dimensions in a similar fashion to Definition 1 by considering vectors that have CRM's as entries.

Definition 4. A vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_d)$ of random measures on $(\mathbb{X}, \mathscr{X})$ is called a *vector of completely random measures* (VCRM) if for any $n > 1$ and disjoint sets $A_1, \ldots, A_n \in \mathscr{X}$ the random vectors $\{(\mu_1(A_i), \ldots, \mu_d(A_i))\}_{i=1}^n$ are mutually independent.

In similarity with the CRM case, we will restrict ourselves to VCRM's that have no deterministic part nor fixed jump locations, again we will treat this case separately; so the VCRM's $\boldsymbol{\mu}$ we consider have a series representation

$$\boldsymbol{\mu} = \left(\sum_{i=1}^{\infty} W_{1,i} \delta_{X_i}, \ldots, \sum_{i=1}^{\infty} W_{d,i} \delta_{X_i} \right) \tag{1.11}$$

for a random collection of vectors $\{(W_{1,i}, \ldots, W_{d,i})\}_{i=1}^{\infty}$ taking values in $(\mathbb{R}^+)^d$ and $\{X_i\}_{i=1}^{\infty}$ taking values in $\mathbb{X}$; such VCRM's are determined by the following Laplace functional transform:

$$\mathbb{E}\left[e^{-\mu_1(f_1) - \ldots - \mu_d(f_d)} \right] = e^{-\int_{(\mathbb{R}^+)^d \times \mathbb{X}} \left(1 - e^{-f_1(x)s_1 - \ldots - f_d(x)s_d}\right) \tilde{\nu}_d(\mathrm{d}\boldsymbol{s}, \mathrm{d}x)} \tag{1.12}$$

where $f_j : \mathbb{X} \to \mathbb{R}^+$, $j \in \{1, \ldots, d\}$, are such that $\mu_j(f_j) < \infty$ and $\tilde{\nu}_d$ is a measure, defined in $\left((\mathbb{R}^+)^d \times \mathbb{X}, \mathscr{B}\left((\mathbb{R}^+)^d\right) \otimes \mathscr{X}\right)$, which must satisfy the integrability condition

$$\int_{(\mathbb{R}^+)^d \times \mathbb{X}} \min\{1, \|\boldsymbol{s}\|\} \tilde{\nu}_d(\mathrm{d}\boldsymbol{s}, \mathrm{d}x) < \infty \tag{1.13}$$

for any bounded set $X \in \mathscr{X}$. We call $\tilde{\nu}_d$ the multivariate Lévy intensity of $\boldsymbol{\mu}$. Two basic examples of VCRM's are given next.

Example 3. Independent entries VCRM
Let $d \in \mathbb{N} \setminus \{0\}$ and $\boldsymbol{\mu}$ a $d-$dimensional VCRM. The entries of $\boldsymbol{\mu}$ are pairwise independent if and only if its Lévy intensity satisfies that

$$\begin{aligned} \tilde{\nu}_d(A, B) = {} & \nu_1(\{s : (s, 0, \ldots, 0) \in A\}, B) + \nu_2(\{s : (0, s, 0, \ldots, 0) \in A\}, B) \\ & + \ldots + \nu_d(\{s : (0, 0, \ldots, s) \in A\}, B) \end{aligned}$$

for some univariate Lévy intensities $\nu_1, \ldots, \nu_d$ and any sets $A \in \mathscr{B}((\mathbb{R}^+)^d)$, $B \in \mathscr{B}(\mathbb{R}^+)$.

Example 4. Completely identical dependence
Let $d \in \mathbb{N} \setminus \{0\}$ and $\boldsymbol{\mu}$ a $d-$dimensional VCRM. The entries of $\boldsymbol{\mu}$ are almost surely equal if and only if its Lévy intensity satisfies that

$$\tilde{\nu}_d(A, B) = \tilde{\nu}(\{s : (s, s, \ldots, s) \in A\}, B)$$

for some univariate Lévy intensities $\tilde{\nu}$ and any sets $A \in \mathscr{B}((\mathbb{R}^+)^d)$, $B \in \mathscr{B}(\mathbb{R}^+)$.

The corresponding homogeneous case arises when $\tilde{\nu}_d$ can be written in the form

$$\tilde{\nu}_d(\mathrm{d}\boldsymbol{s}, \mathrm{d}x) = \tilde{\rho}_d(\mathrm{d}\boldsymbol{s})\kappa(\mathrm{d}x) \tag{1.14}$$

where $\tilde{\rho}_d$ is a measure in $\left((\mathbb{R}^+)^d, \mathscr{B}\left((\mathbb{R}^+)^d\right)\right)$ referring to the vector of jump heights and κ is a non-atomic measure in $(\mathbb{X}, \mathscr{X})$ referring to the shared jump locations in each CRM within the vector. Again, if in the multivariate Laplace transform (1.12) we take $f_i(x) = \lambda_i \mathbb{1}_{\{x \in A\}}$ for $i \in \{1, \ldots, d\}$, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_d) \in (\mathbb{R}^+)^d$ and A a bounded set in $\mathscr{X}$ then we can obtain a simpler version of the multivariate Laplace transform as

$$\mathbb{E}\left[e^{-\lambda_1 \mu_1(A) - \cdots - \lambda_d \mu_d(A)}\right] = e^{-\int_{(\mathbb{R}^+)^d \times A}(1 - e^{-\lambda_1 s_1 - \cdots - \lambda_d s_d})\tilde{\nu}_d(\mathrm{d}\boldsymbol{s}, \mathrm{d}x)}. \tag{1.15}$$

We define the Laplace exponent corresponding to a VCRM as

$$\psi_t(\boldsymbol{\lambda}) = \int_{(\mathbb{R}^+)^d \times (0,t]} (1 - e^{-\langle \boldsymbol{\lambda}, \boldsymbol{s} \rangle})\tilde{\nu}_d(\mathrm{d}\boldsymbol{s}, \mathrm{d}x). \tag{1.16}$$

Again, if the Lév intensity of interest is homogeneous as in (1.14) then the corresponding Laplace exponent can be written as $\psi_t(\boldsymbol{\lambda}) = \gamma(t)\psi(\boldsymbol{\lambda})$ where $\gamma(t) = \int_0^t \kappa(\mathrm{d}x)$ and $\psi(\boldsymbol{\lambda}) = \int_{(\mathbb{R}^+)^d}(1 - e^{-\langle \boldsymbol{\lambda}, \boldsymbol{s} \rangle})\tilde{\rho}_d(\mathrm{d}\boldsymbol{s})$. We observe that each entry in a VCRM $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_d)$ is a CRM with its corresponding Lévy intensity. Indeed, for $j \in \{1, \ldots, d\}$, the Lévy intensity associated to μ_j, the j-th entry of a VCRM $\boldsymbol{\mu}$ with multivariate Lévy intensity $\tilde{\nu}_d$, is given by

$$\nu_j(A, X) = \int_{(\mathbb{R}^+)^{d-1}} \tilde{\nu}_d(\mathrm{d}s_1, \ldots, \mathrm{d}s_{j-1}, A, \mathrm{d}s_{j+1}, \ldots, \mathrm{d}s_d, X) \tag{1.17}$$

with $A \in \mathscr{B}(\mathbb{R}^+)$. We call ν_j the j-th marginal of the d-variate Lévy intensity $\tilde{\nu}_d$. We observe that if $\tilde{\nu}_d$ is homogeneous, taking the form (1.14), then each marginal ν_j associated to $\tilde{\nu}_d$ can be written in the form

$$\nu_j(\mathrm{d}s, \mathrm{d}x) = \rho_j(\mathrm{d}s)\kappa(\mathrm{d}x) \tag{1.18}$$

with ρ_j a measure in $(\mathbb{R}^+, \mathscr{B}(\mathbb{R}^+))$, for each $j \in \{1, \ldots, d\}$. For a Lévy intensity $\rho_d(\mathrm{d}\boldsymbol{s})\mathrm{d}x$ we define the $d-$variate tail integral in analogy with Definition 3:

Definition 5. Let $\rho_d(\mathrm{d}\boldsymbol{s})\mathrm{d}x$ be a $d-$variate Lévy intensity defined in

$$\left((\mathbb{R}^+)^d \times \mathbb{R}^+, \mathscr{B}\left((\mathbb{R}^+)^d\right) \otimes \mathscr{B}(\mathbb{R}^+)\right).$$

The $d-$variate tail integral associated to ρ_d is the function

$$U(\boldsymbol{u}) = \int_{[u_1,\infty)\times\ldots\times[u_d,\infty)} \tilde{\rho}_d(\mathrm{d}\boldsymbol{s}).$$

With the notation as in 1.18 we set the notation for the marginal tail integral, i.e. the tail integral associated to the marginal $\rho_i(\mathrm{d}s)\mathrm{d}x$, as follows

$$U_j(u) = \int_{[u,\infty)} \rho_j(\mathrm{d}s) \tag{1.19}$$

for $j \in \{1,\ldots,d\}$.

1.3 Neutral to the right distributions

In the seminal work of Doksum (1974), the *neutral to the right* (NTR) probability distributions were introduced. Such NTR distributions can be expressed in terms of a CRM μ.

Definition 6. 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

$$\lim_{t\to\infty} \mu(0,t] \stackrel{\text{a.s.}}{=} \infty, \tag{1.20}$$

so Y is almost surely supported in $\mathbb{R}^+$.

NTR distributions have several appealing properties, including the independence of normalized increments

$$F(t_1), \frac{F(t_2)-F(t_1)}{1-F(t_1)}, \cdots, \frac{F(t_K)-F(t_{K-1})}{1-F(t_{K-1})}$$

for $F(t) = 1 - S(t)$ the distribution function associated to the NTR distribution, $0 < t_1 < \ldots < t_K$ for $K \in \mathbb{N}$. Another important property is the posterior characterization for *censored to the right data* which is of great utility in a survival analysis context.

Definition 7. Let there be two independent samples

$$\{Y_i\}_{i=1}^{\infty} \stackrel{\text{i.i.d.}}{\sim} H$$

and

$$\{C_i\}_{i=1}^{\infty} \overset{\text{i.i.d.}}{\sim} G,$$

for H, G cumulative distribution functions. We suppose that we only observe

$$T_i = \min\{Y_i, C_i\}; \quad J_i = \mathbb{1}_{\{Y_i \leq C_i\}}.$$

We denote as censored to the right survival data

$$\mathscr{D} = \{(T_i, J_i)\}_{i=1}^{n}$$

.

Taking into account possible repetitions in the censored to the right observations, we consider the ordered statistics without repetitions, $(T_{(1)}, \ldots, T_{(k)})$, where k is the number of different observations. We set $T_{(0)} = 0$, $T_{(k+1)} = \infty$. In the frequentist statistics literature, the survival function of the observations of interest, in censored to the right data, is fitted with the *Kaplan-Meier* estimator, we will use this frequentist estimation for comparison with the Bayesian nonparametric fits of survival functions in this thesis.

Definition 8. The Kaplan-Meier estimator for the survival function of the events of interest in a censored to the right survival data setting is given by

$$\hat{S}_{\text{KM}}(t) = \prod_{\{j : T_{(j)} \leq t\}} \left(1 - \frac{\#\{i : T_i = T_{(j)}, J_i = 1\}}{\#\{i : T_i \geq T_{(j)}\}}\right) \tag{1.21}$$

The number of exact observations for censored to the right data is $n_e = \sum_{i=1}^{n} J_i$ and the number of censored observations is $n_c = n - n_e$. Define the set functions

$$m^e(A) = \sum_{i=1}^{n} J_i \mathbb{1}_A(T_i) \quad ; \quad m^c(A) = \sum_{i=1}^{n} (1 - J_i \mathbb{1}_A(T_i)).$$

So we define the numbers

$$n_j^e = \#\left\{i : T_i = T_{(j)} \text{ and } J_i = 1\right\} \qquad ; \qquad n_j^c = \#\left\{i : T_i = T_{(j)} \text{ and } J_i = 0\right\}$$

for, respectively, the exact observations and censored observations related to time $T_{(j)}$, $j \in \{1,\dots,k\}$. The next cumulative quantities will also be of use

$$\bar{n}_j^e = \sum_{r=j}^{k} n_r^e \quad ; \quad \bar{n}_j^c = \sum_{r=j}^{k} n_r^c,$$

where $j \in \{1,\dots,k\}$. Using the notation introduced above, the posterior characterization when considering censored to the right data for a NTR distribution is given next.

Theorem 1. *Let μ be a CRM with Lévy intensity of the form $\nu(\mathrm{d}s,\mathrm{d}x) = \nu(s,\mathrm{d}x)\mathrm{d}s$. And let $\mathscr{D}$ be censored to the right data arising from a NTR(μ) model. If for $\eta_t(s) = \nu(s,(0,t])$ the partial derivative $\eta'_{t_0}(s) = \frac{\partial \eta_t(s)}{\partial t}\Big|_{t=t_0}$ exists, then the posterior distribution given censored to the right data $\mathscr{D}$ is again NTR with the next associated CRM*

$$\mu^\circ + \sum_{\{j:T_{(j)} \text{ is an exact observation }\}} M_j \delta_{T_{(j)}} \tag{1.22}$$

where

i) μ° is given by a Lévy intensity ν° such that

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

for $j \in \{1,\dots,k+1\}$.

ii) The random weights $\{M_j\}_{j\in M}$, with

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

are mutually independent and have, respectively, a density given by

$$f_j(s) \propto e^{-(\bar{n}_j^c+\bar{n}_{j+1}^e)s}(1-e^{-s})^{n_j^e}\eta'_{T_{(j)}}(s).$$

iii) The completely random measure μ° is independent of $\{M_j\}_{j\in M}$.

The result above showcases that the posterior distribution of a NTR distribution with respect to censored to the right data is itself NTR again. This is because the measure in (1.22) is completely random; we can distinguish two parts in it, a so called "continuous" part given by μ° and a discrete part with jump locations explicitly given by the exact observations in the survival data and random jump weights. We observe that such result is similar in spirit to the conjugacy results in parametric Bayesian statistics and allows for model tractability. Indeed, we can use the result above to propose an estimator for the survival function of the event times of interest in terms of a posterior mean as follows.

Corollary 1. *In the setting of Theorem 1, let $\bar{H}$ be the survival function associated to the event times of interest, $\{Y_i\}_{i=1}^n$ in the censored to the right data $\mathscr{D}$ and set*

$$M_t = \left\{ j : T_{(j)} \text{ is an exact observation and } T_{(j)} \leq t \right\}.$$

We can estimate $\bar{H}$ with

$$\hat{S}(t) = \mathbb{E}[\mathbb{P}[Y > t \,|\mu] \,|\mathscr{D}] = e^{-\sum_{j=1}^{k+1}\left(\psi^{\circ j}_{t\wedge T_{(j)}}(1) - \psi^{\circ j}_{t\wedge T_{(j-1)}}(1)\right)} \times \prod_{j\in M_t} \frac{\int_0^\infty e^{-(1+\bar{n}^c_j+\bar{n}^e_{j+1})s}(1-e^{-s})^{n^e_j}\eta'_{T_{(j)}}(s)\mathrm{d}s}{\int_0^\infty e^{-(\bar{n}^c_j+\bar{n}^e_{j+1})s}(1-e^{-s})^{n^e_j}\eta'_{T_{(j)}}(s)\mathrm{d}s} \tag{1.23}$$

where $\psi^{\circ j}_t$ is the Laplace exponent of μ° restricted to $(T_{(j-1)}, T_{(j)})$, $j \in \{1, \ldots, k\}$.

Usually the Lévy intensity underlying the CRM in a NTR distribution is parametrized by some real valued vector $\boldsymbol{c}$. It follows from the proof of Theorem 1 that we can get an explicit formula for the likelihood of such vector of hyper-parameters $\boldsymbol{c}$, given survival data $\mathscr{D}$ as before; such likelihood is of key importance for the inferential procedure of the NTR model.

Corollary 2. *In the setting of Theorem 1, given survival data $\mathscr{D}$ if the underlying Lévy intensity $\nu_{\boldsymbol{c}}$, the corresponding partial derivative $\eta'_{t,\boldsymbol{c}}$ and Laplace exponent $\psi_{t,\boldsymbol{c}}$ are parametrized by some vector $\boldsymbol{c}$ with real valued entries then the likelihood on $\boldsymbol{c}$ is given by*

$$l(\boldsymbol{c};\mathscr{D}) = e^{-\sum_{j=1}^{k}\left(\psi_{T_{(j)},\boldsymbol{c}}(\bar{n}^c_j+\bar{n}^e_j) - \psi_{T_{(j-1)},\boldsymbol{c}}(\bar{n}^c_j+\bar{n}^e_j)\right)} \times \prod_{j\in M} \int_0^\infty \left(e^{-(\bar{n}^c_j+\bar{n}^e_{j+1})s}(1-e^{-s})^{n^e_j}\right)\eta'_{T_{(j)},\boldsymbol{c}}(s)\mathrm{d}s.$$

We highlight that if we assign a prior distribution on the vector of hyper-parameters $\boldsymbol{c}$ then we can use the above likelihood to get the posterior distribution; furthermore we can use a Markov Chain Monte-Carlo (MCMC) algorithm to draw samples from $\boldsymbol{c}|\mathscr{D}$. To make easier the evaluation of the estimator in Corollary 1 we present two Propositions next.

Proposition 2. *In the setting of Theorem 1, the Laplace exponent of μ° restricted to $(T_{(j-1)}, T_{(j)})$, $j \in \{1,\dots,k\}$, can be evaluated as*

$$\psi_t^{\circ j}(\lambda) = \left(\psi_{t\wedge T_{(j)}}(\lambda + \bar{n}_j^c + \bar{n}_j^e) - \psi_{t\wedge T_{(j-1)}}(\lambda + \bar{n}_j^c + \bar{n}_j^e)\right) \\ - \left(\psi_{t\wedge T_{(j)}}(\bar{n}_j^c + \bar{n}_j^e) - \psi_{t\wedge T_{(j-1)}}(\bar{n}_j^c + \bar{n}_j^e)\right)$$

We observe that if in the above result the Laplace exponent ψ_t is related to an homogeneous CRM so $\psi_t(\lambda) = \gamma(t)\psi(\lambda)$ then we have the following simplification

$$\psi_t^{\circ j}(\lambda) = \left(\gamma(t\wedge T_{(j)}) - \gamma(t\wedge T_{(j-1)})\right)\left(\psi(\lambda + \bar{n}_j^c + \bar{n}_j^e) - \psi(\bar{n}_j^c + \bar{n}_j^e)\right)$$

Proposition 3. *Let ν be a Lévy measure defining a CRM, let $q \in \mathbb{R}^+$ and $n \in \mathbb{N}\setminus\{0\}$. Then*

$$\int_{\mathbb{R}^+\times(0,t]} e^{-qs}\left(1-e^{-s}\right)^n \nu(\mathrm{d}s,\mathrm{d}x) = \sum_{k=0}^{n-1}\binom{n-1}{k}(-1)^{k+1}\left(\psi_t(k+q) - \psi_t(k+1+q)\right)$$

We observe that in order to calculate the estimator proposed in Corollary 1 we need to explicitly evaluate the Laplace exponent ψ of the underlying CRM in the NTR distribution, this need is further showed in Proposition 3 above where the Laplace exponent is used to evaluate integrals as the ones in the estimator formula 1.23. Similarly to evaluate the likelihood in Corollary 2 the calculation of the Laplace exponent can be of key importance for computations. In the next example we illustrate two ways of of fitting the survival of possibly censored to the right data in a NTR context.

Example 5. NTR fit

We consider synthetic exponentially distributed censored to the right survival data $\mathscr{D}$ where

$$\begin{aligned} Y_i &\sim \text{Exponential}(1.0), && i = 1,\dots,125. \\ C_i &\sim \text{Exponential}(1.0/3.0), && j = 1,\dots,125. \end{aligned}$$

Such choice is useful as it allows for around 80% of the censored to the right observations to be exact, around 100 data points. We use the estimator from Corollary 1 to fit the data. As we take into account a vector of hyper-parameters $\boldsymbol{c}$ in what follows we denote such estimator, (1.23), as $\hat{S}(t;\boldsymbol{c})$. We consider a NTR distribution with an homogeneous Gamma(α,β) as underlying CRM. Hence, our vector of hyper-parameters is $\boldsymbol{c} = (\alpha,\beta)$; to which we assign prior distributions, p_α and p_β respectively, and use a Metropolis-Within Gibbs algorithm to draw samples from the posterior distributions of α and β, i.e. $\alpha|\mathscr{D}$ and $\beta|\mathscr{D}$ accordingly. To this end we use the likelihood $l(\alpha,\beta;\mathscr{D})$ as presented in 2. Given initial values $\alpha^{(0)}$ and $\beta^{(0)}$ the algorithm is as follows.

Algorithm 2 Metropolis within Gibbs for NTR(Gamma (α,β)) model fit example

1: Draw $\alpha^{(i+1)}$ from a Metropolis-Hastings sampler with truncated Normal proposal distribution $g(x'|x) \sim \text{Normal}\big|_{(0,\infty)}(x,1)$ and target distribution

$$l(x,\beta^{(i)};\mathscr{D})p_\alpha(x).$$

2: Draw $\beta^{(i+1)}$ from a Metropolis-Hastings sampler truncated Normal proposal distribution $g(x'|x) \sim \text{Normal}\big|_{(0,\infty)}(x,1)$ and target distribution

$$l(\alpha^{(i+1)},x;\mathscr{D})p_\beta(x).$$

After using the above algorithm to generate a MCMC chain we use the values α^{maxpost} and β^{maxpost} which attain the maximum posterior distribution along the chain to evaluate the estimator in Corollary 1 as

$$\hat{S}_{\text{maxpost}} = \hat{S}(t;\alpha^{\text{maxpost}},\beta^{\text{maxpost}}) \tag{1.24}$$

see Figure . Alternatively we can average the estimator over the values in the chain after a burn-in index.

$$\hat{S}_{\text{averaged}}(t) = \sum_{i=\text{burn-in}}^{\text{chain-length}} \frac{\hat{S}(t;\alpha^{(i)}\beta^{(i)})}{\text{chain-length} - \text{burn-in}}, \tag{1.25}$$

see Figure . The prior distributions we choose for the hyper-parameters are.

$$\alpha \sim \text{LogNormal}(m = 1.0/\bar{T}, \sigma^2 = 1.0)$$
$$\beta \sim \text{LogNormal}(m = 1.0, \sigma^2 = 1.0)$$

where $\bar{T}$ is the mean of the censored to the right observations $\{T_i\}_{i=1}^{125}$ and LogNormal(m,σ^2) denotes the law of LogNormal random variable associated to a Normal distribution with mean m and variance σ^2. With such choices we run Algorithm 2 and show the chains for α and β in Figures 1.3 and 1.4, respectively; the values attaining the maximum a posteriori value are showed in these figures and later used to plug-in in the estimator (1.24) which is showed in Figure 5; even more, we use the chains to produce the estimator (1.25) as showcased in Figure 1.6.

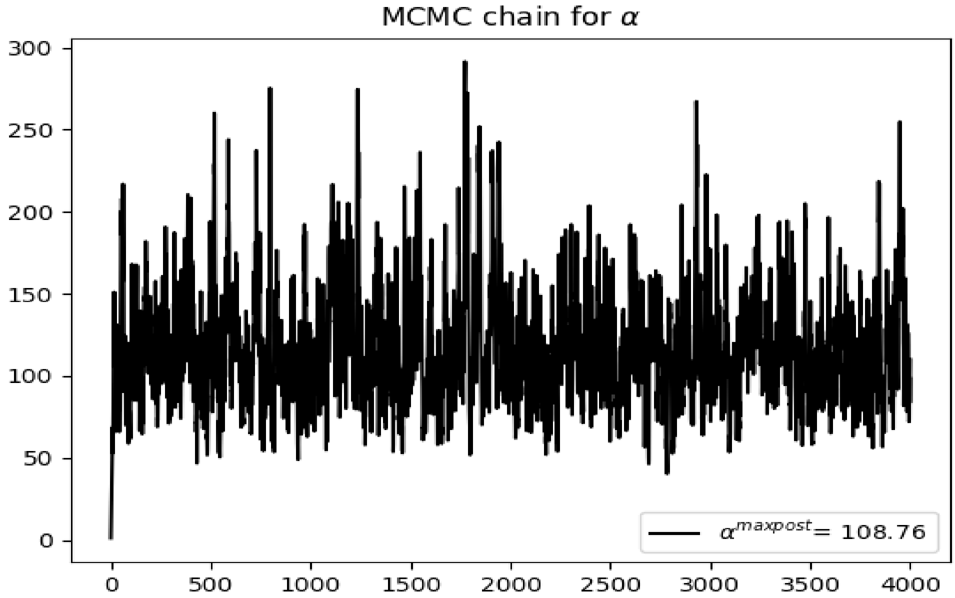


Fig. 1.3 Plot of MCMC chain for α as described in Algorithm 2.

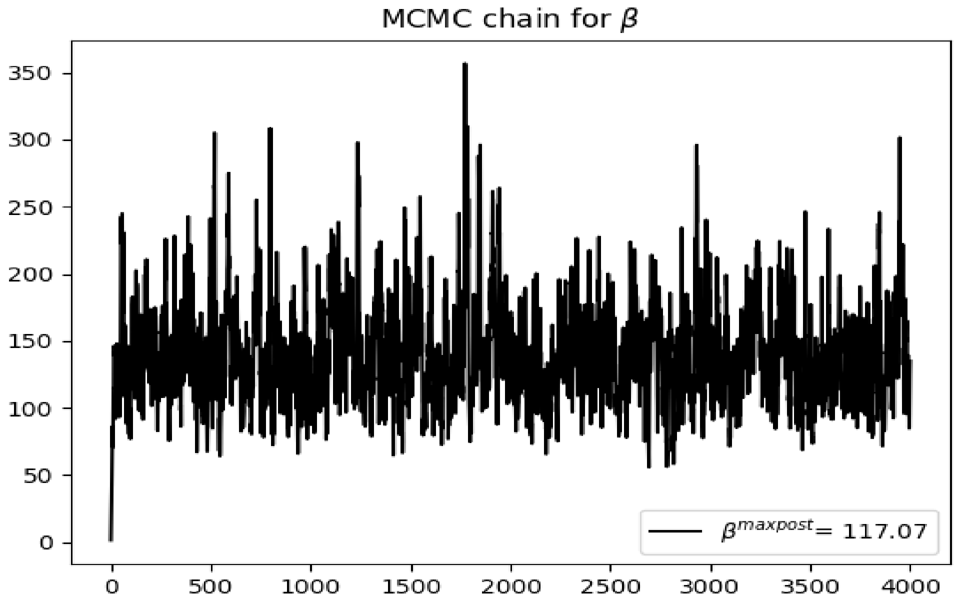


Fig. 1.4 Plot of MCMC chain for β as described in Algorithm 2.

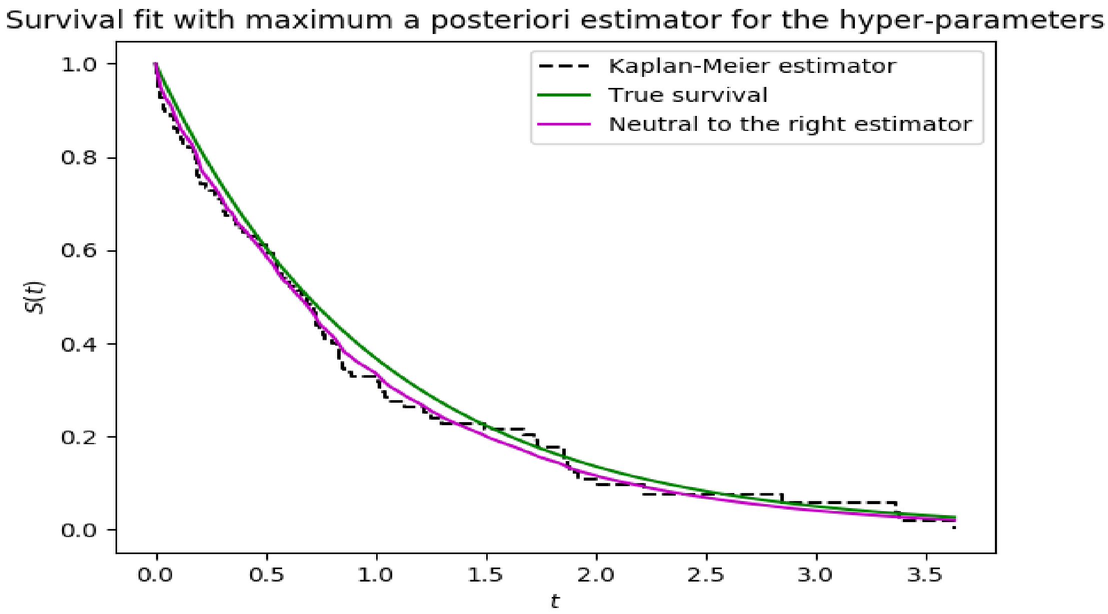


Fig. 1.5 NTR survival fit with the estimator (1.24), compared with the Kaplan-Meier estimator (1.21) and the true survival function of the events of interest, which are exponentially distributed with rate one.

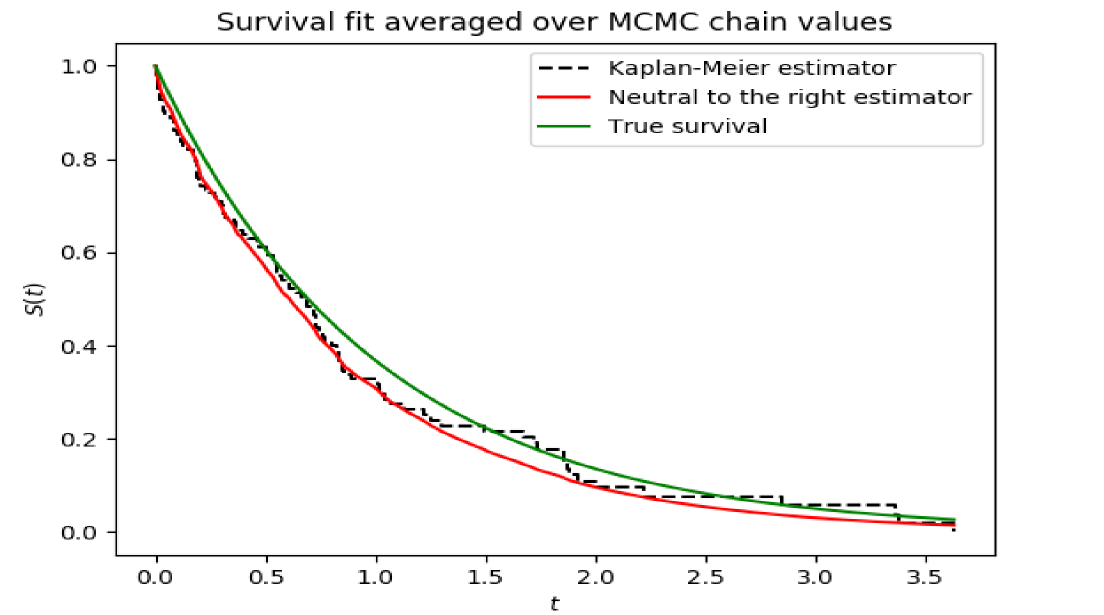


Fig. 1.6 NTR survival fit with the estimator (1.25), burn-in= 200, compared with the Kaplan-Meier estimator (1.21) and the true survival function of the events of interest, which are exponentially distributed with rate one.

The asymptotic analysis of a posterior distribution when the number of observations n tends to infinity is of special importance in Bayesian statistics. *Posterior consistency* is a key concept in this setting.

Definition 9. Given a time horizon $\tau \in \mathbb{R}^+$, n i.i.d. survival times $Y_1,\dots,Y_n$ associated to a true cumulative hazard function Λ_0 which is continuous, a true survival function S_0 such that $\lim_{t\to\infty} S_0(t) = 0$ and $\lim_{t\to 0} S_0(t) = 1$ and corresponding survival data $\mathscr{D}^{(n)} = \{(T_i, \delta_i, \boldsymbol{X}_i)\}_{i=1}^n$, We say that the posterior is consistent if for any $\varepsilon > 0$

$$\mathbb{P}\left[\sup_{t\leq\tau} |S(t) - S_0(t)| < \varepsilon \,|\mathscr{D}^{(n)}\right] \to 1$$

The study of posterior consistency conditions for NTR distributions was performed by Kim and Lee (2001) and Dey et al. (2003). To introduce such conditions we haver first to present the cumulative hazard function for NTR distributions.

Definition 10. Given a random variable with cumulative distribution function F and survival function S, its cumulative hazard functions is given by

$$\Lambda(t) = \int_0^t \frac{F(\mathrm{d}s)}{S(s^-)}.$$

For NTR distributions we have the following characterization of the cumulative hazard.

Proposition 4. *Let $Q^{Exp(1)}(x) = 1 - e^{-x}$ be the quantile functions of an Exponential r.v. with rate parameter 1. If $S \sim NTR(\mu)$ and μ is a CRM with homogeneous Lévy intensity $\nu(\mathrm{d}s, \mathrm{d}x)$ then the cumulative hazard function of S is a CRM with homogeneous Lévy intensity $\xi(\mathrm{d}s, \mathrm{d}x)$ given by*

$$\xi(A,B) = \nu\left(\left(Q^{Exp(1)}\right)^{-1}(A), B\right)$$

for any $A, B \in \mathscr{B}(\mathbb{R}^+)$.

For a proof see Proposition 2 in Dey et al. (2003) and the discussion therein. If ν is homogeneous $\nu(\mathrm{d}s, \mathrm{d}x) = \rho(\mathrm{d}s)\kappa(\mathrm{d}x)$ and furthermore ρ is absolutely continuous with respect to Lebesgue measure, i.e. $\rho(\mathrm{d}s) = \rho(s)\mathrm{d}s$, then $\xi(\mathrm{d}s, dx) = L(\mathrm{d}s)\kappa(\mathrm{d}x)$ and L is absolutely continuous with respect to Lebesgue measure, $L(\mathrm{d}s) = L(s)\mathrm{d}s$, as

$$L(A) = \int_{\left(Q^{\mathrm{Exp}(1)}\right)^{-1}(A)} \rho(s)\mathrm{d}s = \int_A \frac{\rho\left(-\log(1-s)\right)}{1-s}\mathrm{d}s;$$

so $L(s) = \rho\left(-\log(1-s)\right)/(1-s)$ where we have been performing the usual abuse of notation of giving the same name to the measure and the respective Radon-Nikodym derivative.

The next result from Kim and Lee (2001) gives sufficient conditions for posterior consistency when we consider a cumulative hazard rate given by a Lévy process with homogeneous intensity of the form $\xi(s,x)\mathrm{d}s\mathrm{d}x$ in the NTR model.

Proposition 5 (Kim and Lee (2001))**.** *Given a time horizon $\tau \in \mathbb{R}^+$, if S is a survival function such that $S(\mathrm{d}s) = 1 - \Lambda(\mathrm{d}s)$ where Λ is a CRM given by a Lévy intensity ξ such that for*

$$\lambda(x) = \int_0^1 s\xi(s,x)\mathrm{d}s$$

we have that

$$\sup_{x\in[0,\tau],s\in[0,1]} \frac{s(1-s)\xi(s,x)}{\lambda(x)} < \infty$$

and there exists a function $h(x)$ in $[0,\tau]$ such that $0 < \inf_{x\in[0,\tau]} h(x) \leq \sup_{x\in[0,\tau]} h(x) < \infty$ and

$$0 < \lim_{s\to 0} \sup_{x\in[0,\tau]} \left| \frac{s\xi(s,x)}{\lambda(x)} - h(x) \right| < \infty$$

then the posterior distributions of S is consistent.

Example 6 (Kim and Lee (2001))**. Homogeneous Gamma CRM posterior consistency** For the homogeneous gamma CRM we have that

$$\nu(s,x) = \frac{\beta e^{-\alpha s}\kappa(x)}{s}$$

so by Proposition 4 the associated cumulative hazard function Λ has Lévy intensity given by

$$\xi(s,x) = \frac{\beta(1-s)^{\alpha}\kappa(x)}{-\log(1-s)(1-s)}.$$

We observe that

$$\lambda(x) = \int_0^1 \frac{\beta s(1-s)^{\alpha}\kappa(x)}{-\log(1-s)(1-s)}\mathrm{d}s = c\kappa(x)$$

for some $c \in (0,\infty)$. Using $1 - e^{-s} \leq s$

$$\sup_{x\in[0,\tau],s\in[0,1]} \frac{s\beta(1-s)^{\alpha}}{-c\log(1-s)} \leq \sup_{s\in[0,1]} \frac{\beta}{c}(1-s)^{\alpha} = \frac{\beta}{c}.$$

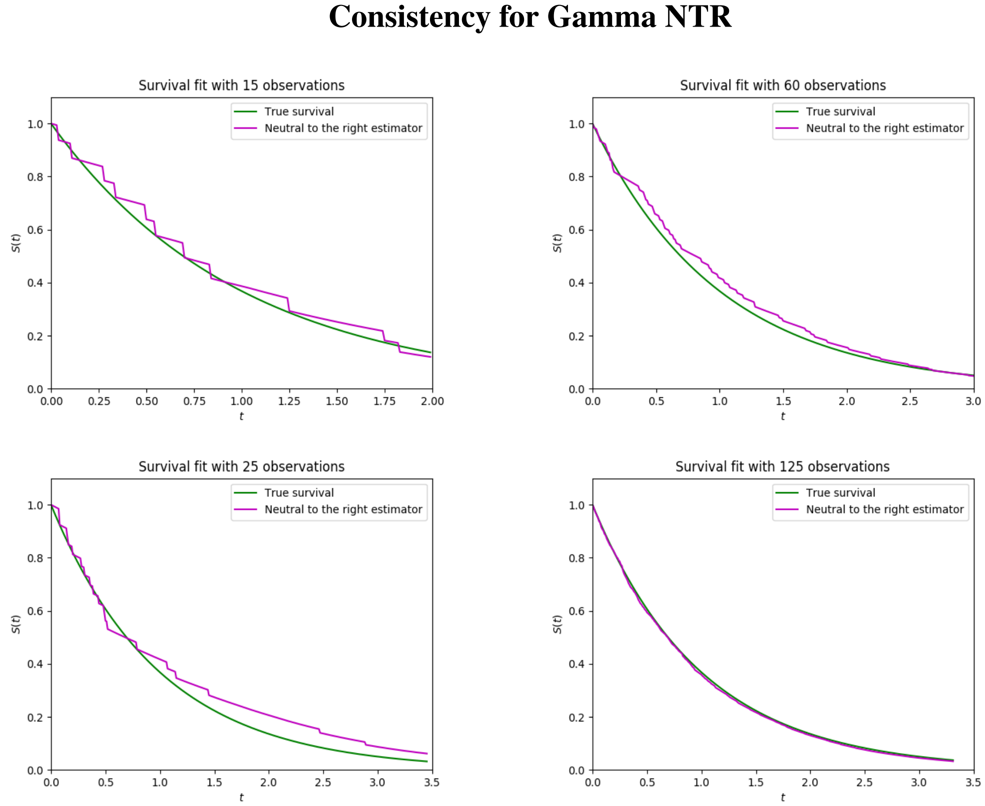


Fig. 1.7 Plot of NTR fits as in Example 5, with (1.24), and true survival for 15, 25, 60 and 125 observations; in each case there are around 80% of the observations are exact and the rest censored to the right.

and choosing $h(x) = \beta/2c$

$$\lim_{s\to 0}\sup_{x\in[0,\tau]}\left|\frac{s\beta(1-s)^{\alpha}}{-c\log(1-s)(1-s)} - \frac{\beta}{2c}\right| = \left|\lim_{s\to 0}\frac{\beta}{c}(1-s) - \frac{\beta}{2c}\right| = \frac{\beta}{2c}$$

So the gamma CRM satisfies the conditions of Proposition 5 and in consequence the consistency of the posterior survival function.

Another interesting asymptotic property in Bayesian statistics is the Bernstein-von Mises theorem which deals with the convergence of the posterior distribution into the law of a frequentist estimator, in our case the Kaplan-Meier estimator of Definition 8. In what follows let $B(\cdot)$ be a standard Brownian motion in $\mathbb{R}^+$ and given survival data with covariates $\mathscr{D}^{(n)}_{\hat{\boldsymbol{X}}}$ denote the associated Kaplan-Meier estimators as $\hat{S}^{(n)}_{\mathrm{KM}}$.

Definition 11. Given a time horizon $\tau \in \mathbb{R}^+$, let $D([0,\tau])$ be the space of cadlag functions on $[0,\tau]$ with the uniform convergence topology and associated Borel σ-algebra. Let there be n i.i.d. survival times $Y_1,\dots,Y_n$ associated to a true cumulative hazard function Λ_0 which is continuous, a true survival function S_0 such that $\lim_{t\to\infty} S_0(t)=0$ and $\lim_{t\to 0} S_0(t)=1$, and corresponding survival data $\mathscr{D}^{(n)}=\{(T_i,\delta_i,\boldsymbol{X}_i)\}_{i=1}^n$; we say that the posterior attains the Bernstein-von Mises theorem if

$$\mathscr{L}\left(\sqrt{n}(S(\cdot)-\hat{S}_{\text{KM}}^{(n)}(\cdot))\,|\,\mathscr{D}_{\hat{\boldsymbol{X}}}^{(n)})\right)\xrightarrow{d}-S_0(\cdot)\mathrm{B}(U_0(\cdot))$$

on $D([0,\tau))$ with probability 1, where $U_0(t)=\int_0^t \mathrm{d}\Lambda_0(s)/Q(s)$ with $Q(t)=\mathbb{P}[Y\geq t]$, $Y\sim\mathscr{L}(S_0)$.

The next result from Kim and Lee (2004) gives sufficient conditions for the Bernstein-von Mises theorem in a NTR setting.

Proposition 6. *(Kim and Lee (2004)) Given a time horizon $\tau\in\mathbb{R}^+$, if S is a survival function such that $S(\mathrm{d}s)=1-\Lambda(\mathrm{d}s)$ where Λ is a CRM given by a Lévy intensity ξ such that for*

$$0<\lambda(x)=\int_0^1 s\xi(s,x)\mathrm{d}s<\infty$$

we have that

$$\sup_{x\in[0,\tau],s\in[0,1]}\frac{s(1-s)\xi(s,x)}{\lambda(x)}<\infty$$

and

$$\sup_{x\in[0,\tau],s\in(0,\varepsilon)}\left|\frac{\partial}{\partial s}\left(\frac{s\xi(s,x)}{\lambda(x)}\right)\right|<\infty$$

for some $\varepsilon>0$. Then the posterior attains the Bernstein-von Mises theorem.

Example 7. (Kim and Lee (2004)) **Homogeneous Gamma CRM Bernstein-von Mises**
In view of Example 6, κ must be bounded and positive in $[0,\tau]$ and we only have to observe that for $g(s)=s\xi(s,x)/\lambda(x)$

$$\begin{aligned} g'(s) &= -\frac{\beta(1-s)^{\alpha-1}}{c\log(1-s)}+\frac{\beta(\alpha-1)s(1-s)^{\alpha-2}}{c\log(1-s)}-\frac{\beta s(1-s)^{\alpha-2}}{\lambda(\log(1-s))^2}\\ &= \frac{\beta(\alpha-1)s(1-s)^{\alpha-2}}{c\log(1-s)}+\frac{\beta(1-s)^{\alpha-1}}{c}\left(\frac{-\log(1-s)-s(1-s)^{-1}}{(\log(1-s))^2}\right)\end{aligned}$$

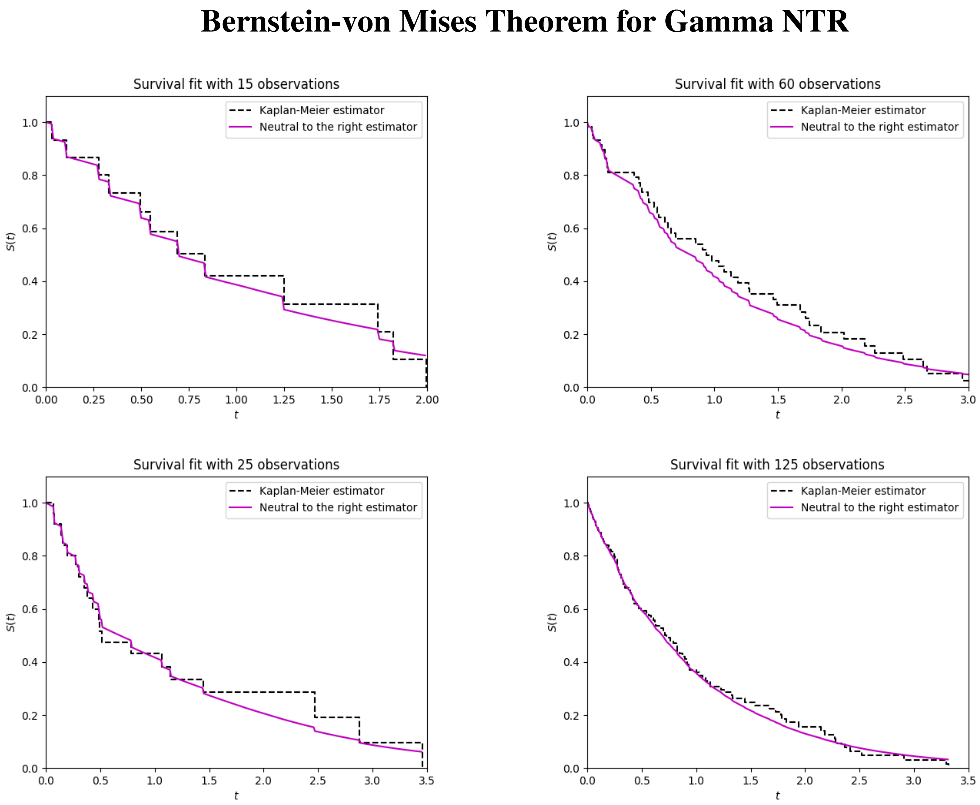


Fig. 1.8 Plot of NTR fits as in Example 5, with (1.24), and Kapla-Meier fits for 15, 25, 60 and 125 observations; in each case around 80% of the observations are exact and the rest censored to the right.

for some $c \in (0,\infty)$. Using L'Hopital's rule we see that

$$\begin{aligned}\lim_{s\to 0} g'(s) &= -\frac{\beta(\alpha-1)}{c} + \frac{\beta}{c}\lim_{s\to 0}\frac{(1-s)^{-1}-(1-s)^{-1}-s(1-s)^{-2}}{-2\log(1-s)(1-s)^{-1}} \\ &= -\frac{\beta(\alpha-1)}{c} - \frac{\beta}{2c}.\end{aligned}$$

So the Bernstein-von Mises result follows.

1.4 Proofs of NTR results

In this section we give the proofs of the results associated to NTR distributions in the previous section.

Proof of Theorem 1

We need the next technical lemma for the proof of the theorem.

Lemma 1. *In the setting of Theorem 1, let $r \in \mathbb{R}^+$, $q \in \mathbb{N} \setminus \{0\}$, $t_0 > 0$ and $0 < \varepsilon < t_0$; then as $\varepsilon \to 0$*

$$\mathbb{E}\left[e^{-r\mu(t_0-\varepsilon,t_0]}(1-e^{\mu(t_0-\varepsilon,t_0]})^q\right] = \varepsilon \int_0^\infty e^{-rs}(1-e^{-s})^q \eta'_{t_0}(s)\mathrm{d}s + o(\varepsilon)$$

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

$$\begin{aligned}
&\mathbb{E}\left[e^{-r\mu(t_0-\varepsilon,t_0]}(1-e^{\mu(t_0-\varepsilon,t_0]})^q\right] = \mathbb{E}\left[\sum_{i=0}^{q}\binom{q}{i}(-1)^i e^{-(r+i)\mu(t_0-\varepsilon,t_0]}\right] \\
&= \sum_{i=0}^{q}\binom{q}{i}(-1)^i e^{-\triangle_{t_0-\varepsilon}^{t_0}\psi_t(r+i)} \\
&= e^{-\triangle_{t_0-\varepsilon}^{t_0}\psi_t(r)}\sum_{i=0}^{q}\binom{q}{i}(-1)^i e^{-\triangle_{t_0-\varepsilon}^{t_0}(\psi_t(r+i)-\psi_t(r))} \\
&= e^{-\triangle_{t_0-\varepsilon}^{t_0}\psi_t(r)}\sum_{i=0}^{q}\binom{q}{i}(-1)^i e^{-\int_0^\infty e^{-rs}(1-e^{-is})\triangle_{t_0-\varepsilon}^{t_0}\eta_t(s)\mathrm{d}s} \\
&= e^{-\triangle_{t_0-\varepsilon}^{t_0}\psi_t(r)}\left(1+\sum_{i=1}^{q}\binom{q}{i}(-1)^i e^{-\int_0^\infty e^{-rs}(1-e^{-is})\triangle_{t_0-\varepsilon}^{t_0}\eta_t(s)\mathrm{d}s}\right) \\
&= e^{-\triangle_{t_0-\varepsilon}^{t_0}\psi_t(r)}\left(1+\sum_{i=1}^{q}\binom{q}{i}(-1)^i\left(1-\varepsilon\int_0^\infty e^{-rs}(1-e^{-is})\eta'_{t_0}(s)\mathrm{d}s+o(\varepsilon)\right)\right) \\
&= e^{-\triangle_{t_0-\varepsilon}^{t_0}\psi_t(r)}\left(-\varepsilon\int_0^\infty e^{-rs}\sum_{i=1}^{q}\binom{q}{i}(-1)^i(1-e^{-is})\eta'_{t_0}(s)\mathrm{d}s+o(\varepsilon)\right) \\
&= e^{-\triangle_{t_0-\varepsilon}^{t_0}\psi_t(r)}\left(\varepsilon\int_0^\infty e^{-rs}(1-e^{-s})^q\eta'_{t_0}(s)\mathrm{d}s+o(\varepsilon)\right) \\
&= (1+o(1))\left(\varepsilon\int_0^\infty e^{-rs}(1-e^{-s})^q\eta'_{t_0}(s)\mathrm{d}s+o(\varepsilon)\right) \\
&= \varepsilon\int_0^\infty e^{-rs}(1-e^{-s})^q\eta'_{t_0}(s)\mathrm{d}s+o(\varepsilon)
\end{aligned}$$

□

We make use of the following Lemma to simplify the proof of the Theorem.

Lemma 2. *In the setting of Theorem 1 suppose the censored to the right data $\mathscr{D}$ is comprised of a sole observation t_1 with frequencies $n^e = n_1^e$ and $n^c = n_1^c$. Let $t < t_1$; then*

$$\mathbb{E}\left[e^{-\lambda\mu(0,t]}|\mu, \mathscr{D}\right] = \lim_{\varepsilon\to 0} \frac{\mathbb{E}\left[e^{-(\lambda n^c + n^e)\mu(0,t]}\right]}{\mathbb{E}\left[e^{-(n^c+n^e)\mu(0,t]}\right]}$$

Proof.

$$\begin{aligned}
\mathbb{E}\left[e^{-\lambda\mu(0,t]}|\mu, \mathscr{D}\right] &= \lim_{\varepsilon\to 0} \frac{\mathbb{E}\left[e^{-\lambda\mu(0,t]}e^{-n^c\mu(0,t_1]}\left(e^{-\mu(0,t_1-\varepsilon]} - e^{-\mu(0,t_1]}\right)^{n^e}\right]}{\mathbb{E}\left[e^{-n^c\mu(0,t_1]}\left(e^{-\mu(0,t_1-\varepsilon]} - e^{-\mu(0,t_1]}\right)^{n^e}\right]} \\
&= \lim_{\varepsilon\to 0} \frac{\mathbb{E}\left[e^{-\lambda\mu(0,t]-n^c\mu(0,t]-n^e\mu(0,t]}\right]\mathbb{E}\left[e^{-n^c\mu(t,t_1]}\left(e^{-\mu(t,t_1-\varepsilon]} - e^{-\mu(t,t_1]}\right)^{n^e}\right]}{\mathbb{E}\left[e^{-n^c\mu(0,t]-n^e\mu(0,t]}\right]\mathbb{E}\left[e^{-n^c\mu(t,t_1]}\left(e^{-\mu(t,t_1-\varepsilon]} - e^{-\mu(t,t_1]}\right)^{n^e}\right]} \\
&= \frac{\mathbb{E}\left[e^{-(\lambda+n^c+n^e)\mu(0,t]}\right]}{\mathbb{E}\left[e^{-(n^c+n^e)\mu(0,t]}\right]} = e^{-(\psi_t(\lambda+n^c+n^e)-\psi_t(n^c+n^e))}
\end{aligned}$$

□

We observe that if ψ_t is the Laplace exponent associated to the Lévy measure ν then $\psi_t^{(k)}(\lambda) = \psi_t(\lambda + k) - \psi_t(k)$ is the Laplace exponent associated to $e^{-ks}\nu(\mathrm{d}s, \mathrm{d}x)$, with this in mind we have that, in the following, without loss of generality for calculations of the posterior Laplace exponent as in the previous Lemma it suffices to consider evaluation of the exponent in a time t which is greater than all the survival times in the survival data $\boldsymbol{D}$.
Lets define

$$\begin{aligned}
\Gamma_{\mathscr{D},\varepsilon} = \bigcap_{j=1}^{k} \{(T_1, J_1, \ldots, T_n, J_n) \, : \, & m^c\left(\{T_{(j)}\}\right) = n_j^c \, , \\
& m^e\left((T_{(j)} - \varepsilon, T_{(j)}]\right) = n_j^e\}
\end{aligned}$$

so that

$$\mathbb{E}\left[e^{-\lambda\mu(0,t]}|\mathscr{D}\right] = \lim_{\varepsilon\to 0} \frac{\mathbb{E}\left[e^{-\lambda\mu(0,t]}\mathbb{1}_{\Gamma_{\mathscr{D},\varepsilon}}(\mathscr{D})\right]}{\mathbb{P}\left[\mathscr{D} \in \Gamma_{\mathscr{D},\varepsilon}\right]}.$$

We have that

$$\mathbb{E}\left[e^{-\lambda\mu(0.t]}\mathbb{1}_{\Gamma_{\mathscr{D},\varepsilon}}(\mathscr{D})\right] = \lim_{\varepsilon\to 0}\mathbb{E}\left[e^{-\lambda\mu(0.t]-\sum_{i=1}^{k} n_i^c\mu(0,T_{(i)}]}\prod_{i=1}^{k}\left(e^{-\mu(0,T_{(i)}-\varepsilon]}-e^{-\mu(0,T_{(i)}]}\right)^{n_i^e}\right]$$

$$= \mathbb{E}\left[e^{-\lambda\mu(0.t]-\sum_{i=1}^{k} n_i^c\mu(0,T_{(i)}]-\sum_{i=1}^{k} n_i^e\mu(0,T_{(i)}-\varepsilon]}\prod_{i=1}^{k}\left(1-e^{-\mu(T_{(i)}-\varepsilon,T_{(i)}]}\right)^{n_i^e}\right]$$

$$= \mathbb{E}\Bigg[e^{-\lambda\mu(T_{(k)}.t]-\sum_{r=1}^{k}(\lambda+\bar{n}_r^c+\bar{n}_r^e)\mu(T_{(r-1)},T_{(r)}-\varepsilon]-\sum_{r=1}^{k}\left(\lambda+\bar{n}_r^c+\bar{n}_{r+1}^e\right)\mu(T_{(r)}-\varepsilon,T_{(r)}]}$$

$$\times\prod_{i=1}^{k}\left(1-e^{-\mu(T_{(i)}-\varepsilon,T_{(i)}]}\right)^{n_i^e}\Bigg]$$

$$= \mathbb{E}\left[e^{-\lambda\mu(T_{(k)}.t]-\sum_{r=1}^{k}(\lambda+\bar{n}_r^c+\bar{n}_r^e)\mu(T_{(r-1)},T_{(r)}-\varepsilon]}\right]$$

$$\times\mathbb{E}\left[e^{-\sum_{r=1}^{k}\left(\lambda+\bar{n}_r^c+\bar{n}_{r+1}^e\right)\mu(T_{(r)}-\varepsilon,T_{(r)}]}\prod_{i=1}^{k}\left(1-e^{-\mu(T_{(i)}-\varepsilon,T_{(i)}]}\right)^{n_i^e}\right]$$

$$= \mathbb{E}[I_1]\,\mathbb{E}[I_2]$$

with

$$I_1 = e^{-\sum_{i=1}^{k+1}\left(\lambda+\bar{n}_i^c+\bar{n}_i^e\right)\mu(T_{(i-1)},T_{(i)}-\varepsilon]}$$

where we set $T_{(k+1)}=t$, $\bar{n}_{k+1}^c=0=\bar{n}_{k+1}^e$. On the other hand

$$I_2 = e^{-\sum_{r=1}^{k}\left(\lambda+\bar{n}_r^c+\bar{n}_{r+1}^e\right)\mu(T_{(r)}-\varepsilon,T_{(r)}]}\prod_{i=1}^{k}\left(1-e^{-\mu(T_{(i)}-\varepsilon,T_{(i)}]}\right)^{n_i^e}. \tag{1.26}$$

For the expectation in I_2 need to calculate quantities of the form

$$\mathbb{E}\left[e^{-r\mu(t_0-\varepsilon,t_0]}(1-e^{\mu(t_0-\varepsilon,t_0]})^q\right]$$

which are already given by the previous lemma for $q>0$; if $q=0$

$$\lim_{\varepsilon\to 0}\mathbb{E}\left[e^{-r\mu(t_0-\varepsilon,t_0]}(1-e^{\mu(t_0-\varepsilon,t_0]})^q\right] = \lim_{\varepsilon\to 0}\mathbb{E}\left[e^{-r\mu(t_0-\varepsilon,t_0]}\right] = 1$$

So if we define $\mathscr{J}=\{j : n_j^e>0\}$ then

$$\lim_{\varepsilon\to 0}\mathbb{E}[I_2] = \lim_{\varepsilon\to 0}\prod_{i\in\mathscr{J}}\left(\varepsilon\int_0^\infty e^{-\left(\lambda+\bar{n}_i^c+\bar{n}_{i+1}^e\right)s}(1-e^{-s})^{n_i^e}\eta'_{T_{(i)}}(s)\mathrm{d}s+o(\varepsilon)\right)$$

We also get that

$$\lim_{\varepsilon\to 0}\mathbb{E}[I_1] = e^{-\sum_{i=1}^{k+1}\left(\psi_{T_{(i)}}(\lambda+\bar{n}_i^c+\bar{n}_i^e)-\psi_{T_{(i-1)}}(\lambda+\bar{n}_i^c+\bar{n}_i^e)\right)}$$

So

$$\mathbb{E}\left[e^{-\lambda\mu_e(0.t]}\mathbb{1}_{\Gamma_{\boldsymbol{D},\varepsilon}}(\mathscr{D})\right] = e^{-\sum_{i=1}^{k+1}\left(\psi_{T_{(i)}}(\lambda+\bar{n}_i^c+\bar{n}_i^e)-\psi_{T_{(i-1)}}(\lambda+\bar{n}_i^c+\bar{n}_i^e)\right)}$$
$$\times\lim_{\varepsilon\to 0}\prod_{i\in\mathscr{J}}\left(\varepsilon\int_0^\infty e^{-\left(\lambda+\bar{n}_r^c+\bar{n}_{r+1}^e\right)s}(1-e^{-s})^{n_i^e}\eta'_{T_{(i)}}(s)\mathrm{d}s+o(\varepsilon)\right)$$

Analogously, or by Monotone convergence when $\lambda\to 0$, we get that

$$\mathbb{E}\left[\mathbb{1}_{\Gamma_{\mathscr{D}}}\right] = e^{-\sum_{i=1}^{k}\left(\psi_{T_i}(\bar{n}_i^c+\bar{n}_i^e)-\psi_{T_{(i-1)}}(\bar{n}_i^c+\bar{n}_i^e)\right)}$$
$$\times\lim_{\varepsilon\to 0}\prod_{i\in\mathscr{J}}\left(\varepsilon\int_0^\infty e^{-\left(\bar{n}_i^c+\bar{n}_{i+1}^e\right)s}(1-e^{-s})^{n_i^e}\eta'_{T_{(i)}}(s)\mathrm{d}s+o(\varepsilon)\right). \tag{1.27}$$

It follows that

$$\mathbb{E}\left[e^{-\lambda\mu_e(0,t]}|\mathscr{D}\right] = e^{-\sum_{i=1}^{k+1}\int_{\mathbb{R}^+\times(T_{(i-1)},T_{(i)}]}(1-e^{-\lambda s})e^{-(\bar{n}_i^c+\bar{n}_i^e)s}\nu(\mathrm{d}s,\mathrm{d}u)}$$
$$\times\prod_{j\in\mathscr{J}}\lim_{\varepsilon\to 0}\left(\frac{\varepsilon\int_0^\infty e^{-\left(\lambda+\bar{n}_r^c+\bar{n}_{r+1}^e\right)s}(1-e^{-s})^{n_i^e}\eta'_{T_{(j)}}(s)\mathrm{d}s+o(\varepsilon)}{\varepsilon\int_0^\infty e^{-\left(\bar{n}_r^c+\bar{n}_{r+1}^e\right)s}(1-e^{-s})^{n_i^e}\eta'_{T_{(j)}}(s)\mathrm{d}s+o(\varepsilon)}\right)$$
$$= e^{-\sum_{i=1}^{k+1}\int_{\mathbb{R}^+\times(T_{(i-1)},T_{(i)}]}(1-e^{-\lambda s})e^{-(\bar{n}_i^c+\bar{n}_i^e)s}\nu(\mathrm{d}s,\mathrm{d}u)}$$
$$\times\prod_{j\in M}\left(\frac{\int_0^\infty e^{-\left(\lambda+\bar{n}_r^c+\bar{n}_{r+1}^e\right)s}(1-e^{-s})^{n_i^e}\eta'_{T_{(i)}}(s)\mathrm{d}s}{\int_0^\infty e^{-\left(\bar{n}_r^c+\bar{n}_{r+1}^e\right)s}(1-e^{-s})^{n_i^e}\eta'_{T_{(i)}}(s)\mathrm{d}s}\right) \tag{1.28}$$

Proof of Corollary 1

In equation 1.28 above, take $\lambda = 1$ and using the discussion after Lemma 2 replace M with M_t to obtain the result.

Proof of Corollary 2

From equation 1.27 in the proof of Theorem 1 we obtain the desired likelihood.

Proof of Proposition 2

In the setting of Theorem 1.

$$\begin{aligned}
\psi_t^{\circ j}(\lambda) &= \int_{\left(t\wedge T_{(j-1)},t\wedge T_{(j)}\right)\times\mathbb{R}^+} (1-e^{-\lambda s})e^{-(\bar{n}_j^c+\bar{n}_j^e)s}\nu(s,\mathrm{d}x)\mathrm{d}s \\
&= \int_{\left(t\wedge T_{(j-1)},t\wedge T_{(j)}\right)\times\mathbb{R}^+} (1-e^{-(\lambda+\bar{n}_j^c+\bar{n}_j^e)s})\nu(s,\mathrm{d}x)\mathrm{d}s \\
&\quad - \int_{\left(t\wedge T_{(j-1)},t\wedge T_{(j)}\right)\times\mathbb{R}^+} (1-e^{-(\bar{n}_j^c+\bar{n}_j^e)s})\nu(s,\mathrm{d}x)\mathrm{d}s \\
&= \left(\psi_{t\wedge T_{(j)}}(\lambda+\bar{n}_j^c+\bar{n}_j^e) - \psi_{t\wedge T_{(j-1)}}(\lambda+\bar{n}_j^c+\bar{n}_j^e)\right) \\
&\quad - \left(\psi_{t\wedge T_{(j)}}(\bar{n}_j^c+\bar{n}_j^e) - \psi_{t\wedge T_{(j-1)}}(\bar{n}_j^c+\bar{n}_j^e)\right).
\end{aligned}$$

Proof of Proposition 3

In the setting of Proposition 3

$$\begin{aligned}
&\int_{\mathbb{R}^+\times(0,t]} e^{-qs}\left(1-e^{-s}\right)^n \nu(\mathrm{d}s,\mathrm{d}x) \\
&= \int_{\mathbb{R}^+\times(0,t]} e^{-qs}(1-e^{-s})\sum_{k=0}^{n-1}\binom{n-1}{k}(-1)^k e^{-ks}\nu(\mathrm{d}s\mathrm{d}x) \\
&= \sum_{k=0}^{n-1}\binom{n-1}{k}(-1)^k\int_{\mathbb{R}^+\times(0,t]} e^{-(k+q)s}(1-e^{-s})\nu(\mathrm{d}s,\mathrm{d}x) \\
&= \sum_{k=0}^{n-1}\binom{n-1}{k}(-1)^{k+1}\left(\int_{\mathbb{R}^+\times(0,t]}(1-e^{-(k+q)s})\nu(\mathrm{d}s,\mathrm{d}x)\right. \\
&\qquad \left. - \int_{\mathbb{R}^+\times(0,t]}(1-e^{-(k+1+q)s})\nu(\mathrm{d}s,\mathrm{d}x)\right) \\
&= \sum_{k=0}^{n-1}\binom{n-1}{k}(-1)^{k+1}\left(\psi_t(k+q)-\psi_t(k+1+q)\right)
\end{aligned}$$

Chapter 2

Compound random measures

2.1 Completely random measures

Griffin and Leisen (2017) introduced a flexible and tractable family of VCRM's. Their key idea was to construct a d-variate VCRM using as building blocks a CRM, i.e. an univariate VCRM, and a d-variate probability distribution. They call *Compound Random Measure* (CoRM) the particular family of VCRM's they propose. The following definition of a CoRM differs from the one in Griffin and Leisen (2017) since it takes into account the *inhomogeneous* case, where the locations and associated weights in the CRM are not independent as in (1.14).

Definition 12. A *Compound Random Measure* (CoRM) is a VCRM with Lévy intensity given by

$$\tilde{\nu}_d(\mathrm{d}\boldsymbol{s}, \mathrm{d}x) = \int_{\mathbb{R}^+} z^{-d} h\left(\frac{s_1}{z}, \ldots, \frac{s_d}{z}\right) \mathrm{d}\boldsymbol{s}\, \nu^\star(\mathrm{d}z, \mathrm{d}x) \tag{2.1}$$

where h is a d-variate probability density function which we call the *score distribution density*, and $\nu^\star$ is a Lévy intensity which we call the *directing Lévy measure*.

By performing a simple change of variable we note that

$$\int_{(\mathbb{R}^+)^d} z^{-d} h\left(\frac{s_1}{z}, \ldots, \frac{s_d}{z}\right) \mathrm{d}\boldsymbol{s} = 1.$$

Therefore, $z^{-d} h\left(\frac{s_1}{z}, \ldots, \frac{s_d}{z}\right)$ can be seen as the density of a distribution function

$$H\left(\frac{\mathrm{d}s_1}{z}, \ldots, \frac{\mathrm{d}s_d}{z}\right).$$

This allows to write the multivariate Lévy intensity in equation (2.1) as

$$\tilde{\nu}_d(\mathrm{d}\boldsymbol{s},\mathrm{d}x) = \int_{\mathbb{R}^+} H\left(\frac{\mathrm{d}s_1}{z},\ldots,\frac{\mathrm{d}s_d}{z}\right)\nu^\star(\mathrm{d}z,\mathrm{d}x), \tag{2.2}$$

and we call such H the *score distribution*. To write the Lévy intensity of a CoRM in terms of distribution functions rather than probability density functions will be convenient for the results presented in the next section. In the following we will say that the CRM associated to the directing Lévy measure $\nu^\star$ of a CoRM $\boldsymbol{\mu}$ is the directing CRM of the CoRM $\boldsymbol{\mu}$.

In Griffin and Leisen (2017), they used a CoRM to model a vector of Dirichlet processes which they used to fit a mixture model for heterogeneous clinical studies; furthermore they extend this approach in Griffin and Leisen (2018) where they use a mixture model based on a normalized CoRM where the score distribution depends on a covariate. On the other hand, Todeschini et al. (2016) recently used CoRM's for the modelling of graphs which allow overlapping communities. Finally in this thesis we will use CoRM's for survival analysis with multiple-sample information, Riva-Palacio and Leisen (2018) as discussed in Chapter 4, and for survival analysis regression in Chapter 5. We conclude the high potential for the use of CoRM's in the analysis of several heterogeneous datasets.

2.2 Integrability conditions

The specification of a CoRM needs the initial choice of a score distribution and a directing Lévy measure. Although this seems straightforward, it is necessary to check that this choices lead to a well defined CoRM. Otherwise, the associated stochastic process, in this case a random measure, is not well defined; in a Bayesian statistics setting where a probability distribution is given in terms of a CoRM we can be under risk of performing inference based on an ill-posed prior if the CoRM we are using is not well defined. In this section we look at two important aspects of Definition 12

1) We provide conditions on the score distribution and the directing Lévy measure for the existence of the marginal Lévy intensities of a CoRM, see Theorem 2 and Corollary 3,

2) We provide conditions on the score distribution and the directing Lévy measure for the existence of the multivariate Lévy intensity of a CoRM, see Theorem 3.

Essentially, Theorem 2 and Corollary 3 ahead focus on the existence of the marginals, (1.17), of a CoRM. On the other hand, Theorem 3 focuses on the global existence of a CoRM. The proofs of the theorems can be found at the end of the chapter, section 2.4.

Let H and $\nu^\star$ be, respectively, a score distribution and a directing Lévy measure which define a CoRM. We denote with H_j, $j \in \{1,\dots,\mathrm{d}\}$, the j-th marginal of a d-dimensional score distribution H. A simple change of variable leads to the j-th marginal of a CoRM, namely

$$\begin{aligned}\nu_j(A,X) &= \int_{\mathbb{R}^+\times X}\int_{A/z} H_j(\mathrm{d}s)\nu^\star(\mathrm{d}z,\mathrm{d}x)\\ &= \int_{\mathbb{R}^+}\int_{A\times X}\nu^\star\left(\frac{\mathrm{d}z}{s},\mathrm{d}x\right)H_j(\mathrm{d}s).\end{aligned} \tag{2.3}$$

We can see the formula above as a mean. Let S_j be a random variable with distribution H_j, then

$$\nu_j(A,X) = \mathbb{E}\left[\nu^\star\left(\frac{A}{S_j},X\right)\right] \tag{2.4}$$

for $A \in \mathscr{B}(\mathbb{R}^+)$. We use the last identity to give conditions for the marginal intensity ν_j to be a proper Lévy intensity, i.e. a measure that satisfies the condition displayed in equation (1.2).

Theorem 2. *Let H be a d-variate score distribution and $\nu^\star$ a directing Lévy measure defining a measure $\tilde{\nu}_d$ as in* (2.2) *with corresponding marginals ν_j for $j \in \{1,\dots,\mathrm{d}\}$. Let X be a bounded set in $\mathscr{X}$, then the measure ν_j satisfies the integrability condition (1.2) if and only if*

$$\int_{(0,1)\times X}\mathbb{P}\left[S_j \geq \frac{1}{z}\right]\nu^\star(\mathrm{d}z,\mathrm{d}x) < \infty \tag{2.5}$$

and

$$\int_{[1,\infty)\times X}\mathbb{P}\left[S_j \leq \frac{1}{z}\right]z\nu^\star(\mathrm{d}z,\mathrm{d}x) < \infty. \tag{2.6}$$

Furthermore if the marginal score H_j satisfies that

$$1 - H_j\left(\frac{1}{z}\right) \leq z\,\forall\, z \in (0,\varepsilon)\ \textit{for some}\ \varepsilon > 0 \tag{2.7}$$

and

$$\lim_{z\to\infty} zH_j\left(\frac{1}{z}\right) < \infty \tag{2.8}$$

then conditions (2.5), (2.6) are satisfied with an arbitrary choice of the directing Lévy measure $\nu^\star$.

As set in Definition 12, we usually work with CoRM's given by a score with a probability density; in such case the following corollary to Theorem 2 follows.

Corollary 3. *If S_j has probability density function h_j then conditions (2.7)-(2.8) reduce to*

$$\lim_{z\to 0} \frac{h_j\left(\frac{1}{z}\right)}{z^2} < 1 \tag{2.9}$$

and

$$\lim_{\varepsilon\to 0} h_j(\varepsilon) < \infty. \tag{2.10}$$

The previous results concerned conditions for the marginals of a CoRM to be well defined, now we focus on such a result for the CoRM. For a score density function h and directing Lévy measure $\nu^\star$ to properly define a CoRM we need to check the condition (1.13) which takes the form

$$\int_{\mathbb{R}^+\times X}\int_{(\mathbb{R}^+)^d} \min\{1,\|\boldsymbol{s}\|\} h\left(\frac{s_1}{z},\ldots,\frac{s_d}{z}\right)\frac{\mathrm{d}\boldsymbol{s}}{z^d}\nu^\star(\mathrm{d}z,\mathrm{d}x) < \infty \tag{2.11}$$

for bounded set $X \in \mathscr{X}$. As stated at the beginning of this section, in the next theorem we provide conditions on the score distribution and the directing Lévy measure for the existence of the multivariate Lévy intensity of a CoRM. This is equivalent to provide conditions such that the above inequality holds true.

Theorem 3. *Consider a CoRM which satisfies conditions (2.5) and (2.6) for each marginal ν_j, $j \in \{1,\ldots,\mathrm{d}\}$, then the integrability condition (2.11) is satisfied.*

We conclude this section by providing three examples of the use of the previous results when considering Gamma, Beta and LogNormal distributed score distributions.

Example 8. Gamma scores
We consider the marginal gamma score case. Let h be the d-variate probability density of the score distribution; for $j \in \{1,\ldots,d\}$ we denote the j-th marginal density h_j and let it correspond to a Gamma distribution with shape and rate parameters α_j, β_j, i.e.

$$h_j(s) = \frac{\beta_j^{\alpha_j} s^{\alpha_j-1} e^{-\beta_j s}}{\Gamma(\alpha_j)}\mathbb{1}_{\{s\in(0,\infty)\}}.$$

We check the constraints (2.7), (2.8) by making use of Corollary 3 as we have probability densities. To check (2.9) we see that

$$\lim_{s\to 0}\frac{h_j\left(\frac{1}{s}\right)}{s^2}=\lim_{s\to 0}\frac{\beta_j^{\alpha_j}e^{-\frac{\beta_j}{s}}}{\Gamma(\alpha_j)s^{\alpha_j+1}}=0$$

and constraint (2.10) is satisfied for arbitrary Lévy directing measure $\nu^\star$ whenever $\alpha_j \geq 1$, as in the examples presented in Griffin and Leisen (2017). However for $\alpha_j < 1$ the associated CoRM will be well posed depending on the choice of $\nu^\star$. If for example we take the directing Lévy measure to be the σ-stable, i.e.

$$\nu^\star(\mathrm{d}z,\mathrm{d}x)=\frac{A\sigma}{\Gamma(1-\sigma)z^{1+\sigma}}\mathrm{d}z\mathrm{d}x$$

then constraint (2.5) in Theorem 2 can be reduced to

$$\frac{A}{\Gamma(1-\sigma)}\int_0^1\int_{\frac{1}{z}}^{\infty}h_j(s)\frac{\sigma}{z^{1+\sigma}}\mathrm{d}s\mathrm{d}z=\frac{A}{\Gamma(1-\sigma)}\int_1^{\infty}\int_{\frac{1}{s}}^{1}h_j(s)\frac{\sigma}{z^{\sigma+1}}\mathrm{d}z\mathrm{d}s$$
$$=\frac{A}{\Gamma(1-\sigma)}\int_1^{\infty}h_j(s)(s^{\sigma}-1)\mathrm{d}s<\infty,$$

which is always satisfied since h_j is a Gamma density. On the other hand, condition (2.6) in Theorem 2 becomes

$$\frac{A}{\Gamma(1-\sigma)}\int_1^{\infty}\int_0^{\frac{1}{z}}h_j(s)\frac{\sigma z}{z^{1+\sigma}}\mathrm{d}s\mathrm{d}z=\frac{A}{\Gamma(1-\sigma)}\int_0^1\int_1^{\frac{1}{s}}h_j(s)\frac{\sigma}{z^{\sigma}}\mathrm{d}z\mathrm{d}s$$
$$=\frac{A\sigma}{\Gamma(2-\sigma)}\int_0^1 h_j(s)(s^{\sigma-1}-1)\mathrm{d}s<\infty$$

which is not satisfied when $\alpha_j+\sigma<1$.

Example 9. Beta scores

In the setting as above, if the marginal scores are Beta distributed, i.e.

$$h_j(s)=\frac{s^{\alpha_j-1}(1-s)^{\beta_j-1}}{B(\alpha_j,\beta_j)}\mathbb{1}_{\{s\in(0,1)\}}$$

then constraint (2.9) becomes

$$\lim_{s\to 0}\frac{h_j\left(\frac{1}{s}\right)}{s^2}=\lim_{s\to 0}\frac{(s-1)^{\beta_j-1}}{s^{\alpha_j+\beta_j}B(\alpha_j,\beta_j)}=0,$$

so it is always satisfied; and condition (2.8) is satisfied whenever $\alpha_j \geq 1$. We consider again a σ-stable Lévy intensity for $\nu^\star$ when $\alpha_j < 1$. Proceeding as in the previous example, constraint (2.5) becomes

$$\frac{A}{\Gamma(1-\sigma)}\int_1^\infty \int_{\frac{1}{s}}^1 h_j(s)\frac{\sigma}{z^{\sigma+1}}\mathrm{d}z\mathrm{d}s < \infty$$

so it always holds; and constraint (2.6) becomes

$$\frac{A}{\Gamma(1-\sigma)}\int_0^1 \int_1^{\frac{1}{s}} h_j(s)\frac{\sigma}{z^{\sigma}}\mathrm{d}z\mathrm{d}s < \infty,$$

which holds for $\sigma + \alpha_j > 1$.

Example 10. LogNormal scores

We check conditions (2.9) and (2.10) for an arbitrary LogNormal distribution with density

$$h(z) = \frac{1}{z\sigma\sqrt{2\pi}}e^{-\frac{(\ln(z)-\mu)^2}{2\sigma^2}}.$$

We have that

$$\begin{aligned}\lim_{z\to 0}\frac{h(1/z)}{z^2} &= \lim_{z\to 0}\frac{1}{z\sigma\sqrt{2\pi}}e^{-\frac{(\ln(1/z)-\mu)^2}{2\sigma^2}}\\ &= \lim_{z\to 0}\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-\ln(1/z)^2+2(\mu+\sigma^2)\ln(1/z)-\mu^2}{2\sigma^2}} = 0\end{aligned}$$

and

$$\begin{aligned}\lim_{z\to 0}h(z) &= \lim_{z\to 0}\left(\frac{1}{z\sigma\sqrt{2\pi}}e^{-\frac{(\ln(z)-\mu)^2}{2\sigma^2}}\right)\lim_{z\to 0}\left(\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-\ln(z)^2+2(\mu-\sigma^2)\ln(z)-\mu^2}{2\sigma^2}}\right)\\ &= 0\end{aligned}$$

So LogNormal scores define a CoRM for any choice of the directing Lévy measure $\nu^\star$.

Example 11. Mixture distribution scores

We will see that a mixture of score distributions which satisfy the conditions of Corollary 3 attains the integrability conditions.

Definition 13. Let $n \in \mathbb{N}\setminus\{0\}$ and $H^{(i)}$ be a $d-$variate probability distribution for each $i \in \{1,\ldots,n\}$. Given $\boldsymbol{w} = (w_1,\ldots,w_n)$ in the n-dimensional simplex, i.e. with positive entries and such that $\sum_{i=1}^n w_i = 1$, then we say that the mixture of $H^{(1)},\ldots,H^{(n)}$ with weights

$\boldsymbol{w}$ is the distribution given by

$$H = \sum_{i=1}^{n} w_i H^{(i)},$$

denoted $w_1 \mathscr{L}(H^{(1)}) + \ldots + w_n \mathscr{L}(H^{(n)})$ where we can refer to the law of the probability distributions by their name, cumulative distribution, survival function or density function.

It is straightforward that if $\{h^{(i)}\}_{i=1}^{n}$ are probability densities satisfying the conditions in Corollary 3, then then any mixture distribution of them has a probability density that attain the Corollary 3 as well.

In what follows we denote a CoRM with score distribution "distribution" and directing Lévy measure "directing" as "distribution"-"directing" CoRM; for example we can have a LogNormal-Gamma CoRM or a Beta-σ-stable CoRM. If the "distribution" label used corresponds to univariate distribution we assume that the vector given by the score has mutually independent entries with marginal distribution as the label; furthermore the directing Lévy measure is taken to be homogeneous if it is not otherwise indicated in the label.

2.3 Other interesting properties

The aim of this section is to investigate four interesting properties of CoRM's. First, we focus on CoRM's which arise from regurlarly varying directing Lévy measures. This result is motivated by the recent papers of Caron and Fox (2017) and Todeschini et al. (2016) which made use of regularly varying Lévy measures to construct sparse random graphs. Second, we provide an explicit expression of the multivariate Lévy intensity of a CoRM with independent exponential scores. This result is interesting when compared with Theorem 3.2 in Zhu and Leisen (2015) and Corollary 2 in Griffin and Leisen (2017) which provide, respectively, the Lévy copula representation and the Laplace exponent of CoRM's with independent exponential scores. Third we give a formula for the Laplace exponent of a CoRM which can be approximated via Monte-Carlo methods. Finally, we give a series representation for a CoRM which can be used for simulation purposes and that will be of use later on to prove results concerning the model in Chapter 5. The proofs of the results can be found at the end of the chapter.

For the results that deal with a d-variate CoRM given by an homogeneous directing Lévy intensity $\nu^{\star}(\mathrm{d}z, \mathrm{d}x) = \rho^{\star}(\mathrm{d}z)\kappa(\mathrm{d}x)$, we observe that the corresponding marginals of the CoRM can be written in the form $\nu_j(\mathrm{d}s, \mathrm{d}x) = \rho_j(\mathrm{d}s)\kappa(\mathrm{d}x)$ as discussed in (1.18).

2.3.1 Regularly varying directing Lévy measure

In this section we focus on CoRM's given by a directing Lévy measure that is regularly varying. We recall that a real valued function L is slowly varying if $\lim_{t\to\infty} L(at)/L(t) = 1\ \forall a > 0$.

Definition 14. An homogeneous Lévy measure $\rho^\star(\mathrm{d}z)\kappa(\mathrm{d}x)$ in $\mathbb{R}^+ \times \mathbb{X}$ is said to be regularly varying if the tail integral $U^\star(y) = \int_y^\infty \rho^\star(\mathrm{d}s)$ is a regularly varying function, i.e. it satisfies

$$U^\star(y) = L\left(\frac{1}{y}\right)\frac{1}{y^\sigma}$$

for some $\sigma \in [0,1)$, which we call the regular variation index, and L a slowly varying function.

The following Theorem highlights an interesting link between the directing Lévy measure and the marginal Lévy intensities in terms of the regularly varying property.

Theorem 4. *Consider a CoRM with an homogeneous directing Lévy measure $\rho^\star(\mathrm{d}s)\kappa(\mathrm{d}x)$ such that the conditions of Theorem 3 are satisfied. If $\rho^\star$ is regularly varying with tail integral U then the marginals ρ_j, $j \in \{1,\dots,d\}$, are regularly varying.*

Example. σ-stable directing Lévy measure
Consider a σ-stable directing Lévy measure

$$\nu^\star(\mathrm{d}s,\mathrm{d}x) = \frac{\sigma}{\Gamma(1-\sigma)s^{\sigma+1}}\mathrm{d}s\mathrm{d}x.$$

The related tail integral is

$$U^\star(y) = \frac{1}{\Gamma(1-\sigma)y^\sigma}$$

which is a regularly varying function with index σ and slowly varying function $L(y) = \frac{1}{\Gamma(1-\sigma)}$. We see that the regularly varying tail integrals related to the CoRM construction arise as a factor of $U^\star$, namely $U_j(y) = \mathbb{E}\left[S_j^\sigma\right]U^\star(y)$ and from Theorem 4 the associated marginal tail integrals are regularly varying.

Regularly varying CRM's are of interest in the work of Caron and Fox (2017) and Todeschini et al. (2016) as they are related to the asymptotic properties of their models, see Caron and Rousseau (2017).

2.3.2 Independent Exponential scores

Consider a d-variate CoRM given by an homogeneous directing Lévy measure $\nu^\star(\mathrm{d}z,\mathrm{d}x) = \rho^\star(\mathrm{d}z)\kappa(\mathrm{d}x)$ and a score distribution corresponding to d independent standard exponential distributions, i.e.

$$h(s_1,\ldots,s_d) = \prod_{i=1}^{d} e^{-s_i}.$$

We observe that each associated marginal takes the form $\nu_j(\mathrm{d}s,\mathrm{d}x) = f(s)\mathrm{d}s\kappa(\mathrm{d}x)$, where $f(s) = \int_0^\infty z^{-1}e^{-\frac{s}{z}}\rho^\star(\mathrm{d}z)$. The following Theorem provides a characterization for this class of CoRM's.

Theorem 5. *Consider a CoRM as described above; the corresponding d-variate Lévy intensity* $\tilde{\nu}_d(\mathrm{d}\boldsymbol{s},\mathrm{d}x) = \tilde{\rho}_d(\boldsymbol{s})\mathrm{d}\boldsymbol{s}\kappa(\mathrm{d}x)$ *is such that*

$$\tilde{\rho}_d(\boldsymbol{s}) = (-1)^{d-1}\frac{\partial^{d-1}}{\partial s^{d-1}}f(s)\bigg|_{s=s_1+\ldots+s_d}.$$

2.3.3 Laplace exponent of a CoRM.

We can express the Laplace exponent of a CoRM in terms of the Laplace exponent of its directing Lévy measure as follows.

Theorem 6. *Let* $\boldsymbol{\mu}$ *be a d-variate CoRM given by a score distribution H and directing Lévy intensity* $\nu^\star$ *associated to a Laplace exponent* $\psi_t^\star$*, then the Laplace exponent is given by*

$$\psi_t(\lambda_1,\ldots,\lambda_d) = \mathbb{E}[\psi_t^\star(\lambda_1 W_1 - \ldots - \lambda_d W_d)]$$

where $(W_1,\ldots,W_d) \sim H$.

The inferential schemes for the models presented in Chapters 4 and 5 rely heavily on the evaluation of the Laplace exponent of an underlying VCRM; so the above result is of special interest for using a CoRM in such models when its Laplace exponent is not explicitly available but the one corresponding to the directing Lévy measure is.

2.3.4 Series representation of a CoRM

We can use the second part of Proposition 1 to obtain a series representation of a CoRM given a series representation of the directing CRM. We illustrate such procedure in the next theorem.

Theorem 7. *Let $\boldsymbol{\mu}$ be a CoRM given by a score distribution H and a directing CRM $\mu^\star$ with a series representations*

$$\mu^\star = \sum_{i=1}^{\infty} W_i^\star \delta_{X_i}.$$

If

$$\left(Z_{1,i},\ldots,Z_{d,i}\right) \overset{i.i.d.}{\sim} H,\ i \in \{1,2,\ldots\}$$

then

$$\boldsymbol{\mu} \overset{a.s.}{=} \sum_{i=1}^{\infty} \left(Z_{1,i}W_i,\ldots,Z_{d,i}W_i\right) \delta_{X_i}. \tag{2.12}$$

We can use the above result to set an algorithm for simulation of a CoRM with score distribution H and directing CRM $\mu^\star$ as follows

Algorithm 3 CoRM simulation

1: Use the Ferguson-Klass algorithm, Algorithm 1, or some other algorithm to generate a truncated series approximation of the directing CRM

$$\mu^\star \approx \sum_{i=1}^{K} W_i^\star \delta_{X_i}.$$

for some $K \in \mathbb{N}$.

2: Sample

$$\left(Z_{1,i},\ldots,Z_{d,i}\right) \overset{\text{i.i.d.}}{\sim} H,\ i \in \{1,\ldots,K\}.$$

3: Approximate $\boldsymbol{\mu}$ by using

$$\boldsymbol{\mu} \approx \sum_{i=1}^{K} \left(Z_{1,i}W_i,\ldots,Z_{d,i}W_i\right) \delta_{X_i}.$$

We denote LogNormal$(\boldsymbol{m},\boldsymbol{\Sigma})$ for a d-variate LogNormal distribution associated to a d-variate Normal distribution with vector of means $\boldsymbol{m}$ and variance-covariance matrix $\boldsymbol{\Sigma}$. Let $I^{(d)}$ be the d-variate identity matrix. With such notation we plot a LogNormal-Gamma CoRM in Figure 2.1.

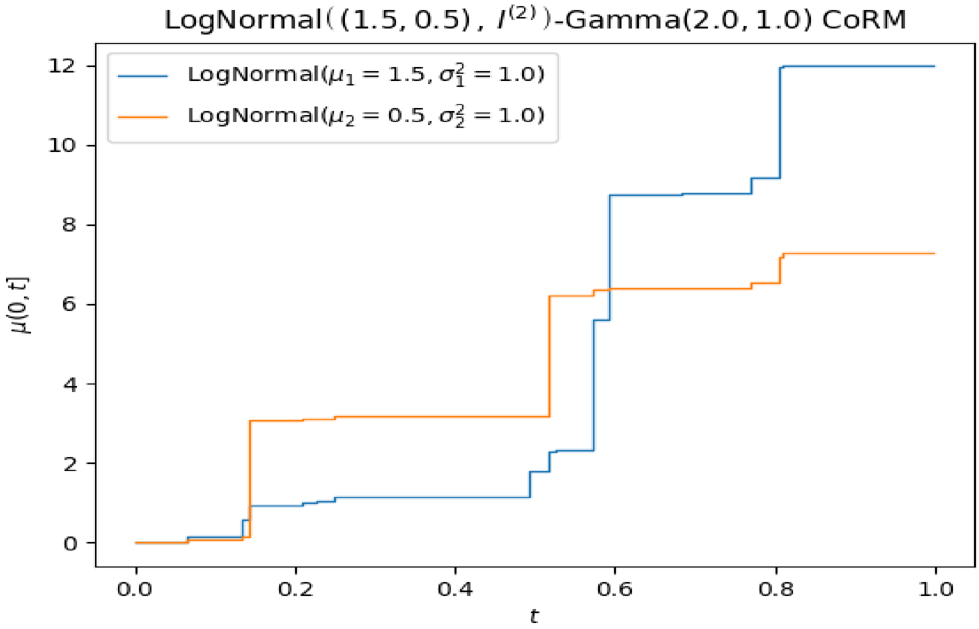


Fig. 2.1 Plot of the entries of $\boldsymbol{\mu}\left((0,t]\times(0,t]\right)$ when a LogNormal $\left(\boldsymbol{m}=(1.5,0.5)\,,I^{(2)}\right)$-Gamma$(2,1)$ CoRM is considered, the Gamma directing Lévy measure is homogeneous, i.e. $\kappa(\mathrm{d}x)=\mathrm{d}x$ as it was not otherwise stated. The random vector related to the secore distribution has mutually independent entries due to the choice of variance-covariance matrix but they are not identically distributed due to the vector of means choice. The simulation was obtained by using Algorithm 3. The underlying Gamma process was obtained by using Algorithm 1 as indicated in step 1 of the CoRM simulation algorithm.

2.4 Chapter 2 proofs

Proof of Theorem 2

We recall that $\nu^\star$ satisfies (1.2) since it is a Lévy intensity. Using (2.4), condition (1.2) for ν_j becomes

$$\mathbb{E}\left[\int_{\mathbb{R}^+\times X}\min\{1,z\}\nu^\star\left(\frac{\mathrm{d}z}{S_j},\mathrm{d}x\right)\right]$$
$$=\mathbb{E}\left[\int_{\left(0,\frac{1}{S_j}\right)\times X} z\nu^\star(\mathrm{d}z,\mathrm{d}x)\right]+\mathbb{E}\left[\int_{\left[\frac{1}{S_j},\infty\right)\times X}\nu^\star(\mathrm{d}z,\mathrm{d}x)\right]<\infty. \tag{2.13}$$

Therefore, ν_j satisfies (1.2) if and only if

$$\mathbb{E}\left[\int_{\left(0,\frac{1}{S_j}\right)\times X} z\nu^\star(\mathrm{d}z,\mathrm{d}x)\right] < \infty \tag{2.14}$$

and

$$\mathbb{E}\left[\int_{\left[\frac{1}{S_j},\infty\right)\times X} \nu^\star(\mathrm{d}z,\mathrm{d}x)\right] < \infty. \tag{2.15}$$

The former can be decomposed using the Fubini-Tonelli theorem in

$$\mathbb{E}\left[\int_{\left(0,\frac{1}{S_j}\right)\times X} z\nu^\star(\mathrm{d}z,\mathrm{d}x)\right] = \int_{\mathbb{R}^+\times X} \mathbb{P}\left[S_j \le \frac{1}{z}\right] z\nu^\star(\mathrm{d}z,\mathrm{d}x)$$
$$= \int_{(0,1)\times X} \mathbb{P}\left[S_j \le \frac{1}{z}\right] z\nu^\star(\mathrm{d}z,\mathrm{d}x) + \int_{[1,\infty)\times X} \mathbb{P}\left[S_j \le \frac{1}{z}\right] z\nu^\star(\mathrm{d}z,\mathrm{d}x).$$

Condition (2.6) ensures that the second term of the above equation is finite. It is easy to see that the first term is finite as well. Indeed,

$$\int_{(0,1)\times X} \mathbb{P}\left[S_j \le \frac{1}{z}\right] z\nu^\star(\mathrm{d}z,\mathrm{d}x) \le \int_{(0,1)\times X} z\nu^\star(\mathrm{d}z,\mathrm{d}x) < \infty.$$

On the other hand, the second term in (2.13) can be decomposed in

$$\mathbb{E}\left[\int_{\left[\frac{1}{S_j},\infty\right)\times X} \nu^\star(\mathrm{d}z,\mathrm{d}x)\right] = \int_{\mathbb{R}^+\times X} \mathbb{P}\left[\frac{1}{z} \le S_j\right] \nu^\star(\mathrm{d}z,\mathrm{d}x)$$
$$= \int_{(0,1)\times X} \mathbb{P}\left[S_j \ge \frac{1}{z}\right] \nu^\star(\mathrm{d}z,\mathrm{d}x) + \int_{[1,\infty)\times X} \mathbb{P}\left[S_j \ge \frac{1}{z}\right] \nu^\star(\mathrm{d}z,\mathrm{d}x).$$

Condition (2.5) ensures that the first term of the above equation is finite. It is easy to see that the second term is finite as well. Indeed,

$$\int_{[1,\infty)\times X} \mathbb{P}\left[S_j \ge \frac{1}{z}\right] \nu^\star(\mathrm{d}z,\mathrm{d}x) \le \int_{[1,\infty)\times X} \nu^\star(\mathrm{d}z,\mathrm{d}x) < \infty.$$

Therefore, the first part of the theorem follows from (2.13), (2.14) and (2.15).
For the remaining part of the Theorem we use that (1.2) is attained when considering the

directing Lévy measure $\nu^\star$. Indeed, if

$$\lim_{z\to\infty} z\mathbb{P}\left[S_j \leq \frac{1}{z}\right] < \infty \tag{2.16}$$

then as $\nu^\star$ is a Lévy intensity

$$\int_{[1,\infty)\times X} \mathbb{P}\left[S_j \leq \frac{1}{z}\right] z\nu^\star(\mathrm{d}z, \mathrm{d}x) < \infty. \tag{2.17}$$

so (2.14) holds. And if there exists $\varepsilon > 0$ such that $1 - H_j\left(\frac{1}{z}\right) \leq z \,\forall\, z \in (0, \varepsilon)$ then

$$\int_{(0,1)\times X} \mathbb{P}\left[\frac{1}{z} \leq S_j\right] \nu^\star(\mathrm{d}z, \mathrm{d}x) < \int_{(0,1)\times X} z\nu^\star(\mathrm{d}z, \mathrm{d}x) < \infty,$$

so (2.15) also holds. From the first part of the theorem the CoRM marginal ν_j satisfies the integrability conditions for arbitrary $\nu^\star$.

Proof of Corollary 3

We define $f(z) = z - \left(1 - H_j\left(\frac{1}{z}\right)\right)$ and observe that $f(0^+) = 0$ so the existence of $f'(0^+) > 0$ implies (2.7). As S_j has a probability density we get that $f'(0^+)$ exists and (2.7) is equivalent to $f'(0^+) > 0$ which we write as

$$\lim_{z\to 0} \frac{h_j\left(\frac{1}{z}\right)}{z^2} < 1.$$

Using the fundamental theorem of calculus we see that (2.8) reduces to

$$\lim_{z\to\infty} z\mathbb{P}\left[S_j \leq \frac{1}{z}\right] = \lim_{\varepsilon\to 0} h_j(\varepsilon) < \infty$$

which is satisfied when h_j is continuous at zero.

Proof of Theorem 3

Denote $P_j = \{\boldsymbol{s} \in (\mathbb{R}^+)^d \,:\, \max\{s_1, \ldots, s_d\} = s_j\}$ for $j \in \{1, \ldots, \mathrm{d}\}$; then, by using (2.3) and the fact that each ν_j is a Lévy intensity we get that for any bounded set X in $\mathscr{X}$

$$
\begin{aligned}
&\int_{\mathbb{R}^+\times X}\int_{(\mathbb{R}^+)^d} \min\{1, \|\boldsymbol{s}\|\} h\left(\frac{s_1}{z}, \ldots, \frac{s_d}{z}\right) \frac{\mathrm{d}\boldsymbol{s}}{z^d} \nu^\star(\mathrm{d}z, \mathrm{d}x) \\
&= \sum_{j=1}^d \int_{\mathbb{R}^+\times X}\int_{P_j} \min\{1, \|\boldsymbol{s}\|\} h\left(\frac{s_1}{z}, \ldots, \frac{s_d}{z}\right) \frac{\mathrm{d}\boldsymbol{s}}{z^d} \nu^\star(\mathrm{d}z, \mathrm{d}x) \\
&\le \sum_{j=1}^d \int_{\mathbb{R}^+\times X}\int_{P_j} \min\{1, \sqrt{d} s_j\} h\left(\frac{s_1}{z}, \ldots, \frac{s_d}{z}\right) \frac{\mathrm{d}\boldsymbol{s}}{z^d} \nu^\star(\mathrm{d}z, \mathrm{d}x) \\
&\le \sum_{j=1}^d \int_{\mathbb{R}^+\times X}\int_{(\mathbb{R}^+)^d} \min\{1, \sqrt{d} s_j\} h\left(\frac{s_1}{z}, \ldots, \frac{s_d}{z}\right) \frac{\mathrm{d}\boldsymbol{s}}{z^d} \nu^\star(\mathrm{d}z, \mathrm{d}x) \\
&= \sum_{j=1}^d \int_{\mathbb{R}^+\times X}\int_{\mathbb{R}^+} \min\{1, \sqrt{d} s\} h_j\left(\frac{s}{z}\right) \frac{\mathrm{d}s}{z} \nu^\star(\mathrm{d}z, \mathrm{d}x) \\
&= \sum_{j=1}^d \int_{\mathbb{R}^+\times X} \min\{1, \sqrt{d} s\} \nu_j(\mathrm{d}s, \mathrm{d}x) \\
&\le \sum_{j=1}^d \int_{\mathbb{R}^+\times X} \min\{\sqrt{d}, \sqrt{d} s\} \nu_j(\mathrm{d}s, \mathrm{d}x) \\
&= \sqrt{d} \sum_{j=1}^d \int_{\mathbb{R}^+\times X} \min\{1, s\} \nu_j(\mathrm{d}s, \mathrm{d}x) < \infty.
\end{aligned}
$$

Proof of Theorem 4

We recall that for the case at hand

$$
U^\star(y) = L\left(\frac{1}{y}\right)\frac{1}{y^\sigma} \tag{2.18}
$$

is a tail integral.

Proof. We note that equation (2.4) implies that

$$
\rho_j(A) = \mathbb{E}\left[\rho^\star\left(\frac{A}{S_j}\right)\right]
$$

It follows that the marginals of the CoRM are given by

$$\begin{aligned}
U_j(y) &= \rho_j\left((y,\infty)\right) \\
&= \mathbb{E}\left[U^\star\left(\frac{y}{S_j}\right)\right] \\
&= \mathbb{E}\left[L\left(\frac{S_j}{y}\right)\left(\frac{S_j}{y}\right)^\sigma\right] \\
&= \mathbb{E}\left[L\left(\frac{S_j}{y}\right)S_j^\sigma\right]\frac{1}{y^\sigma}.
\end{aligned}$$

Hence, it is enough to check if the function $l(z) = \mathbb{E}\left[L(S_j z)S_j^\sigma\right]$ is slowly varying for L a slowly varying function. Let $a > 0$, we need to check

$$\lim_{t\to\infty}\frac{l(at)}{l(t)} = \lim_{t\to\infty}\frac{\mathbb{E}\left[L(atS_j)S_j^\sigma\right]}{\mathbb{E}\left[L(tS_j)S_j^\sigma\right]} = 1.$$

For a fixed $\varepsilon > 0$ we can choose t_0 such that $\forall u > t_0$

$$|L(au)/L(u) - 1| < \frac{\varepsilon}{2},$$

since L is slowly varying. Then for $t > t_0$

$$\begin{aligned}
&\left|\frac{\mathbb{E}\left[L(atS_j)S_j^\sigma\right]}{\mathbb{E}\left[L(tS_j)S_j^\sigma\right]} - 1\right| = \left|\frac{\mathbb{E}\left[S_j^\sigma\left(L(atS_j) - L(tS_j)\right)\right]}{\mathbb{E}\left[L(tS_j)S_j^\sigma\right]}\right| \\
&\leq \frac{\mathbb{E}\left[\mathbb{1}_{\left\{S_j > \frac{t_0}{t}\right\}}S_j^\sigma\left|L(atS_j) - L(tS_j)\right|\right]}{\mathbb{E}\left[L(tS_j)S_j^\sigma\right]} + \frac{\mathbb{E}\left[\mathbb{1}_{\left\{S_j \leq \frac{t_0}{t}\right\}}S_j^\sigma\left|L(atS_j) - L(tS_j)\right|\right]}{\mathbb{E}\left[L(tS_j)S_j^\sigma\right]} \\
&< \frac{\mathbb{E}\left[\mathbb{1}_{\left\{S_j > \frac{t_0}{t}\right\}}S_j^\sigma\frac{\varepsilon}{2}L(tS_j)\right]}{\mathbb{E}\left[L(tS_j)S_j^\sigma\right]} + \frac{\mathbb{E}\left[\mathbb{1}_{\left\{S_j \leq \frac{t_0}{t}\right\}}S_j^\sigma\left|L(atS_j) - L(tS_j)\right|\right]}{\mathbb{E}\left[L(tS_j)S_j^\sigma\right]} \\
&< \frac{\varepsilon}{2} + \frac{\mathbb{E}\left[\mathbb{1}_{\left\{S_j \leq \frac{t_0}{t}\right\}}S_j^\sigma\left|L(atS_j) - L(tS_j)\right|\right]}{\mathbb{E}\left[L(tS_j)S_j^\sigma\right]}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{\varepsilon}{2} + \frac{\mathbb{E}\left[\mathbb{1}_{\left\{S_j \le \frac{t_0}{t}\right\}} S_j^{\sigma} t^{\sigma} \left|L(atS_j) - L(tS_j)\right|\right]}{U_j(\frac{1}{t})} \\
&= \frac{\varepsilon}{2} + \frac{\int_{(0,\frac{t_0}{t}]} s^{\sigma} t^{\sigma} \left|L(ats) - L(ts)\right| H_j(\mathrm{d}s)}{U_j(\frac{1}{t})} \\
&= \frac{\varepsilon}{2} + \frac{\int_{(0,t_0]} u^{\sigma} \left|L(au) - L(u)\right| H_j\left(\frac{\mathrm{d}u}{t}\right)}{U_j(\frac{1}{t})} \qquad (2.19)
\end{aligned}
$$

We observe that $\lim_{x\to 0} U_j(x) = \infty$. since U_j is a tail integral. From (2.18) it follows that $\lim_{x\to 0} x^{\sigma} (L(ax) - L(x)) = 0$. Hence, the function $g(x) = x^{\sigma} (L(ax) - L(x))$ is bounded in $[0,t_0]$ by a constant K_{1,t_0}. Finally we observe that for $t > t_0$

$$
\int_{(0,t_0]} H_j\left(\frac{\mathrm{d}u}{t}\right) < \int_{(0,1]} H_j(\mathrm{d}u) \le 1.
$$

We set $t_1 > t_0$ such that for $u > t_1$

$$
\frac{2K_{1,t_0}}{\varepsilon} < U_j(1/u).
$$

Choosing $t > t_1$ we get

$$
\frac{\int_{(0,t_0]} u^{\sigma} \left|L(au) - L(u)\right| H_j\left(\frac{\mathrm{d}u}{t}\right)}{U_j(\frac{1}{t})} < \frac{K_{1,t_0}}{U_j(\frac{1}{t})} < \frac{\varepsilon}{2}
$$

It follows from (2.19) that

$$
\left|\frac{\mathbb{E}\left[L(atS_j)S_j^{\sigma}\right]}{\mathbb{E}\left[L(tS_j)S_j^{\sigma}\right]} - 1\right| < \varepsilon.
$$

Consequently, l defined above is slowly varying, implying that the marginal tail integral U_j is regularly varying. □

Proof of Theorem 5

Let

$$
f(s) = \int_0^{\infty} z^{-1} e^{-\frac{s}{z}} \rho^{\star}(\mathrm{d}z).
$$

From the setting of the independent exponential multivariate score distribution it is straightforward to see that

$$\tilde{\rho}_d(\boldsymbol{s}) = \int_0^\infty z^{-d} h\left(\frac{s_1}{z}, \cdots, \frac{s_d}{z}\right) \rho^\star(\mathrm{d}z) = \int_0^\infty z^{-d} e^{-\frac{s_1+\ldots+s_d}{z}} \rho^\star(\mathrm{d}z).$$

From Example 8 we know that, for arbitrary $\rho^\star$ and $d \in \mathbb{N} \setminus \{0\}$, the previous integral is finite. Therefore for $s \neq 0$

$$\int_0^\infty \left| \frac{\partial^j}{\partial s^j} z^{-1} e^{-\frac{s}{z}} \right| \rho^\star(\mathrm{d}z) = \tilde{\rho}_{j+1}(s, 0, \ldots, 0) < \infty; \tag{2.20}$$

it follows using the Dominated Convergence Theorem that we can take the derivative under the integral sign as

$$\begin{aligned}
&(-1)^{d-1} \frac{\partial^{d-1}}{\partial s^{d-1}} f(s) \bigg|_{s=s_1+\ldots+s_d} \\
&= (-1)^{d-1} \int_0^\infty \frac{\partial^{d-1}}{\partial s^{d-1}} \left(z^{-1} e^{-\frac{s}{z}}\right) \rho^\star(\mathrm{d}z) \bigg|_{s=s_1+\ldots+s_d} \\
&= \int_0^\infty z^{-d} e^{-\frac{s_1+\ldots+s_d}{z}} \rho^\star(\mathrm{d}z).
\end{aligned} \tag{2.21}$$

Using (2.20) and (2.21) we conclude the proof.

Proof of Theorem 6

$$\begin{aligned}
&\psi_t(\lambda_1, \ldots, \lambda_d) = \int_0^t \int_0^\infty \cdots \int_0^\infty (1 - e^{-\lambda_1 u_1 - \ldots - \lambda_d u_d}) \nu(\mathrm{d}u_1, \ldots, \mathrm{d}u_d \mathrm{d}x) \\
&= \int_0^t \int_0^\infty \cdots \int_0^\infty (1 - e^{-\lambda_1 w_1 z - \ldots - \lambda_d w_d z}) h(w_1, \ldots, w_d) \mathrm{d}w_1 \ldots \mathrm{d}w_d \nu^\star(\mathrm{d}z, \mathrm{d}x) \\
&= \int_0^\infty \cdots \int_0^\infty \int_0^t \int_0^\infty (1 - e^{-(\lambda_1 w_1 - \ldots - \lambda_d w_d) z}) \nu^\star(\mathrm{d}z, \mathrm{d}x) h(w_1, \ldots, w_d) \mathrm{d}w_1 \ldots \mathrm{d}w_d \\
&= \mathbb{E}\left[\int_0^t \int_0^\infty (1 - e^{-(\lambda_1 W_1 - \ldots - \lambda_d W_d) z}) \nu^\star(\mathrm{d}z, \mathrm{d}x)\right] \\
&= \mathbb{E}[\psi_t^\star(\lambda_1 W_1 - \ldots - \lambda_d W_d)]
\end{aligned}$$

Proof of Theorem 7

Let $\mu^\star$ and H as in the hypothesis of the theorem and $\nu^\star$ the Lévy intensity of $\mu^\star$. We consider a PRM

$$M = \sum_{i=1}^{\infty} \delta_{(\mathbf{Z}_i, W_i, X_i)},$$

where $\{\mathbf{Z}_i\}_{i=1}^{\infty} \overset{\text{i.i.d.}}{\sim} H$ and $\{(W_i, X_i)\}_{i=1}^{\infty}$ are such that

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

as a PRM has intensity $\nu^\star$. It follows that M has intensity $\mu = H \times \nu^\star$. We define by

$$h(\mathbf{z}, w, x) = (z_1 w, z_2 w, \ldots, z_d w, x)$$

Then for $A_1, \ldots, A_d, B \in \mathscr{B}(\mathbb{R}^+)$

$$h^{-1}\left((A_1 \times \ldots \times A_d) \times B\right) =$$
$$\left\{(\frac{a_1}{w}, \frac{a_2}{w}, \ldots, \frac{a_d}{w}, w, x) \text{ such that } x \in B,\ a_1 \in A_1, \ldots a_d \in A_d,\ w \in \mathbb{R}^+\right\}$$

So the pullback measure $\eta = \mu \circ h^{-1}$ is given by

$$\begin{aligned}
\eta\left((A_1 \times \ldots \times \times A_d) \times B\right) &= \int_{h^{-1}((A_1 \times \ldots \times \times A_d) \times B)} \mathrm{d}\mu \\
&= \int_{A_1/z \times A_2/z \times \ldots \times A_d/z \times (0,\infty) \times B} H(\mathrm{d}s_1, \ldots, \mathrm{d}s_d) \nu^\star(\mathrm{d}z, \mathrm{d}x) \\
&= \int_{A_1 \times A_2 \times \ldots \times A_d \times (0,\infty) \times B} H\left(\frac{\mathrm{d}s_1}{z}, \ldots, \frac{\mathrm{d}s_d}{z}\right) \nu^\star(\mathrm{d}z, \mathrm{d}x)
\end{aligned}$$

So extending the measure we conclude that

$$N = \sum_{i=1}^{\infty} \delta_{(Z_{1,i} W_i, Z_{2,i} W_i, \ldots, Z_{d,i} W_i, X_i)}$$

is a CoRM given by the score distribution H and the directing Lévy measure $\nu^\star$ due to Proposition 1.

Chapter 3

Lévy copulas from compound random measures

3.1 Lévy copulas

A widely used approach for setting the dependence structure in a VCRM is the *Lévy copula* approach. Lévy copulas were proposed in Kallsen and Tankov (2006) and serve as an analogue of the distributional copulas. For Lévy copulas interest is placed on a multivariate Lévy intensity while distributional copulas interest is placed on a multivariate probability distribution. For a full review of Lévy copulas see Cont and Tankov (2004) and for a full review of distributional copulas see Nelsen (2007). In this chapter we discuss the link between a Lévy copula and a CoRM, exhibited in Theorem 9, and introduce a new class of Lévy copulas which generalize the widely used Clayton Lévy copula. These results are of interest in a Bayesian nonparametric context as we will see in Chapter 4. However, Lévy copulas are also of interest in the frequentist literature. In fact, they were first used in this framework for modelling dependent Lévy processes, see for instance Esmaeili and Klüppelberg (2010). We introduce some preliminary concepts for the discussion in this chapter. Let $d \in \mathbb{N}$, in the following we say for $\boldsymbol{a} = (a_1, \dots, a_d)$ and $\boldsymbol{b} = (b_1, \dots, b_d)$ both in $(\mathbb{R}^+)^d$ that a $d-$box denoted $[\boldsymbol{a}, \boldsymbol{b}]$ is given by

$$[\boldsymbol{a}, \boldsymbol{b}] = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_d, b_d].$$

We say that the vertices of a $d-$box $[\boldsymbol{a},\boldsymbol{b}]$ are the points $\boldsymbol{c}=(c_1,\dots,c_d)$ such that $c_k\in[a_k,b_k]$ for $k\in\{1,\dots,d\}$. We also define the function

$$\operatorname{sgn}(\boldsymbol{c})=\begin{cases}1, & \text{if } c_k=a_k \text{ for an even number of } k\text{'s}.\\ -1, & \text{if } c_k=a_k \text{ for an odd number of } k\text{'s}.\end{cases}$$

We need the next two definitions for the setting up of the Lévy copula concept.

Definition 15. Let f be a d-variate function $\operatorname{Dom}(f)\subset(\mathbb{R}^+)^d$ and $[\boldsymbol{a},\boldsymbol{b}]$ with all its vertices in $\operatorname{Dom}(f)$; we say that the $f-$volume of $[\boldsymbol{a},\boldsymbol{b}]$ is

$$V_f([\boldsymbol{a},\boldsymbol{b}])=\sum_{\{\boldsymbol{c}:\boldsymbol{c}\text{ is a vertex of }[\boldsymbol{a},\boldsymbol{b}]\}}\operatorname{sgn}(\boldsymbol{c})f(\boldsymbol{c})$$

Definition 16. We say that a d-variate function f with domain $\operatorname{Dom}(f)\subset(\mathbb{R}^+)^d$ is d-increasing in $(\mathbb{R}^+)^d$ if $V_f([\boldsymbol{a},\boldsymbol{b}])>0$ for all $d-$boxes $[\boldsymbol{a},\boldsymbol{b}]$ with all their vertices in $\operatorname{Dom}(f)$.

More precisely a Lévy copula is defined as follows.

Definition 17. A $d-$variate positive Lévy copula is a function $\mathscr{C}(s_1,\dots,s_d):[0,\infty]^d\to[0,\infty]$ which satisfies

1. $\mathscr{C}(s_1,\dots,s_d)<\infty$ for $(s_1,\dots,s_d)\neq(\infty,\dots,\infty)$.
2. $\mathscr{C}$ is $d-$increasing.
3. $\mathscr{C}(s_1,\dots,s_d)=0$ if $u_k=0$ for any $k\in\{1,\dots,d\}$
4. $\mathscr{C}_k(s)=\mathscr{C}(y_1^{(k)},\dots,y_{k-1}^{(k)},s,y_{k+1},\dots,y_d^{(k)})=s$ for $k\in\{1,\dots,d\}$, $s\in\mathbb{R}^+$, where $y_1^{(k)}=\dots=y_{k-1}^{(k)}=y_{k+1}^{(k)}=\dots=y_d^{(k)}=\infty$.

The relation between the $d-$variate tail integral, see Definition 5, the marginal tail integrals, see equation (1.19), and the Lévy copula is made explicit in the next result, which can be seen as the Lévy copula analogue of the Sklar Theorem for distributional copulas.

Theorem 8. *(Cont and Tankov (2004)) Let U be a d-variate tail integral with margins $\{U_i\}_{i=1}^d$ then there exists a Lévy copula $\mathscr{C}$ such that*

$$U(s_1,\dots,s_d)=\mathscr{C}(U_1(s_1),\dots,U_d(s_d))$$

If $\{U_i\}_{i=1}^d$ are continuous $\mathscr{C}$ is unique, otherwise it is unique in $Ran(U_1)\times\dots\times Ran(U_d)$.

For a proof see Cont and Tankov (2004) where a full review of Lévy copulas and their link to Lévy processes is given. If the Lévy copula is smooth then from Theorem 8 and the definition of the multivariate tail integral we have that the underlying multivariate Lévy intensity can be expressed as

$$\tilde{\rho}_d(\boldsymbol{s}) = \frac{\partial^d}{\partial u_1 \cdots \partial u_d} \mathscr{C}(\boldsymbol{u}) \Big|_{u_1=U_1(s_1),\cdots,x_d=U_d(s_d)} \rho_1(s_1) \cdots \rho_d(s_d), \tag{3.1}$$

where ρ_i, $i \in \{1, \ldots, d\}$, are the corresponding marginal tail integrals. Furthermore if $\mathscr{C}$ is a two dimensional Lévy copula and $\{(W_{1,i}, W_{2,i})\}_{i=1}^{\infty}$ are the random weights of a series representation for the associated CRM, equation (1.11), then the law of $S_{1,i} = U_1(W_{1,i})$ conditioned on $S_{2,i} = U_2(W_{2,i}) = s_2 \in \mathbb{R}^+ \setminus \{0\}$ is given by the distribution function

$$\hat{F}_{S_1|S_2=s_1}(s_1) = \frac{\partial}{\partial s_2} \mathscr{C}(s_1, s_2) \tag{3.2}$$

and the law of $S_{2,i} = U_2(W_{2,i})$ conditioned on $S_{1,i} = U_1(W_{1,i}) = s_1 \in \mathbb{R}^+ \setminus \{0\}$ is given by the distribution function

$$\hat{F}_{S_2|S_1=s_1}(s_2) = \frac{\partial}{\partial s_1} \mathscr{C}(s_1, s_2); \tag{3.3}$$

see Theorem 6.3 in Cont and Tankov (2004) for a proof.

Some examples of d-variate positive Lévy copulas are the following:

Example 12. Independence Lévy copula.

$$\mathscr{C}_{\perp}(s_1, \ldots, s_d) = \sum_{i=1}^{d} s_i \prod_{j \neq i} \mathbb{1}_{\{s_j = \infty\}}.$$

In this case the random measures $\{\mu_i\}_{i=1}^{d}$ are pairwise independent.

Example 13. Complete dependence Lévy copula.

$$\mathscr{C}_{||}(s_1, , \ldots, s_d) = \min\{s_1, \ldots, s_d\}.$$

In this case the random measures $\{\mu_i\}_{i=1}^{d}$ are completely dependent in the sense that the jumps weights of the VCRM associated to each location, the vectors $\{(W_{1,i}, \ldots, W_{d,i})\}_{i=1}^{\infty}$ as in (1.11), are in a set S such that whenever $\boldsymbol{v}, \boldsymbol{u} \in S$ then either $v_j < u_j$ or $u_j < v_j$ for all $j \in \{1, \ldots, d\}$.

Example 14. Clayton Lévy copula.

$$\mathscr{C}_\theta(s_1,\ldots,s_d) = \left(s_1^{-\theta} + \ldots + s_d^{-\theta}\right)^{-1/\theta}; \quad \theta > 0.$$

The Clayton example above is of great interest as its parameter θ enable us to modulate between the independence and complete dependence cases; indeed

$$\lim_{\theta\to 0} \mathscr{C}_\theta(s_1,\ldots,s_d) = \mathscr{C}_\perp(s_1,\ldots,s_d)$$

and

$$\lim_{\theta\to \infty} \mathscr{C}_\theta(s_1,\ldots,s_d) = \mathscr{C}_{||}(s_1,,\ldots,s_d).$$

Lévy copulas are useful to construct VCRM's in such a way that the marginal behaviour can be fixed and the dependence structure can be modeled separately, see for example Grothe and Nicklas (2013) Leisen and Lijoi (2011), Leisen et al. (2013) and Zhu and Leisen (2015).

Example 15. Clayton Lévy copula with σ-stable marginals.

We focus on an homogeneous VCRM with dependence in the weights given by the Clayton Lévy copula and with σ-stable margins, Example 1 with $\kappa(\mathrm{d}x) = \mathrm{d}x$. If we consider the Lévy intensity arising from (3.1) when considering the d-dimensional Clayton Lévy copula, above, and σ-stable marginals with the same parameter, (1.7), we obtain

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

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

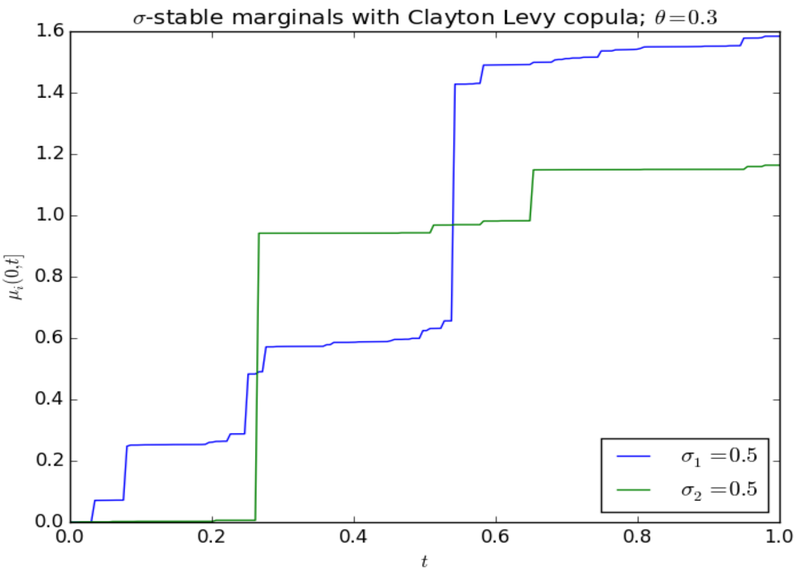
$$\tilde{\rho}_{d,A,\sigma}(\boldsymbol{s}) = \frac{A(\sigma+1)(\sigma+2)\cdots(\sigma+d-1)\sigma}{\Gamma(1-\sigma)\,(s_1+\cdots+s_d)^{\sigma+d}}. \tag{3.4}$$

Such intensity corresponds as well to a CoRM with Gamma$(1,1)$ score distribution, Example 2, which is restrained to have Gamma marginals; this example arises when taking $\phi = 1$ in equation (4.4) of Griffin and Leisen (2017). 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

$$\psi_{d,A,\sigma}(\boldsymbol{\lambda}) = \sum_{i=1}^{d} \frac{\lambda_i^{\sigma+d-1}}{\prod_{j=1,\,j\neq i}^{d}(\lambda_i-\lambda_j)}; \quad \lambda_i \neq \lambda_j \text{ for } j\neq i, \tag{3.5}$$

where we take the appropriate limits when $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_d)$ is such that $\lambda_i = \lambda_j$ for distinct $i, j \in \{1, \dots, d\}$.

The Lévy copula approach is also convenient for the proposal of general simulation schemes for VCRM's. We can use the identities (3.2) and (3.3) to construct a simulation algorithm as follows, see Cont and Tankov (2004).



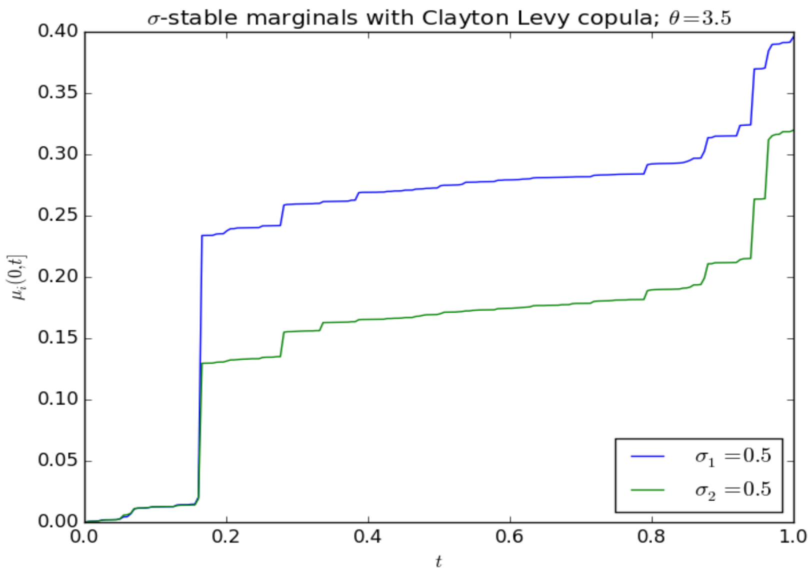


Fig. 3.1 Plots of VCRM's given by a Clayton Lévy copula and σ-stable margins; simulation was peformed using Algorithm 4 with Algorithm 1 for the marginal CRM simulation.

Algorithm 4 Lévy copula based simulation for two dimensional VCRM

1: Draw a marginal CRM.
2: Use the conditional distributions (3.2) or (3.3) to draw the weights of the remaining CRM weights given the weights of the CRM drawn in the last step.

Illustration of the dependence structure given by a Clayton Lévy copula is presented in Figure 3.1. As expected, when $\theta = 0.3$, we are close to independent behaviour. 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.

3.2 A new class of Lévy copulas from CoRM's

Griffin and Leisen (2017) highlighted the Lévy copula structure of a CoRM when the score distribution has independent and identically distributed marginal distributions; further exploration of the Lévy copulas corresponding to a CoRM was not performed. In a more general setting where the score distribution has an arbitrary d-variate density function h which we determine by its associated distributional survival Copula $\hat{C}$ and marginal survival functions $S_1,\dots,S_d$, see Nelsen (2007) Section 2.6, we have the next result to recover the Lévy Copula

Theorem 9. *Let $\boldsymbol{\mu}$ be a CoRM given by a directing Lévy measure $\nu^\star$ and a score distribution with distributional survival Copula $\hat{C}$ and marginal survival functions $S_1,\dots,S_d$, then the Lévy copula, $\mathscr{C}$, associated to $\boldsymbol{\mu}$ is given by*

$$\mathscr{C}(s_1,\dots,s_d) = \int_0^\infty \hat{C}\left(S_1\left(\frac{U_1^{-1}(s_1)}{z}\right),\cdots,S_d\left(\frac{U_2^{-1}(s_2)}{z}\right)\right)\rho^\star(\mathrm{d}z)$$

where the marginal tail integrals U_i can be expressed as

$$U_i(x) = \int_0^\infty S_i\left(\frac{x}{z}\right)\rho^\star(\mathrm{d}z)$$

for $i \in \{1,\dots d\}$.

The above result can be used to propose new families of Lévy copulas which arise from a CoRM. With the aim of proposing a new family of Lévy copulas we focus on the next family of bivariate CoRM's which was previously studied in Griffin and Leisen (2017).

Example 16. If $\sigma \in (0,1)$, $\phi > 0$ and

- $h(y_1,y_2) = \frac{(y_1y_2)^{\phi-1}e^{-y_1-y_2}}{\Gamma(\phi)^2}$
- $\rho^\star(z) = \frac{z^{-\sigma-1}\Gamma(\phi)}{\Gamma(\phi+\sigma)\Gamma(1-\sigma)}$

Then by Theorem 4.1 in Griffin and Leisen (2017) the corresponding bivariate CoRM has $\sigma-$stable marginals and Lévy intensity given by

$$\rho_{\sigma,\phi}(\mathrm{d}s_1,\mathrm{d}s_2) = \frac{\sigma(s_1s_2)^{\phi-1}\Gamma(\sigma+2\phi)(s_1+s_2)^{-\sigma-2\phi}}{\Gamma(\phi)\Gamma(\sigma+\phi)\Gamma(1-\sigma)}\mathrm{d}s_1\mathrm{d}s_2 \tag{3.6}$$

We show the Lévy copula associated to the above CoRM in the next result.

Theorem 10. *Let* $\sigma \in (0,1)$ *and* $\phi > 0$*. We set*

$$\begin{aligned}\mathscr{C}_{\sigma,\phi}(s_1,s_2) &= \frac{\sigma\Gamma(\sigma+2\phi)(s_1^{-\frac{1}{\sigma}}+s_2^{-\frac{1}{\sigma}})^{-\sigma}}{2\Gamma(\phi)\Gamma(\sigma+\phi)} \\ &\times \sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\binom{\phi-1}{k}\binom{\phi+k-1}{j}(-1)^{k+j}\frac{\left(s_1^{-\frac{j}{\sigma}}+s_2^{-\frac{j}{\sigma}}\right)\left(s_1^{-\frac{1}{\sigma}}+s_2^{-\frac{1}{\sigma}}\right)^{-j}}{(\sigma+j)(\sigma+\phi+k)}.\end{aligned} \tag{3.7}$$

Then the Lévy Copula associated to $\rho_{\sigma,\phi}$*, in* (3.6)*, is* $\mathscr{C}_{\sigma,\phi}$*.*

For $\phi \in \mathbb{N}\setminus\{0\}$ we observe that $\mathscr{C}_{\sigma,\phi}$ reduces to

$$\begin{aligned}\mathscr{C}_{\sigma,\phi}(s_1,s_2) &= \frac{\sigma\Gamma(\sigma+2\phi)(s_1^{-\frac{1}{\sigma}}+s_2^{-\frac{1}{\sigma}})^{-\sigma}}{\Gamma(\phi)\Gamma(\sigma+\phi)} \\ &\times \sum_{k=0}^{\phi-1}\binom{\phi-1}{k}\sum_{j=0}^{\phi+k-1}\binom{\phi+k-1}{j}(-1)^{k+j}\frac{\left(s_1^{-\frac{j}{\sigma}}+s_2^{-\frac{j}{\sigma}}\right)\left(s_1^{-\frac{1}{\sigma}}+s_2^{-\frac{1}{\sigma}}\right)^{-j}}{(\sigma+j)(\sigma+\phi+k)}\end{aligned}$$

Furthermore, we observe that under the reparametrization $\theta = \frac{1}{\sigma}$ the Lévy copula (3.7) has the Clayton Lévy copula as a factor. Although, from Theorem 10 we see that $\theta \in (1,\infty)$ as $\rho_{\sigma,\phi}$ was only defined for $\sigma \in (0,1)$. Surprisingly, this Lévy copula can be extended for $\theta \in (0,1]$ as showcased in the next theorem.

Theorem 11. *Let $\phi \in \mathbb{N}$ and $\theta \in (0,\infty)$, then*

$$\mathscr{C}_{\theta,\phi}(s_1,s_2) = \frac{\Gamma(\frac{1}{\theta}+2\phi)(s_1^{-\theta}+s_2^{-\theta})^{-\frac{1}{\theta}}}{2\theta\Gamma(\phi)\Gamma(\frac{1}{\theta}+\phi)}$$
$$\times \sum_{k=0}^{\phi-1} \binom{\phi-1}{k} \sum_{j=0}^{\phi+k-1} \binom{\phi+k-1}{j} (-1)^{k+j} \frac{\left(s_1^{-j\theta}+s_2^{-j\theta}\right)\left(s_1^{-\theta}+s_2^{-\theta}\right)^{-j}}{(\sigma+j)(\sigma+\phi+k)}$$

is a Lévy copula.

3.3 Future work

We follow Esmaeili and Klüppelberg (2010) to perform parameter estimation of a bivariate compound Poisson process. We will restrict ourselves to compound Poisson processes with positive increments.

Definition 18. Given $\lambda_1 \lambda_2 \in \mathbb{R}^+ \setminus \{0\}$ and probability distributions F_1, F_2 in $\mathbb{R}^+$, a bivariate compound Poisson process with positive increments is a bivariate vector of stochastic processes (X_1, X_2) given by

$$X_i(t) = \mu_i(0,t]$$

for $t \in \mathbb{R}^+$ and $\boldsymbol{\mu} = (\mu_1, \mu_2)$ a bivariate VCRM such that marginally μ_i has Lévy intensity

$$\nu_i(\mathrm{d}s\mathrm{d}x) = \lambda_i F_i(\mathrm{d}s)\mathrm{d}x.$$

We observe that the associated marginal tail integrals are bounded in $\mathbb{R}^+$ so almost surely the associated series representation has finite jumps. We will focus on bivariate compound processes of the form

$$(X_1(t), X_2(t)) = \left(\sum_{i=1}^{N_1(t)} W_{1,i} \,,\, \sum_{i=1}^{N_2(t)} W_{2,i} \right)$$
$$= \left(\sum_{i=1}^{N_1^{\perp}(t)} W_{1,i}^{\perp} + \sum_{i=1}^{N^{\parallel}(t)} W_{1,i}^{\parallel} \,,\, \sum_{i=1}^{N_2^{\perp}(t)} W_{2,i}^{\perp} + \sum_{i=1}^{N^{\parallel}(t)} W_{2,i}^{\parallel} \right)$$

where $N_1(t)$, $N_2(t)$, $N_1^{\perp}(t)$, $N_2^{\perp}(t)$ and $N^{\parallel}(t)$ can be seen as PRM's in $\mathbb{R}^+$, evaluated in $(0,t]$, see Definition 2. We observe that in each vector component the first sum is related to independent weights while the second sum in each component is related to dependent weights, we call the former the independent part and the latter the dependent part. The dependent part

$$\left(\sum_{i=1}^{N^{\|}(t)} W_{1,i}^{\|}, \sum_{i=1}^{N^{\|}(t)} W_{2,i}^{\|} \right)$$

can be modelled by a Lévy copula with conditional distributions in the weights (3.2) and (3.3) such that they do not assign a point mass at zero; which is the case for $\mathscr{C}_{\theta,\phi}$. For a full review of Poisson processes we refer to Kingman (2005). We will assume the next observation scheme for bivariate compound Poisson processes.

Definition 19. We say that we observe the bivariate compound process continuously through time if we are able to observe all the jump times and jump weights in a given time interval.

Let $\{(w_{1,i}, w_{2,i})\}_{i=1}^{n}$ be the jump sizes of a continuously observed bivariate bivariate compound Poisson process, with $n \in \mathbb{N}$ the number of jumps. We denote

$$\{w_{1,i}^{\perp}\}_{i=1}^{n_1^{\perp}} = \{(w_1, w_2) \ : \ w_{1,i} \neq 0\,,\, w_{2,i} = 0\}, \quad \{w_{2,i}^{\perp}\}_{i=1}^{n_2^{\perp}} = \{(w_1, w_2) \ : \ w_{1,i} = 0\ ,\, w_{2,i} \neq 0\}$$

and

$$\{(w_{1,i}^{\|}, w_{2,i}^{\|})\}_{i=1}^{n^{\|}} = \{(w_1, w_2) \ : \ w_{1,i} \neq 0\,,\, w_{2,i} \neq 0\}$$

with $n_1^{\perp}, n_2^{\perp}, n^{\|} \in \mathbb{N}$, respectively, the number of jumps in only the first dimension, number of jumps in only the second dimension and number of jumps in both dimensions. With the above notation we can give the likelihood for the continuous through time observations.

Proposition 7 (Esmaeili and Klüppelberg (2010)). *Let $T \in \mathbb{R}^{+}$ if a bivariate compound Poisson process has jump rates λ_1, λ_2 and dependent part modelled by a Lévy copula $\mathscr{C}_{\boldsymbol{c}}$ parametrized by a real valued vector $\boldsymbol{c}$ such that $\frac{\partial^2}{\partial u_1 \partial u_2}\mathscr{C}$ exists for every $(u_1, u_2) \in (0, \lambda_1) \times (0, \lambda_2)$, then setting $\lambda^{\|} = \mathscr{C}_{\boldsymbol{c}}(\lambda_1, \lambda_2)$, $\lambda_i^{\perp} = \lambda_i - \lambda^{\|}$, marginal jump weight distributions F_i associated to survival functions S_i and probability densities f_i parametrized by real valued vector $\boldsymbol{\alpha}_i$, $i \in \{1,2\}$, the likelihood function for continuously observed bivariate compound*

Poisson processes in $(0,t]$ *is given by*

$$
\begin{aligned}
L(\lambda_1,\lambda_2,\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{c}) = \lambda_1^{n_1^{\perp}} e^{-\lambda_1^{\perp}T} \prod_{i=1}^{n_1^{\perp}} \left(f_1(w_{1,i}^{\perp};\boldsymbol{\alpha}_1) \left(1 - \frac{\partial}{\partial u_1}\mathscr{C}_{\boldsymbol{c}}(u_1,\lambda_2)\Big|_{u_1=\lambda_1 S_1(w_{1,i}^{\perp};\boldsymbol{\alpha}_1)}\right)\right) \\
\times \lambda_2^{n_2^{\perp}} e^{-\lambda_2^{\perp}T} \prod_{i=1}^{n_2^{\perp}} \left(f_2(w_{2,i}^{\perp};\boldsymbol{\alpha}_2) \left(1 - \frac{\partial}{\partial u_2}\mathscr{C}_{\boldsymbol{c}}(\lambda_1,u_2)\Big|_{u_2=\lambda_2 S_2(w_{2,i}^{\perp};\boldsymbol{\alpha}_2)}\right)\right) \\
\times (\lambda_1\lambda_2)^{n^{\parallel}} e^{-\lambda^{\parallel}T} \prod_{i=1}^{n^{\parallel}} \Big(f_1(w_{1,i}^{\parallel};\boldsymbol{\alpha}_1) f_2(w_{2,i}^{\parallel};\boldsymbol{\alpha}_2) \\
\times \frac{\partial^2}{\partial u_1 \partial u_2}\mathscr{C}_{\boldsymbol{c}}(u_1,u_2)\Big|_{u_1=\lambda_1 S_1(w_{1,i}^{\parallel};\boldsymbol{\alpha}_1),\, u_2=\lambda_2 S_2(w_{2,i}^{\parallel};\boldsymbol{\alpha}_2)}\Big)
\end{aligned}
$$

The application of our extension of the Clayton Lévy copula $\mathscr{C}_{\theta,,\phi}$ is of interest for the above model as it can offer more flexibility in the likelihood above. Esmaeili and Klüppelberg (2010) performed a real data analysis of the Danish fire insurance dataset. With the same data set and the use of our Copula we have numerically found values of δ for each $2 \leq \phi \leq 12$ which attain a higher likelihood value than the maximum likelihood restrained to $\phi = 1$, i.e. the usual Clayton Lévy copula. However numerical issues arise in the maximization routine as the parameter ϕ grows. We plan to keep working on the use of our new copula for bivariate compound Poisson process modelling

3.4 Chapter 3 proofs

Proof of Theorem 9

By definition

$$
\begin{aligned}
U(s_1,\dots,s_d) &= \int_0^{\infty}\int_{s_1}^{\infty}\cdots\int_{s_d}^{\infty} H\left(\frac{\mathrm{d}u_1}{z},\dots,\frac{\mathrm{d}u_d}{z}\right)\rho^{\star}(\mathrm{d}z) \\
&= \int_0^{\infty}\int_{\frac{s_1}{z}}^{\infty}\cdots\int_{\frac{s_d}{z}}^{\infty} H(\mathrm{d}u_1,\dots,\mathrm{d}u_d)\,\rho^{\star}(\mathrm{d}z) \\
&= \int_0^{\infty} S\left(\frac{s_1}{z},\dots,\frac{s_d}{z}\right)\rho^{\star}(\mathrm{d}z) \\
&= \int_0^{\infty} \hat{C}\left(S_1\left(\frac{s_1}{z}\right),\dots,S_d\left(\frac{s_d}{z}\right)\right)\rho^{\star}(\mathrm{d}z).
\end{aligned}
$$

Where in the last equation we have used the Sklar theorem for survival copulas

$$S(u_1,\dots,u_d) = \hat{C}\left(S_1(u_1),\dots S_d(u_d)\right)$$

From the Sklar theorem for Lévy copulas, Theorem 8, we conclude the proof.

Proof of Theorem 10

Let U be the bivariate tail integral of $\rho_{\sigma,\phi}$ as in the hypothesis.

$$U(s_1,s_2) = \int_{s_1}^{\infty}\int_{s_2}^{\infty} \frac{\sigma(y_1y_2)^{\phi-1}\Gamma(\sigma+2\phi)(y_1+y_2)^{-\sigma-2\phi}}{\Gamma(\phi)\Gamma(\sigma+\phi)\Gamma(1-\sigma)}\mathrm{d}y_1\mathrm{d}y_2$$

We consider the change of variable

$$\begin{aligned}\boldsymbol{h}(y_1,y_2) &= (y_1+y_2, y_1/(y_1+y_2)) = (\rho, z_1)\\ \mathrm{d}\rho\mathrm{d}z_1 &= \left|\det(\frac{\mathrm{d}\boldsymbol{h}}{\mathrm{d}\boldsymbol{y}})\right|\mathrm{d}y_1\mathrm{d}y_2 = (y_1+y_2)^{-1}\mathrm{d}y_1\mathrm{d}y_2\end{aligned}$$

so

$$\begin{aligned}U(s_1,s_2) &= \frac{\sigma\Gamma(\sigma+2\phi)}{\Gamma(\phi)\Gamma(\sigma+\phi)\Gamma(1-\sigma)}\\ &\times \int_{h(\{y_1,y_2:s_1\le y_1,s_2\le y_2\})}(z_1-z_1^2)^{\phi-1}\rho^{-\sigma-1}\mathrm{d}\rho\mathrm{d}z_1.\end{aligned}$$

For notation purposes let for the rest of this proof

$$c_{\sigma,\phi} = \frac{\sigma\Gamma(\sigma+2\phi)}{\Gamma(\phi)\Gamma(\sigma+\phi)\Gamma(1-\sigma)}$$

For the region of integration we consider the curves

$$\begin{aligned}\hat{\omega}(\hat{t}) &= h(s_1,s_2+\hat{t}) = (s_1+s_2+\hat{t}, s_1/(s_1+s_2+\hat{t}))\\ \hat{\gamma}(\hat{t}) &= h(s_1+\hat{t},s_2) = (s_1+s_2+\hat{t}, (s_1+\hat{t})/(s_1+s_2+\hat{t}))\end{aligned}$$

with $\hat{t} \geq 0$; so for $t \geq s_1 + s_2$ we can get the reparametrized curves $\omega(t) = (t, s_1/t)$ and $\gamma(t) = (t, 1 - s_2/t)$ to delimit the integration area, hence using Fubini theorem

$$
\begin{aligned}
U(s_1,s_2) &= c_{\sigma,\phi} \int_{s_1+s_2}^{\infty} \int_{s_1/\rho}^{1-s_2/\rho} (z_1 - z_1^2)^{\phi-1} \rho^{-\sigma-1} \mathrm{d}z_1 \mathrm{d}\rho \\
&= c_{\sigma,\phi} \int_{s_1+s_2}^{\infty} \int_{s_1/\rho}^{1-s_2/\rho} (z_1 - z_1^2)^{\phi-1} \rho^{-\sigma-1} \mathrm{d}z_1 \mathrm{d}\rho \\
&= c_{\sigma,\phi} \int_{s_1+s_2}^{\infty} \rho^{-\sigma-1} \int_{s_1/\rho}^{1-s_2/\rho} \sum_{k=0}^{\infty} \binom{\phi-1}{k} (-1)^k z_1^{\phi-1+k} \mathrm{d}z_1 \mathrm{d}\rho \\
&\overset{\text{Fubini}}{=} c_{\sigma,\phi} \sum_{k=0}^{\infty} \int_{s_1+s_2}^{\infty} \rho^{-\sigma-1} \binom{\phi-1}{k} (-1)^k \left. \frac{z_1^{\phi+k}}{\phi+k} \right|_{s_1/\rho}^{1-s_2/\rho} \mathrm{d}\rho \\
&= c_{\sigma,\phi} \sum_{k=0}^{\infty} \int_{s_1+s_2}^{\infty} \rho^{-\sigma-1} \binom{\phi-1}{k} (-1)^k \left[\frac{(1-\frac{s_2}{\rho})^{\phi+k} - (\frac{s_1}{\rho})^{\phi+k}}{\phi+k} \right] \mathrm{d}\rho \\
&= c_{\sigma,\phi} \sum_{k=0}^{\infty} \binom{\phi-1}{k} \frac{(-1)^k}{\phi+k} \int_{s_1+s_2}^{\infty} \left(\rho^{-\sigma-1} (1-\frac{s_2}{\rho})^{\phi+k} - s_1^{\phi+k} \rho^{-\sigma-1-\phi-k} \right) \mathrm{d}\rho \\
&= c_{\sigma,\phi} \sum_{k=0}^{\infty} \binom{\phi-1}{k} \frac{(-1)^k}{\phi+k} \int_{s_1+s_2}^{\infty} \left(\rho^{-\sigma-1} \sum_{j=0}^{\infty} \binom{\phi+k}{j} (-1)^j \left(\frac{s_2}{\rho}\right)^j \right. \\
&\quad \left. - s_1^{\phi+k} \rho^{-\sigma-1-\phi-k} \right) \mathrm{d}\rho \\
&= c_{\sigma,\phi} \sum_{k=0}^{\infty} \binom{\phi-1}{k} \frac{(-1)^k}{\phi+k} \left(\sum_{j=0}^{\infty} \binom{\phi+k}{j} (-1)^j s_2^j \int_{s_1+s_2}^{\infty} \rho^{-\sigma-1-j} \mathrm{d}\rho \right. \\
&\quad \left. - s_1^{\phi+k} \int_{s_1+s_2}^{\infty} \rho^{-\sigma-1-\phi-k} \mathrm{d}\rho \right) \\
&= c_{\sigma,\phi} \sum_{k=0}^{\infty} \binom{\phi-1}{k} \frac{(-1)^k}{\phi+k} \left[\sum_{j=0}^{\infty} \binom{\phi+k}{j} (-1)^{j+1} s_2^j \left(\left. \frac{\rho^{-\sigma-j}}{(\sigma+j)} \right|_{s_1+s_2}^{\infty} \right) \right. \\
&\quad \left. + s_1^{\phi+k} \left(\left. \frac{\rho^{-\sigma-\phi-k}}{(\sigma+\phi+k)} \right|_{s_1+s_2}^{\infty} \right) \right] \\
&= c_{\sigma,\phi} \sum_{k=0}^{\infty} \binom{\phi-1}{k} \frac{(-1)^k}{\phi+k} \left[\sum_{j=0}^{\infty} \binom{\phi+k}{j} (-1)^j s_2^j \frac{(s_1+s_2)^{-\sigma-j}}{(\sigma+j)} \right. \\
&\quad \left. - s_1^{\phi+k} \frac{(s_1+s_2)^{-\sigma-\phi-k}}{(\sigma+\phi+k)} \right]
\end{aligned}
$$

$$= c_{\sigma,\phi}\sum_{k=0}^{\infty}\binom{\phi-1}{k}\frac{(-1)^k}{\phi+k}\left(\sum_{j=0}^{\infty}\binom{\phi+k}{j}(-1)^j(s_1+s_2)^{-\sigma}\frac{(\frac{s_2}{s_1+s_2})^j}{(\sigma+j)}\right.$$
$$\left.-\frac{(s_1+s_2)^{-\sigma}(1-\frac{s_2}{s_1+s_2})^{\phi+k}}{(\sigma+\phi+k)}\right)$$
$$= c_{\sigma,\phi}\sum_{k=0}^{\infty}\binom{\phi-1}{k}\frac{(-1)^k}{\phi+k}\left(\sum_{j=0}^{\infty}\binom{\phi+k}{j}(-1)^j(s_1+s_2)^{-\sigma}\frac{(\frac{s_2}{s_1+s_2})^j}{(\sigma+j)}\right.$$
$$\left.-\sum_{j=0}^{\infty}\binom{\phi+k}{j}(-1)^j(s_1+s_2)^{-\sigma}\frac{(\frac{s_2}{s_1+s_2})^j}{(\sigma+\phi+k)}\right)$$
$$= c_{\sigma,\phi}\sum_{k=0}^{\infty}\binom{\phi-1}{k}\frac{(-1)^k}{\phi+k}\left(\sum_{j=0}^{\infty}\binom{\phi+k}{j}(-1)^j\right.$$
$$\left.(s_1+s_2)^{-\sigma}\frac{(\frac{s_2}{s_1+s_2})^j(\phi+k-j)}{(\sigma+j)(\sigma+\phi+k)}\right)$$

To get the copula we evaluate the above tail integral in

$$\left(U^{-1}(s_1),U^{-1}(s_2)\right)=\left((\Gamma(1-\sigma)s_1)^{-\frac{1}{\sigma}},(\Gamma(1-\sigma)s_2)^{-\frac{1}{\sigma}}\right);$$

entailing the associated Lévy copula

$$\mathscr{C}_{\sigma,\phi}(s_1,s_2)=\frac{\sigma\Gamma(\sigma+2\phi)(s_1^{-\frac{1}{\sigma}}+s_2^{-\frac{1}{\sigma}})^{-\sigma}}{\Gamma(\phi)\Gamma(\sigma+\phi)}$$
$$\sum_{k=0}^{\phi-1}\binom{\phi-1}{k}\sum_{j=0}^{\phi+k-1}\binom{\phi+k-1}{j}(-1)^{k+j}\frac{\left(\frac{s_2^{-\frac{1}{\sigma}}}{s_1^{-\frac{1}{\sigma}}+s_2^{-\frac{1}{\sigma}}}\right)^j}{(\sigma+j)(\sigma+\phi+k)}.$$

Exploiting that by construction $\mathscr{C}_{\sigma,\phi}$ is symmetric, we get

$$\mathscr{C}_{\sigma,\phi}(s_1,s_2)=\frac{\sigma\Gamma(\sigma+2\phi)(s_1^{-\frac{1}{\sigma}}+s_2^{-\frac{1}{\sigma}})^{-\sigma}}{2\Gamma(\phi)\Gamma(\sigma+\phi)}$$
$$\sum_{k=0}^{\phi-1}\binom{\phi-1}{k}\sum_{j=0}^{\phi+k-1}\binom{\phi+k-1}{j}(-1)^{k+j}\frac{\left(s_1^{-\frac{j}{\sigma}}+s_2^{-\frac{j}{\sigma}}\right)\left(s_1^{-\frac{1}{\sigma}}+s_2^{-\frac{1}{\sigma}}\right)^{-j}}{(\sigma+j)(\sigma+\phi+k)}.$$

Proof of Theorem 11

From equation (3.1) we have that

$$\frac{\rho_{\sigma,\phi}\left(U^{-1}_{\sigma-\text{stab}}(s_1),U^{-1}_{\sigma-\text{stab}}(s_2)\right)}{\rho^{\sigma-\text{stab}}\left(U^{-1}_{\sigma-\text{stab}}(s_1)\right)\rho^{\sigma-\text{stab}}\left(U^{-1}_{\sigma-\text{stab}}(s_2)\right)}=\frac{\partial}{\partial s_2}\frac{\partial}{\partial s_1}\mathscr{C}_{\sigma,\phi}(s_1,s_2),$$

where $\rho^{\sigma-\text{stab}}$ is the Lévy intensity of an homogeneous σ-stable CRM and $U^{-1}_{\sigma-\text{stab}}$ its corresponding generalized inverse tail integral, see Example 1. It follows that

$$\begin{aligned}
&\frac{\partial}{\partial s_2}\frac{\partial}{\partial s_1}\mathscr{C}_{\sigma,\phi}(s_1,s_2)=\\
&\frac{\sigma\left(U^{-1}_{\sigma-\text{stab}}(s_1)U^{-1}_{\sigma-\text{stab}}(s_2)\right)^{\phi-1}\Gamma(\sigma+2\phi)\left(U^{-1}_{\sigma-\text{stab}}(s_1)+U^{-1}_{\sigma-\text{stab}}(s_2)\right)^{-\sigma-2\phi}}{\Gamma(\phi)\Gamma(\sigma+\phi)\Gamma(1-\sigma)\rho^{\sigma-\text{stab}}\left(U^{-1}_{\sigma-\text{stab}}(s_1)\right)\rho^{\sigma-\text{stab}}\left(U^{-1}_{\sigma-\text{stab}}(s_2)\right)}\\
&=\frac{\sigma\left((\Gamma(1-\sigma))^{-\frac{2}{\sigma}}s_1^{-\frac{1}{\sigma}}s_2^{-\frac{1}{\sigma}}\right)^{\phi-1}\Gamma(\sigma+2\phi)\left(s_1^{-\frac{1}{\sigma}}+s_2^{-\frac{1}{\sigma}}\right)^{-\sigma-2\phi}(\Gamma(1-\sigma))^{1+\frac{2\phi}{\sigma}}}{\Gamma(\phi)\Gamma(\sigma+\phi)\Gamma(1-\sigma)\rho^{\sigma-\text{stab}}\left(U^{-1}_{\sigma-\text{stab}}(s_1)\right)\rho^{\sigma-\text{stab}}\left(U^{-1}_{\sigma-\text{stab}}(s_2)\right)}
\end{aligned}$$

$$\begin{aligned}
&=\frac{\sigma\left(s_1^{-\frac{1}{\sigma}}s_2^{-\frac{1}{\sigma}}\right)^{\phi-1}\Gamma(\sigma+2\phi)\left(s_1^{-\frac{1}{\sigma}}+s_2^{-\frac{1}{\sigma}}\right)^{-\sigma-2\phi}(\Gamma(1-\sigma))^{1+\frac{2}{\sigma}}}{\Gamma(\phi)\Gamma(\sigma+\phi)\Gamma(1-\sigma)\rho^{\sigma-\text{stab}}\left(U^{-1}_{\sigma-\text{stab}}(s_1)\right)\rho^{\sigma-\text{stab}}\left(U^{-1}_{\sigma-\text{stab}}(s_2)\right)}\\
&=\frac{\sigma\left(s_1^{-\frac{1}{\sigma}}s_2^{-\frac{1}{\sigma}}\right)^{\phi-1}\Gamma(\sigma+2\phi)\left(s_1^{-\frac{1}{\sigma}}+s_2^{-\frac{1}{\sigma}}\right)^{-\sigma-2\phi}(\Gamma(1-\sigma))^{3+\frac{2}{\sigma}}}{\Gamma(\phi)\Gamma(\sigma+\phi)\Gamma(1-\sigma)\sigma^2\left(\Gamma(1-\sigma)s_1\right)^{1+\frac{1}{\sigma}}\left(\Gamma(1-\sigma)s_2\right)^{1+\frac{1}{\sigma}}}\\
&=\frac{\Gamma(\sigma+2\phi)\left(s_1^{-\frac{1}{\sigma}}+s_2^{-\frac{1}{\sigma}}\right)^{-\sigma-2\phi}}{\Gamma(\phi)\Gamma(\sigma+\phi)\sigma\left(s_1s_2\right)^{1+\frac{\phi}{\sigma}}}
\end{aligned}$$

which is greater than zero for $s_2,s_1>0$ and is well defined for $0<\sigma$.
By symmetry and using (3.3) it suffices to check that

$$\lim_{s_2\to 0}\hat{F}_{S_2|S_1=s_1}(s_2)\lim_{s_2\to 0}=\frac{\partial}{\partial s_1}C_{\sigma,\phi}(s_1,s_2)=0$$

and

$$\lim_{s_2\to\infty}\hat{F}_{S_2|S_1=s_1}(s_2)=\lim_{s_2\to\infty}\frac{\partial}{\partial s_1}\mathscr{C}_{\sigma,\phi}(s_1,s_2)=1.$$

For the first limit we observe that

$$\frac{\partial}{\partial s_1}\mathscr{C}_{\sigma,\phi}(s_1,s_2) =$$

$$\frac{\sigma\Gamma(\sigma+2\phi)}{\Gamma(\phi)\Gamma(\sigma+\phi)}\sum_{k=0}^{\phi-1}\binom{\phi-1}{k}\sum_{j=0}^{\phi+k-1}\binom{\phi+k-1}{j}(-1)^{k+j}\frac{\partial}{\partial s_1}\left(\frac{s_2^{-\frac{j}{\sigma}}\left(s_1^{-\frac{1}{\sigma}}+s_2^{-\frac{1}{\sigma}}\right)^{-\sigma-j}}{(\sigma+j)(\sigma+\phi+k)}\right)$$

$$= \frac{\sigma\Gamma(\sigma+2\phi)}{\Gamma(\phi)\Gamma(\sigma+\phi)}\sum_{k=0}^{\phi-1}\binom{\phi-1}{k}\sum_{j=0}^{\phi+k-1}\binom{\phi+k-1}{j}$$
$$\times(-1)^{k+j}\frac{\left(1+\frac{j}{\sigma}\right)s_2^{-\frac{j}{\sigma}}s_1^{-\frac{1}{\sigma}-1}\left(s_1^{-\frac{1}{\sigma}}+s_2^{-\frac{1}{\sigma}}\right)^{-\sigma-j-1}}{(\sigma+j)(\sigma+\phi+k)}$$

and for any $j\in\mathbb{N}$

$$\lim_{s_2\to 0}\left(s_2^{-\frac{j}{\sigma}}s_1^{-\frac{1}{\sigma}-1}\left(s_1^{-\frac{1}{\sigma}}+s_2^{-\frac{1}{\sigma}}\right)^{-\sigma-j-1}\right)=\lim_{s_2\to 0}\left(s_2^{-\frac{j}{\sigma}}s_1^{-\frac{1}{\sigma}-1}\left(\frac{s_1^{\frac{1}{\sigma}}s_2^{\frac{1}{\sigma}}}{s_1^{\frac{1}{\sigma}}+s_2^{\frac{1}{\sigma}}}\right)^{\sigma+j+1}\right)$$
$$= s_1^{-\frac{1}{\sigma}-1}\lim_{s_2\to 0}\left(s_2^{1+\frac{1}{\sigma}}\right)\lim_{s_2\to 0}\left(\frac{s_1^{\frac{1}{\sigma}}}{s_1^{\frac{1}{\sigma}}+s_2^{\frac{1}{\sigma}}}\right)^{\sigma+j+1} = 0$$

So

$$\lim_{s_2\to 0}\hat{F}_{S_2|S_1=s_1}(s_2)=0.$$

On the other hand

$$\lim_{s_2\to\infty}\frac{s_2^{1+\frac{1}{\sigma}}}{\left(s_1^{\frac{1}{\sigma}}+s_2^{\frac{1}{\sigma}}\right)^{\sigma+1}}=\lim_{s_2\to\infty}\left(\frac{s_2^{\frac{1}{\sigma}}}{s_1^{\frac{1}{\sigma}}+s_2^{\frac{1}{\sigma}}}\right)^{\sigma+1}=1$$

So for $j \in \mathbb{N}$

$$\lim_{s_2\to\infty}\left(s_2^{-\frac{j}{\sigma}}s_1^{-\frac{1}{\sigma}-1}\left(s_1^{-\frac{1}{\sigma}}+s_2^{-\frac{1}{\sigma}}\right)^{-\sigma-j-1}\right) = s_1^{\frac{j}{\sigma}}\lim_{s_2\to\infty}\frac{s_2^{1+\frac{1}{\sigma}}}{\left(s_1^{\frac{1}{\sigma}}+s_2^{\frac{1}{\sigma}}\right)^{\sigma+j+1}}$$

$$= s_1^{\frac{j}{\sigma}}\lim_{s_2\to\infty}\frac{1}{\left(s_1^{\frac{1}{\sigma}}+s_2^{\frac{1}{\sigma}}\right)^{j}} = \begin{cases}0, & \text{for } j\neq 0\\ 1, & \text{for } j=0\end{cases}$$

It follows that

$$\lim_{s_2\to\infty}\hat{F}_{S_2|S_1=s_1}(s_2) = \frac{\sigma\Gamma(\sigma+2\phi)}{\Gamma(\phi)\Gamma(\sigma+\phi)}\sum_{k=0}^{\phi-1}\binom{\phi-1}{k}(-1)^k\frac{1}{\sigma(\sigma+\phi+k)}$$

From formula 0.160.2 in Ryzhik and Gradshteyn (1965) we have that

$$\sum_{k=0}^{n}\binom{n}{k}(-1)^k\frac{\Gamma(k+c)}{\Gamma(k+c+1)} = \sum_{k=0}^{n}\binom{n}{k}(-1)^k\frac{1}{k+c} = \mathrm{B}(n+1,c)$$

So we conclude that

$$\lim_{s_2\to\infty}\hat{F}_{S_2|S_1=s_1}(s_2) = \frac{\Gamma(\sigma+2\phi)\mathrm{B}(\phi,\sigma+\phi)}{\Gamma(\phi)\Gamma(\sigma+\phi)} = \frac{\Gamma(\sigma+2\phi)\mathrm{B}(\phi,\sigma+\phi)}{\Gamma(\phi)\Gamma(\sigma+\phi)} = 1.$$

As this limits do not depend on what values σ takes in $(0,\infty)$ we conclude that we can construct a CRM with the desired Lévy copula for any $\sigma \in (0,\infty)$.

Chapter 4

Multiple-sample Neutral to the Right Model

In this chapter we generalize the model of Epifani and Lijoi (2010) to an arbitrary dimension on the underlying VCRM in their model. We provide the posterior characterization for the model, see Theorem 12. We retain the conjugacy property of NTR type models as discussed in section 1.3. Extensions of some results in Epifani and Lijoi (2010) and Doksum (1974) are also provided. The derivation of such results is not trivial when considering an arbitrary dimension. Proposition 8 gives a general expression for the Laplace exponent when a Lévy copula is considered to set the dependence of the VCRM underlying the multiple-sample NTR model; Proposition 10 gives an alternative characterization of the model. 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 chapter is divided as follows: 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.

4.1 Exchangeability and Partial exchangeability

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

Definition 20. 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

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

In several modelling problems the exchangeability assumption appears to be far too restrictive. In particular, if we consider observations airsing from d different populations where the order in which they are collected within each population is irrelevant. To describe this setting we resort to the notion of partial exchangeability, which was introduced in de Finetti (1938), as set forth in de Finetti (1980). Partial exchangeability formalizes the idea of partitioning a set of observations into a certain number of classes, say d, in such a way that exchangeability is attained within each class. For ease of exposition, we confine ourselves to consider the case where $d = 2$.

Definition 21. 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

$$\big(Z_1^{(1)},\ldots,Z_{m_1}^{(1)},Z_1^{(2)},\ldots,Z_{m_2}^{(2)}\big) \stackrel{\mathrm{d}}{=} \big(Z_{\pi_1(1)}^{(1)},\ldots,Z_{\pi_1(m_1)}^{(1)},Z_{\pi_2(1)}^{(2)},\ldots,Z_{\pi_2(m_2)}^{(2)}\big). \tag{4.1}$$

A fundamental result regarding exchangeability is de Finetti's representation theorem which states that a sequence of random variables is exchangeable if and only if it is conditionally i.i.d., see for instance Kallenberg (2006). For example if

$$Y_i|\mu \stackrel{\text{i.i.d.}}{\sim} \mathrm{NTR}(\mu), \quad i \in \{1,\ldots\};$$

then it follows that $\{Y_i\}_{i=1}^{\infty}$ are conditionally i.i.d. and hence exchangeable. An extension of the NTR model into a partially exchangeable setting was given by Epifani and Lijoi (2010) for the $2-$dimensional case. They considered two populations

$$\{Y_i^{(1)}\}_{i=1}^{\infty}, \quad \{Y_i^{(2)}\}_{i=1}^{\infty}$$

such that for $t_1,t_2 \in \mathbb{R}^+$ and $i,j \in \mathbb{N}$

$$\mathbb{P}\Big[Y_i^{(1)} > t_1, Y_j^{(2)} > t_2 \,|\, (\mu_1,\mu_2)\Big] = e^{-\mu_1(0,t_1]-\mu_2(0,t_2]}.$$

So

$$Y_i^{(1)}|\mu_1 \overset{\text{i.i.d.}}{\sim} \mathrm{NTR}(\mu), \quad i \in \{1, \dots\},$$

and

$$Y_i^{(2)}|\mu_2 \overset{\text{i.i.d.}}{\sim} \mathrm{NTR}(\mu_2), \quad i \in \{1, \dots\},$$

implying that each population $\{Y_i^{(j)}\}_{i=1}^{\infty}$ is exchangeable, $j \in \{1,2\}$ although $\{Y_i^{(1)}\}_{i=1}^{\infty} \cup \{Y_i^{(2)}\}_{i=1}^{\infty}$ is not necessarily exchangeable as $Y_i^{(1)}$, $Y_j^{(2)}$, $i, j \in \mathbb{N}$ are not identically distributed when $\mu_1 \overset{\text{a.s.}}{\neq} \mu_2$. It follows that $\{Y_i^{(1)}, Y_i^{(2)}\}_{i=1}^{\infty}$ are partially exchangeable.

4.2 Multiple-sample NTR model

In the present work, we follow the approach of Epifani and Lijoi (2010) and focus on models based on a d-dimensional VCRM.

Definition 22. Let $d \in \mathbb{N} \setminus \{0\}$ and $\boldsymbol{\mu} = (\mu_1, \dots, \mu_d)$ a d-variate VCRM such that

$$\lim_{t \to \infty} \mu_i(0,t] \overset{\text{a.s.}}{=} \infty$$

for any $i \in \{1, \dots, d\}$. We say that d collections of survival times

$$\{Y_j^{(1)}\}_{j=1}^{\infty}, \dots, \{Y_j^{(d)}\}_{j=1}^{\infty}$$

follow a multiple-sample NTR distribution, denoted $\mathrm{NTR}(\boldsymbol{\mu})$, if for $t_{i,j} \in (\mathbb{R}^+)^d$, $n_i \in \mathbb{N} \setminus \{0\}$, $(i, j) \in \{(i, j)\,;\, 1 \leq j \leq n_i,\, 1 \leq i \leq d\}$.

$$\mathbb{P}\left[Y_1^{(1)} > t_{1,1}, \dots, Y_{n_1}^{(1)} > t_{1,n_1}, \dots, Y_1^{(d)} > t_{d,1}, \dots, Y_{n_d}^{(d)} > t_{d,n_d}\,|\boldsymbol{\mu}\right] = \prod_{i=1}^{d} \prod_{j=1}^{n_i} e^{-\mu_i(0,t_{i,j}]}. \tag{4.2}$$

In particular for $\boldsymbol{t} = (t_1, , \dots t_d) \in (\mathbb{R}^+)^d$

$$S(\boldsymbol{t}) = \mathbb{P}\left[Y_{i_1}^{(1)} > t_1, \dots, Y_{i_d}^{(d)} > t_d\,|(\mu_1, \dots, \mu_d)\right] = e^{-\mu_1(0,t_1] - \dots - \mu_d(0,t_d]}, \tag{4.3}$$

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

$$Y_1^{(i)}, \dots, Y_{n_i}^{(i)} \overset{\text{i.i.d.}}{\sim} \text{NTR}(\mu_i)$$

with $i \in \{1, \dots, d\}$, $n_i \in \mathbb{N} \setminus \{0\}$; we observe that this is a clear extension of the model in Epifani and Lijoi (2010) to an arbitrary dimension d. In (4.3) we want to model the dependence in the entries of the VCRM $\boldsymbol{\mu}$ in a way that allows us to fix a marginal behaviour so we can exploit the fact that marginally we recover a NTR model; Lévy copulas as set in Definition 17 are a natural framework to model the dependence structure of a VCRM's entries in such way so we will be using them in some of the following results. We remember that Lévy copulas as set forth in Chapter 3 assume that the related Lévy measure is homogeneous, so for instance the Laplace exponent can be written as $\psi_t(\boldsymbol{\lambda}) = \gamma(t)\psi(\boldsymbol{\lambda})$ for some $\gamma : \mathbb{R}^+ \to \mathbb{R}^+$ which must satisfy $\lim_{t\to\infty} \gamma(t) = \infty$. The family of Clayton Lévy copulas, Example 14, 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 is useful since, as we will see, an explicit calculation of ψ is of key importance to implement the Bayesian inference in the model above.

Let $\tilde{\rho}_{d,\theta}$ be the Lévy intensity associated via Sklar Theorem 8 to the Clayton Lévy copula $\mathscr{C}_{\theta,d}$ and fixed marginal Lévy intensities $\rho_1, \dots, \rho_d$ with corresponding Laplace transforms $\psi_1, \dots, \psi_d$. We denote the vector of tail integrals corresponding to the marginal Lévy intensities as $\boldsymbol{U}_d(\boldsymbol{x}) = (U_1(x_1), \dots, U_d(x_d))$ and fix the notation

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

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

Proposition 8. *Suppose that* $d \in \{2, 3, \dots\}$ *and*

$$\int_{\|\boldsymbol{s}\| \leq 1} \|\boldsymbol{s}\| \tilde{\rho}_{d,\theta}(\boldsymbol{s}) \mathrm{d}\boldsymbol{s} < \infty$$

then

$$\psi(\boldsymbol{\lambda}) = \int_{(\mathbb{R}^+)^d} (1 - e^{-<\boldsymbol{\lambda},\boldsymbol{s}>}) \frac{\partial^d}{\partial u_d \cdots \partial u_1} \mathscr{C}_{\theta,d}(\boldsymbol{u})\big|_{\boldsymbol{u}=\boldsymbol{U}_d(\boldsymbol{s})} \rho_1(s_1)\cdots\rho_d(s_d)\mathrm{d}\boldsymbol{s}$$

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

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

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

Calculation of the Laplace exponent as above is important for the evaluation of various quantities of interest in the setting of our multiple-sample NTR model. For example, in the next result we give a formula for the prior survival in our model where we integrate out the underlying VCRM. With such purpose, we introduce the following notation

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

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

Proposition 9. *In the context of 4.2, let* $\mathbf{1} = (1,\dots,1)$. *For* $t_1 \le \cdots \le t_d$ *and* $i_1,\dots,i_d \in \{1,\dots,d\}$ *such that* $t_{i_1} \le \cdots \le t_{i_d}$ *then*

$$\mathbb{P}\Big[Y^{(1)} > t_1,\dots,Y^{(d)} > t_d\Big] = e^{-\gamma(t_{i_1})\psi(\mathbf{1}) - (\gamma(t_{i_2})-\gamma(t_{i_1}))\psi_{i_2,\dots,i_d}(\mathbf{1})\cdots - \left(\gamma(t_{i_d})-\gamma(t_{i_{d-1}})\right)\psi_{i_d}(\mathbf{1})}.$$

This result showcases the importance of the Laplace exponent ψ for calculating probabilities in the multiple-sample information NTR model and the impact of the function $\gamma(t)$, related to the time depending part of the Laplace exponent, in the survival function. In the Applications section of this chapter, 4.3, 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 $\boldsymbol{\mu} = (\mu_1,\dots,\mu_d)$, denote $\mu_i(t) = \mu_i((0,t])$ with $i \in \{1,\dots,d\}$. Then, we have the following multivariate extension to Theorem 3.1 in Doksum (1974) and Proposition 4 in Epifani and Lijoi (2010).

Proposition 10. *$F(\boldsymbol{t})$, with $\boldsymbol{t} = (t_1, \dots, t_d)$, has the same distribution as*

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

for some d-variate CRM $\boldsymbol{\mu} = (\mu_1, \dots, \mu_d)$ if and only if for $h \in \{1, 2, \dots\}$ and vectors $\boldsymbol{t}_1 = (t_{1,1}, \dots, t_{d,1}), \dots, \boldsymbol{t}_h = (t_{1,h}, \dots, t_{d,h})$ 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 $(V_{1,1}, \dots V_{d,1}), \dots, (V_{1,h}, \dots V_{d,h})$ such that

$$\begin{aligned}
F(\boldsymbol{t}_1) &\stackrel{d}{=} V_{1,1} \cdots V_{d,1} \\
F(\boldsymbol{t}_2) &\stackrel{d}{=} [1 - \bar{V}_{1,1}\bar{V}_{1,2}] \cdots [1 - \bar{V}_{d,1}\bar{V}_{d,2}] \\
&\vdots \\
F(\boldsymbol{t}_h) &\stackrel{d}{=} [1 - \prod_{j=1}^{h} \bar{V}_{1,j}] \cdots [1 - \prod_{j=1}^{h} \bar{V}_{d,j}]
\end{aligned} \tag{4.4}$$

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

When possibly censored to the right survival data is considered for our model in (4.2) we have to generalize the setting of survival data, Definition 7, for the multiple sample setting.

Definition 23. Let $d, n_1, \dots, n_d, n \in \mathbb{N} \setminus \{0\}$ such that $n = \sum_{i=1}^{d} n_i$ and $\{Y_j^{(1)}\}_{j=1}^{\infty}, \dots, \{Y_j^{(d)}\}_{j=1}^{\infty}$ be d groups of observations following a multiple-sample NTR distribution. Associate to each group of observations $\{Y_j^{(i)}\}_{j=1}^{\infty}$ possibly censored to the right data

$$\mathscr{D}^{(i)} = \left\{ \left(T_j^{(i)}, J_j^{(i)} \right) \right\}_{j=1}^{n_i},$$

$i \in \{1, \dots, d\}$. We say that

$$\mathscr{D}_d = \bigcup_{i=1}^{d} \mathscr{D}^{(i)}$$

is censored to the right data with multiple-samples.

We establish some notation in order to address the posterior distribution arising from the model in (4.2) The number of exact observations is

$$n_e = \sum_{i=1}^{d} \sum_{j=1}^{n_i} J_i^{(j)}$$

and the number of censored observations is $n_c = n - n_e$. Taking into account the possible repetition of values among the observations

$$\{T_j^{(1)}\}_{j=1}^{n_1}, \dots, \{T_j^{(d)}\}_{j=1}^{n_d}$$

we consider the order statistics $(T_{(1)}, \dots, T_{(k)})$ of the distinct observations where k is the number of distinct observed times among all groups; we set $T_{(0)} = 0$ and $T_{(k+1)} = \infty$. Let define the set functions

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

for $i \in \{1, \dots, d\}$, which denote the number of, respectively, exact and censored marginal observations in A, with respect to group i. We define the numbers

$$n_{i,j}^e = m_i^e(\{T_{(j)}\}) \quad ; \quad n_{i,j}^c = m_i^c(\{T_{(j)}\})$$

for, respectively, the exact and censored observations in group i related to the time $T_{(j)}$, $i \in 1, \dots, d\}$ and $j \in \{1, \dots, k\}$. The following cumulative quantities will also be of use

$$\bar{n}_{i,j}^e = \sum_{r=j}^{k} n_{i,r}^e \quad ; \quad \bar{n}_{i,j}^c = \sum_{r=j}^{k} n_{i,r}^c$$

and the corresponding vectors

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

$i \in 1, \dots, d\}$ and $j \in \{1, \dots, k\}$. The next theorem determines the calculation of the posterior characterization for the undelrying VCRM in the multiple-sample NTR model given possibly censored to the right survival data; we highlight that it applies to a general VCRM as the assumption that the respective Lévy intensity is homogeneous has been dropped.

Theorem 12. *Let $\boldsymbol{\mu} = (\mu_1, \dots, \mu_d)$ be a d-variate VCRM with corresponding Lévy intensity of the form $\tilde{\nu}_d(\mathrm{d}\boldsymbol{s}, \mathrm{d}x) = \tilde{\nu}_d(\boldsymbol{s}, \mathrm{d}x)\mathrm{d}\boldsymbol{s}$ and let $\mathscr{D}_d$ be survival data with d multiple-samples arising from a multiple-sample NTR($\boldsymbol{\mu}$) distribution. If for $\eta_t = \tilde{\nu}_d(\boldsymbol{s}, (0,t])$ and arbitrary $t_0 \in \mathbb{R}^+ \setminus \{0\}$ the partial derivative $\eta'_{t_0}(\boldsymbol{s}) = \partial \eta_t(\boldsymbol{s}) / \partial t \big|_{t=t_0}$ exists then the posterior distribution of $\boldsymbol{\mu}$ given survival data $\mathscr{D}_d$ is the distribution of the random measure*

$$(\mu_1^\circ, \dots, \mu_d^\circ) + \sum_{\{j : T_{(j)} \text{is an exact observation}\}} (M_{1,j}\delta_{T_{(j)}}, \dots, M_{d,j}\delta_{T_{(j)}})$$

where

i) $\boldsymbol{\mu}^\circ = (\mu_1^\circ, \ldots, \mu_d^\circ)$ *is a d-variate VCRM with Lévy intensity* ν_d° *such that*

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

for $j \in \{1, \ldots, k+1\}$.

ii) The vectors of jumps $\{(M_{1,j}, \ldots, M_{d,j})\}_{j \in I^{(e)}}$, *with*

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

are mutually independent and have, respectively, a d-variate probability density function given by

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

iii) The random measure $\boldsymbol{\mu}^\circ$ *is independent of* $\{(M_{1,j}, \ldots, M_{d,j})\}_{j \in M}$.

The previous result showcases that the posterior distribution arising from (4.2) can be modelled in the same framework using the VCRM $\boldsymbol{\mu}^\circ$ instead of $\boldsymbol{\mu}$ and adding as extra component a series of δ measures with random weights and locations on the exact observation times. This Theorem is enough to provide a scheme for posterior inference. In particular, we want to estimate the corresponding posterior survival function of the from

$$S(\boldsymbol{t}) = S(t_1, \ldots, t_d) = \mathbb{P}\left[Y^{(1)} > t_1, \ldots, Y^{(d)} > t_d \,|\, \boldsymbol{\mu}\right]$$

when multiple-sample survival data 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 possibly censored to the right data $\mathscr{D}_d$, we calculate the mean of the survival function given the data by marginalizing over the VCRM $\boldsymbol{\mu}$. As a result of Theorem 12 we can calculate such quantity. The next corollary allows us to implement the necessary inferential scheme for performing the estimation of the survival function as a posterior mean. We set $S_L(t) = S(t \sum_{l \in L} \boldsymbol{e}_l)$ for $t > 0$, $\emptyset \neq L \subset \{1, \ldots, d\}$. In view of the independent increments of the CRM's in a VCRM, the 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 4. *In the setting of Theorem 12, let* $\emptyset \neq L \subset \{1, \dots, d\}$ *and set*

$$I_t^{(e)} = \{j : T_{(j)} \text{ is an exact observation and } T_{(j)} \leq t\}$$

Then

$$\hat{S}_L(t) = \mathbb{E}[\mathbb{E}[S_L(t)|\boldsymbol{\mu}]\,|\mathscr{D}_d] = e^{-\sum_{j=1}^{k+1}\left(\psi^{\circ j}_{t\wedge T_{(j)}}(\sum_{l\in L}\boldsymbol{e}_l)-\psi^{\circ j}_{t\wedge T_{(j-1)}}(\sum_{l\in L}\boldsymbol{e}_l)\right)\mathbb{1}_{\{T_{(j-1)}<t\}}}$$
$$\times \prod_{j\in I_t^{(e)}} \frac{\int_{(\mathbb{R}^+)^d}\prod_{i=1}^d\left(e^{-(\mathbb{1}_{\{i\in L\}}+\bar{n}^c_{i,j}+\bar{n}^e_{i,j+1})s_i}(1-e^{-s_i})^{n^e_{i,j}}\right)\eta'_{T_{(j)}}(\boldsymbol{s})\mathrm{d}\boldsymbol{s}}{\int_{(\mathbb{R}^+)^d}\prod_{i=1}^d\left(e^{-(\bar{n}^c_{i,j}+\bar{n}^e_{i,j+1})s_i}(1-e^{-s_i})^{n^e_{i,j}}\right)\eta'_{T_{(j)}}(\boldsymbol{s})\mathrm{d}\boldsymbol{s}}$$

where ψ_t° *is the Laplace exponent of* μ°.

From the independence of increments of CRM's, it follows the next corollary which gives us an estimator for $S(\boldsymbol{t})$ for arbitrary $\boldsymbol{t} \in (\mathbb{R}^+)^d$ in terms of the estimates defined in the previous corollary.

Corollary 5. *In the setting of Theorem 12, let* $S(\boldsymbol{t})$ *be the survival function associated to a* d*-variate multiple-sample NTR distribution. For* $\boldsymbol{t} = (t_1, \dots, t_d)$ *and* π *the permutation of* $\{1, \dots, d\}$ *such that* $t_{\pi(1)} \leq t_{\pi(2)} \leq \dots \leq t_{\pi(d)}$. *We define, for* $i \in \{1, \dots, d-1\}$, *the sets*

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

then the posterior mean of the survival function $S(\boldsymbol{t})$ *given multiple-sample survival data* $\mathscr{D}_d$ *is*

$$\hat{S}(\boldsymbol{t}) = \mathbb{E}[\mathbb{E}[S(\boldsymbol{t})|\boldsymbol{\mu}]\,|\mathscr{D}_d] = \hat{S}_{L_1}(t_{\pi(1)})\prod_{i=1}^{d-1}\frac{\hat{S}_{L_i}(t_{\pi(i+1)})}{\hat{S}_{L_i}(t_{\pi(i)})} \tag{4.5}$$

for arbitrary $\boldsymbol{t} \in (\mathbb{R}^+)^d$.

Usually, we deal with Lévy intensities which exhibit some dependences in a vector of hyper-parameters $\boldsymbol{c}$. In the proof of Theorem 12, it is outlined how, given multiple-sample survival data $\mathscr{D}_d$, we can 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 6. *In the setting of Theorem 12 with survival data* $\mathscr{D}_d$, *underlying Lévy intensity* $\tilde{\nu}_{d,\boldsymbol{c}}$, *associated partial derivative* $\eta'_{t,\boldsymbol{c}}$ *and Laplace exponent* $\psi_{t,\boldsymbol{c}}$ *for some real valued vector* *of hyper-parameters* $\boldsymbol{c}$*, we get the likelihood on* $\boldsymbol{c}$

$$l(\boldsymbol{c};\mathscr{D}_d) = e^{-\sum_{j=1}^{k}\left(\psi_{T_{(j)},\boldsymbol{c}}(\bar{\boldsymbol{n}}_j^c+\bar{\boldsymbol{n}}_j^e)-\psi_{T_{(j-1)},\boldsymbol{c}}(\bar{\boldsymbol{n}}_j^c+\bar{\boldsymbol{n}}_j^e)\right)}$$
$$\times\prod_{j\in M}\int_{(\mathbb{R}^+)^d}\prod_{i=1}^{d}\left(e^{-(\bar{n}_{i,j}^c+\bar{n}_{i,j+1}^e)s_i}(1-e^{-s_i})^{n_{i,j}^e}\right)\eta'_{T_{(j)},\boldsymbol{c}}(\boldsymbol{s})\mathrm{d}\boldsymbol{s}.$$

The next proposition is useful for the evaluation of the estimator in Corollary 5.

Proposition 11. *In the setting of Theorem 12 the Laplace exponent of* $\boldsymbol{\mu}^\circ$ *restricted to* $(T_{(j-1)},T_{(j)})$*,* $j\in\{1,\ldots,d\}$*, can be evaluated as*

$$\psi_t^\circ(\boldsymbol{\lambda})\big|_{[T_{(j-1)},T_{(j)})} = \int_{(0,t]\times(\mathbb{R}^+)^d}\left(1-e^{-\langle\boldsymbol{\lambda},\boldsymbol{s}\rangle}\right)e^{-\langle\bar{\boldsymbol{n}}_j^c+\bar{\boldsymbol{n}}_j^e,\boldsymbol{s}\rangle}\,\tilde{\nu}_d(\boldsymbol{s},\mathrm{d}x)\mathrm{d}\boldsymbol{s}$$
$$= \psi_t(\boldsymbol{\lambda}+\bar{\boldsymbol{n}}_j^c+\bar{\boldsymbol{n}}_j^e)-\psi_t(\bar{\boldsymbol{n}}_j^c+\bar{\boldsymbol{n}}_j^e).$$

The next proposition provides a useful identity for the computation of the integrals in Corollary 5 and Corollary 6.

Proposition 12. *In the setting of Theorem 12, let* $n_1,\ldots,n_d\in\mathbb{N}$ *and* $j\in\{1,\ldots,d\}$ *be such that* $n_j>0$*. Set* $n=\sum_{i=1}^d n_i$ *and a multiset* $I=\{i_1,\ldots,i_n\}\subset\{1\le k\le d : n_k\neq 0\}$ *such that* $\#\{i\in I : i=k\}=n_k$*; then*

$$\int_{(\mathbb{R}^+)^d\times(0,t]}e^{-\langle\boldsymbol{q},\boldsymbol{s}\rangle}\prod_{i=1}^{d}\left(1-e^{-s_i}\right)^{n_i}\tilde{\nu}_d(\mathrm{d}\boldsymbol{s},\mathrm{d}x)$$
$$=\sum_{S\subset I\setminus j}(-1)^{\#(S)}\Big(\psi_t(\boldsymbol{e}_j+\sum_{l\in S}\boldsymbol{e}_l+\boldsymbol{q})-\psi_t(\sum_{l\in S}\boldsymbol{e}_l+\boldsymbol{q})\Big)$$

where $I\setminus j=\left\{i_1,\ldots,i_{j-1},i_{j+1},\ldots,i_n\right\}$*.*

The previous results highlight how the implementation of the inferential procedure depends on whether we can perform evaluations of the Laplace exponent or not, this will be of key importance in the next section.

4.3 Multiple-sample NTR model application

In this section we perform the fitting of a multivariate survival function given possibly censored to the right multiple-sample survival data in the framework of (4.2). As mentioned previously, the evaluation of the Laplace exponent of $\boldsymbol{\mu}$ in (4.2) is necessary to evaluate the posterior mean in Corollary 5 and the likelihood in Corollary 6; with this in mind, we choose the random measure $\boldsymbol{\mu}$ given by the Lévy intensity showcased in (3.4), so that the corresponding Laplace exponent is readily given by (3.5). For illustration purposes, we use 4-dimensional data arising from a distributional copula with fixed marginal distributions, see Nelsen (2007) for an overview of distributional copulas. More precisely, we generate simulated data $\boldsymbol{Y} = (Y_1, ..., Y_4)$ 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 $\boldsymbol{C} = (C_1, \ldots, C_4)$ consisting of independent exponential random variables with parameter λ_C. We will consider 150 synthetic observation drawn as follows

$$\begin{aligned} &\left(Y_i^{(1)}, Y_i^{(2)}, Y_i^{(3)}, Y_i^{(4)}\right) \sim F_{\theta=0.3,\lambda=1.}, && j = 1, \ldots, 150 \\ &C_j^{(i)} \sim \mathrm{Exp}(\lambda_c = 3.7), && i = 1, ..., 4; \quad j = 1, \ldots, 150 \end{aligned}$$

We chose $\lambda_c = 3.7$ so we have at least 75% of exact observations for $\boldsymbol{T}$ in each dimension. Set

$$J_j^{(i)} = \mathbb{1}_{\{Y_j^{(i)} \leq C_j^{(i)}\}}$$

and

$$T_j^{(i)} = \min\{Y_j^{(i)}, C_j^{(i)}\}.$$

We obtain the multiple-sample censored to the right data

$$\mathscr{D}_4 = \bigcup_{i=1}^{4} \left\{ \left(T_j^{(i)}, J_j^{(i)} \right) \right\}_{j=1}^{150} /$$

The construction of $F_{\theta,\lambda}$ through a distributional Clayton copula allows us to calculate explicitly the associated survival function. Indeed if $C_{d,\theta}$ 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 with the previous marginals is given by

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

see Section 2.6 in Nelsen (2007). 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 (4.5). For fitting the data, we use the 4-dimensional Lévy intensity given by (3.4) and assign priors for the corresponding hyper-parameters σ and A which we respectively denote p_σ and p_A. We choose a LogNormal 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, i.e. $\sigma|\mathscr{D}_d$ and $A|\mathscr{D}_d$, by making use of the likelihood $l(\sigma,A;\mathscr{D}_d)$ showed in Corollary 6. Given initial values $\sigma^{(0)}$, $A^{0)}$, the algorithm is as follows

Algorithm 5 Metropolis within Gibbs for Multiple-Smaple NTR model

1: Draw $A^{(i+1)}$ from a Metropolis-Hastings sampler with proposal distribution $g(x'|x) \sim$ LogNormal$(\log(x),1)$ and target distribution $l(\sigma^{(i)},x;\mathscr{D}_d)p_A(x)$.
2: Draw $\sigma^{(i+1)}$ from a Metropolis-Hastings sampler with Uniform proposal distribution and target distribution $l(x,A^{(i+1)};\mathscr{D}_d)p_\sigma(x)$.

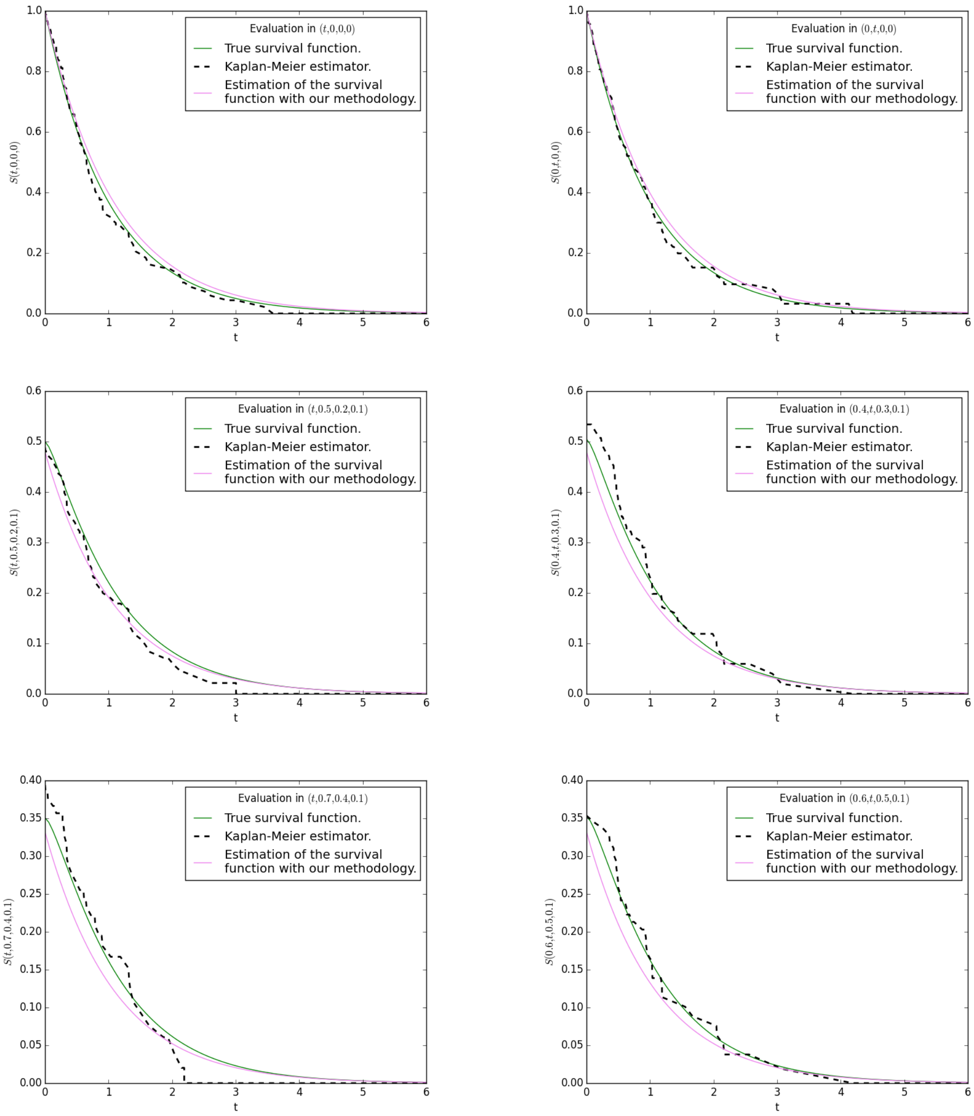
After using the above algorithm we perform a Monte Carlo approximation of the estimator (4.5), where we average over the posterior draws of A and σ. For a full review of Markov Chain Monte Carlo methods refer, for example, to Robert and Casella (2010). The prior distributions we choose for the hyperparameters are

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

In Figures 4.1 and 4.2 we show the fit for 150 possibly right censored observations where we performed 100 iterations for the inner Metropolis-Hasting sampler and 1000 iterations for the overall Gibbs sampler in algorithm 5. The estimated survival functions approximate well the true functions. For comparison purposes, we presented a Kaplan-Meier type of estimator for the true survival function. As there is no multivariate Kaplan-Meier, we use the next estimator for a multivariate survival function:

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

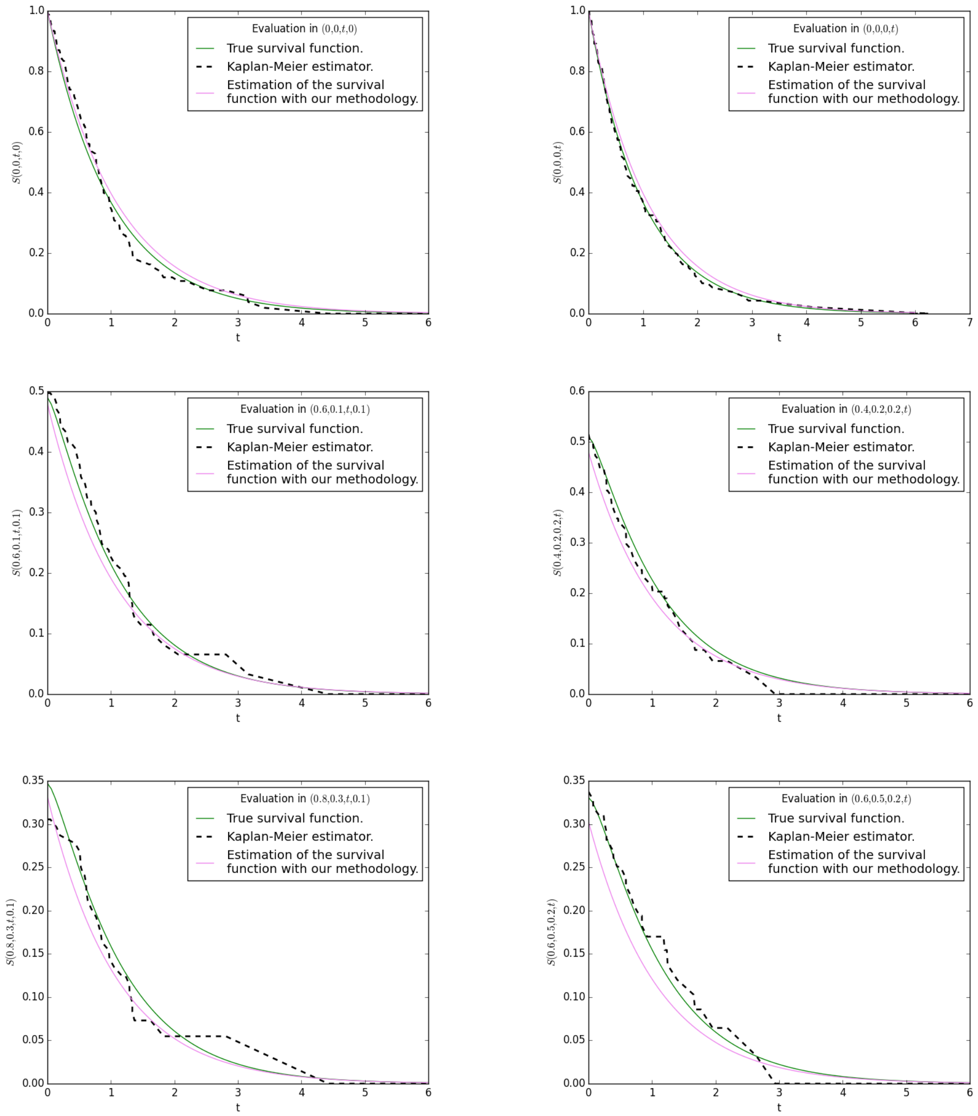
where each S_{KM} estimator is treated as a univariate Kaplan-Meier estimator restricted to the corresponding set of observations. In Figure 4.1 and Figure 4.2, we can appreciate in the last subplots of each column that the Kaplan-Meier can fit poorly as there are less observations on the conditioned Kaplan-Meier functions, as presented in the formula above.



(a) Fits with first dimension not fixed.

(b) Fits with second dimension not fixed.

Fig. 4.1 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.



(a) Fits with third dimension not fixed.

(b) Fits with fourth dimension not fixed.

Fig. 4.2 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.

4.4 Proofs of multiple-sample NTR model results

Proof of Proposition 8

Given $d \in \{2,3,\dots\}$, we use the notation $\rho_{-i}(\boldsymbol{s}) = \prod_{j=i+1}^{d} \rho_j(s_j)$ and

$$\boldsymbol{U}_{k:d}(\boldsymbol{s}) = (U_k(s_1),\dots,U_d(s_{d-k+1}))$$

for $\boldsymbol{s} \in (\mathbb{R}^+)^d$. Furthermore we define integrals

$$a_{0,m}(\boldsymbol{\lambda}) = \int_{(\mathbb{R}^+)^m} (1-e^{-<\boldsymbol{\lambda},\boldsymbol{s}>}) \frac{\partial^d}{\partial u_d \cdots \partial u_1} \mathscr{C}_{\theta,m}(\boldsymbol{u})\big|_{\boldsymbol{u}=\boldsymbol{U}_{d-m+1:d}(\boldsymbol{s})} \rho_{-0}(\boldsymbol{s}) \mathrm{d}\boldsymbol{s}$$

and

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

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.
An integration by parts shows that

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

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

$$a_{r,d}(\boldsymbol{\lambda}) = a_{r,d-1}(\boldsymbol{\lambda}_{-(r+1)}) + a_{r+1,d}(\boldsymbol{\lambda}). \tag{4.6}$$

We prove the next technical lemma which provides $d+1$ identities; of which the first one, index 0, will be of use for the proof of the Proposition and the rest are useful to prove by induction the $d+1$ identities of the Lemma.

Lemma 3. *If $a_{0,d}(\boldsymbol{\lambda}) < \infty$ then the next $d+1$ identities hold*

$$
\begin{aligned}
a_{0,d}(\boldsymbol{\lambda}) &= \sum_{i=1}^{d} \psi_i(\lambda_i) - \sum_{\substack{\boldsymbol{i}=(i_1,i_2)\in\{1,\dots,d\}^2 \\ i_1<i_2}} \kappa(\theta;\boldsymbol{\lambda},\boldsymbol{i}) + \cdots \\
&\cdots + (-1)^d \sum_{\substack{\boldsymbol{i}=(i_1,\dots,i_{d-1})\in\{1,\dots,d\}^{d-1} \\ i_1<\cdots<i_{d-1}}} \kappa(\theta;\boldsymbol{\lambda},(i_1,\dots,i_{d-1})) \\
&+ (-1)^{d+1}\kappa(\theta;\boldsymbol{\lambda},(1,\dots,d)) \\
a_{1,d}(\boldsymbol{\lambda}) &= \psi_1(\lambda_1) - \sum_{i=2}^{d} \kappa(\theta;\boldsymbol{\lambda},(1,i)) + \sum_{\substack{i_1,i_2\in\{2,\dots,d\} \\ i_1<i_2}} \kappa(\theta;\boldsymbol{\lambda},(1,i_1,i_2)) + \cdots \\
&\cdots + (-1)^d \sum_{\substack{i_1,\dots,i_{d-2}\in\{2,\dots,d\} \\ i_1<\cdots<i_{d-2}}} \kappa((\theta;\boldsymbol{\lambda},(1,i_1,\dots,i_{d-2})) \\
&+ (-1)^{d+1}\kappa(\theta;\boldsymbol{\lambda},(1,\dots,d)) \\
&\vdots \\
a_{d-1,d}(\boldsymbol{\lambda}) &= (-1)^d\kappa(\theta;\boldsymbol{\lambda},(1,\dots,d-1)) + (-1)^{d+1}\kappa(\theta;\boldsymbol{\lambda},(1,\dots,d)) \\
a_{d,d}(\boldsymbol{\lambda}) &= (-1)^{d+1}\kappa(\theta;\boldsymbol{\lambda},(1,\dots,d))
\end{aligned}
\tag{4.7}
$$

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

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

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

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

Therefore, we get the validity of the equations in (4.7) for the case $d=2$. Now, suppose that (4.7) is true for $d=m-1$, we must show the validity for $d=m$. From the recursion formula

(4.6) 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 (4.7) for $d = m$ follows from the validity for $d = m-1$ and a combinatorial argument. □

Proposition 8 follows by considering the first equation in the Lemma statement and the definition of $a_{0,d}$.

Proof of Proposition 9.

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

$$\begin{aligned}
&\mathbb{P}\Big[Y^{(1)} > t_1, \ldots, Y^{(d)} > t_d\Big] = \mathbb{E}\Big[e^{-\mu_1(0,t_1]-\cdots-\mu_d(0,t_d]}\Big] \\
&= \mathbb{E}\Big[e^{-\mu_{i_1}(0,t_{i_1}]-\cdots-\mu_{i_d}(0,t_{i_1}]}\Big]\,\mathbb{E}\Big[e^{-\mu_{i_2}(t_{i_1},t_{i_2}]-\cdots-\mu_{i_d}(t_{i_1},t_{i_2}]}\Big]\cdots \\
&\quad \cdots \times \mathbb{E}\Big[e^{-\mu_{i_d}(t_{i_{d-1}},t_{i_d}]}\Big] \\
&= e^{-\gamma(t_{i_1})\psi(\mathbf{1})} e^{-[\gamma(t_{i_2})-\gamma(t_{i_1})]\psi_{i_2,\ldots,i_d}(\mathbf{1})} \cdots e^{-[\gamma(t_{i_d})-\gamma(t_{i_{d-1}})]\psi_{i_d}(\mathbf{1})}
\end{aligned}$$

Proof of Proposition 10.

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 $(F_1(t_1), \ldots, F_d(t_d)) \stackrel{d}{=} (1 - e^{-\mu_1(t_1)}, \ldots, 1 - e^{-\mu_d(t_d)})$ we have

$$\begin{aligned}
F(t_{1,1}, \ldots, t_{d,1}) &\stackrel{d}{=} [1 - e^{-\mu_1(t_{1,1})}] \cdots [1 - e^{-\mu_d(t_{d,1})}] \\
&= [1 - e^{-[\mu_1(t_{1,1})-\mu_1(t_{1,0})]}] \cdots [1 - e^{-[\mu_d(t_{d,1})-\mu_d(t_{d,0}]}] \\
&= V_{1,1} \cdots V_{d,1}
\end{aligned}$$

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 $i \in \{2, \dots d\}$

$$F(t_{1,i}, \dots, t_{d,i}) \stackrel{d}{=} [1 - e^{-\mu_1(t_{1,i})}] \cdots [1 - e^{-\mu_d(t_{d,i})}]$$
$$= [1 - \prod_{j=1}^{i} \bar{V}_{1,j}] \cdots [1 - \prod_{j=1}^{i} \bar{V}_{d,j}].$$

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

$$(V_{1,1}, \dots V_{d,1}), \dots, (V_{1,h}, \dots V_{d,h})$$

such that we have (4.4).
Marginalizing in (4.4), we can apply Theorem 3.1 of Doksum (1974) to each F_i so we obtain that $F_i \sim \mathrm{NTR}(\mu_i)$ for some CRM μ_i that is stochastically continuous, almost surely non-decreasing and has the appropriate limit behaviour.
We observe that

$$\mu_1(t_j) - \mu_1(t_{j-1}) \stackrel{d}{=} -\log(1 - V_{1,j})$$
$$\vdots$$
$$\mu_d(t_j) - \mu_d(t_{j-1}) \stackrel{d}{=} -\log(1 - V_{d,j})$$

Hence $(\mu_1, \dots, \mu_d)$ defines a VCRM.

Proof of Theorem 12.

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 (1.16). In order to prove the theorem we use the next technical lemma.

Lemma 4. *Let $(\mu_1, \dots, \mu_d)$ be a d-variate CRM such that $\mu_1, \dots, \mu_d$ are not independent and let the Lévy intensity $\tilde{\nu}_d(\boldsymbol{s}, \mathrm{d}t)\mathrm{d}\boldsymbol{s}$ of $(\mu_1, \dots, \mu_d)$ be such that $\eta_t = \tilde{\nu}_d(\boldsymbol{x}, (0, t])$ 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 $\boldsymbol{q} = (q_1, \dots, q_d) \in \mathbb{N}^d$ are such that* $\max\{q_1, \dots, q_d\} \geq 1$ *and $\boldsymbol{r} = (r_1, \dots, r_d) \in (\mathbb{R}^+)^d$ are* *such that* $\min\{r_1,\dots,r_d\}\geq 1$, *then*

$$\begin{aligned}&\mathbb{E}\Big[e^{-r_1\mu_1(A_\varepsilon)-\cdots-r_d\mu_d(A_\varepsilon)}\Big(1-e^{-\mu_1(A_\varepsilon)}\Big)^{q_1}\cdots\Big(1-e^{-\mu_d(A_\varepsilon)}\Big)^{q_d}\Big]\\&=\varepsilon\int_{(\mathbb{R}^+)^d}e^{-\langle \boldsymbol{r},\boldsymbol{s}\rangle}(1-e^{-s_1})^{q_1}\cdots(1-e^{-s_d})^{q_d}\eta'_{t_0}(\boldsymbol{s})\mathrm{d}\boldsymbol{s}+o(\varepsilon)\end{aligned}$$

as $0<\varepsilon\to 0$, *with* $A_\varepsilon=(t_0-\varepsilon,t_0]$ *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

$$\begin{aligned}&\sum_{j_1=0}^{q_1}\cdots\sum_{j_d=0}^{q_d}\binom{q_1}{j_1}\cdots\binom{q_d}{j_d}(-1)^{j_1+\cdots+j_d}e^{-\big(\psi_{t_0}(\boldsymbol{r}+(j_1,\dots,j_d))-\psi_{t_0-\varepsilon}(\boldsymbol{r}+(j_1,\dots,j_d))\big)}\\&=e^{-\triangle_{t_0-\varepsilon}^{t_0}\psi_t(\boldsymbol{r})}+e^{-\triangle_{t_0-\varepsilon}^{t_0}\psi_t(\boldsymbol{r})}\Bigg\{\sum_{i=1}^{d}\sum_{j=1}^{q_i}\binom{q_i}{j}(-1)^je^{-\triangle_{t_0-\varepsilon}^{t_0}(\psi_t(\boldsymbol{r}+j\boldsymbol{e}_i)-\psi_t(\boldsymbol{r}))}\\&\quad+\sum_{\substack{i_1,i_2\in\{1,\dots,d\}\\i_1<i_2}}\sum_{j_1=1}^{q_{i_1}}\sum_{j_2=1}^{q_{i_2}}\binom{q_{i_1}}{j_1}\binom{q_{i_2}}{j_2}(-1)^{j_1+j_2}e^{-\triangle_{t_0-\varepsilon}^{t_0}\big(\psi_t(\boldsymbol{r}+j_1\boldsymbol{e}_{i_1}+j_2\boldsymbol{e}_{i_2})-\psi_t(\boldsymbol{r})\big)}\\&\quad+\cdots+\sum_{j_1=1}^{q_1}\cdots\sum_{j_d=1}^{q_d}\binom{q_1}{j_1}\cdots\binom{q_d}{j_d}(-1)^{\langle \boldsymbol{1},\boldsymbol{j}\rangle}e^{-\triangle_{t_0-\varepsilon}^{t_0}(\psi_t(\boldsymbol{r}+\boldsymbol{j})-\psi_t(\boldsymbol{r}))}\Bigg\}\end{aligned}\tag{4.8}$$

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

$$\begin{aligned}&e^{-\triangle_{t_0-\varepsilon}^{t_0}[\psi_t(\boldsymbol{r}+\boldsymbol{j})-\psi_t(\boldsymbol{r})]}=e^{-\int_{(\mathbb{R}^+)^d}e^{-\langle \boldsymbol{r},\boldsymbol{x}\rangle}(1-e^{-\langle \boldsymbol{j},\boldsymbol{s}\rangle})\triangle_{t_0-\varepsilon}^{t_0}\eta_t(\boldsymbol{s})\mathrm{d}\boldsymbol{s}}\\&=1-\varepsilon\int_{(\mathbb{R}^+)^d}e^{-\langle \boldsymbol{r},\boldsymbol{s}\rangle}(1-e^{-\langle \boldsymbol{j},\boldsymbol{s}\rangle})\eta'_{t_0}(\boldsymbol{s})\mathrm{d}\boldsymbol{s}+o(\varepsilon)\end{aligned}\tag{4.9}$$

Furthermore by the binomial theorem we get the next d identities

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

(2) $$\begin{aligned}&\sum_{\substack{i_1,i_2\in\{1,\dots,d\}\\i_1<i_2}}\sum_{j_1=1}^{q_{i_1}}\sum_{j_2=1}^{q_{i_2}}\binom{q_{i_1}}{j_1}\binom{q_{i_2}}{j_2}(-1)^{j_1+j_2}(1-e^{-j_1s_{i_1}-j_2s_{i_2}})\\&=\sum_{\substack{i_1,i_2\in\{1,\dots,d\}\\i_1<i_2}}\big\{(1-e^{-s_{i_1}})^{q_{i_1}}+(1-e^{-s_{i_2}})^{q_{i_2}}\\&\quad-(1-e^{-s_{i_1}})^{q_{i_1}}(1-e^{-s_{i_2}})^{q_{i_2}}\big\}\end{aligned}$$

$$\vdots$$

(d-1) $$\sum_{\substack{i_1,\dots,i_{d-1}\in\{1,\dots,d\}\\ i_1<\dots<i_{d-1}}} \sum_{j_1=1}^{q_{i_1}} \cdots \sum_{j_{d-1}=1}^{q_{i_{d-1}}} \binom{q_{i_1}}{j_1} \cdots \binom{q_{i_{d-1}}}{j_{d-1}} (-1)^{j_1+\cdots+j_{d-1}}$$
$$\times (1-e^{-j_1 s_{i_1}-\cdots-j_{d-1}s_{i_{d-1}}})$$
$$= \sum_{\substack{i_1,\dots,i_{d-1}\in\{1,\dots,d\}\\ i_1<\dots<i_{d-1}}} \Big\{(-1)^{d-1}\sum_{j=1}^{d-1}(1-e^{-s_{i_j}})^{q_{i_j}} +$$
$$(-1)^{d-2} \sum_{\substack{j_1,j_2\in\{i_1,\dots,i_{d-1}\}\\ j_1<j_2}} (1-e^{-s_{j_1}})^{q_{j_1}}(1-e^{-s_{j_2}})^{q_{j_2}} + \cdots$$
$$\cdots + (-1)(1-e^{-s_{i_1}})^{q_{i_1}}\cdots(1-e^{-s_{i_{d-1}}})^{q_{i_{d-1}}}\Big\}$$

(d) $$\sum_{j_1=1}^{q_1} \cdots \sum_{j_d=1}^{q_d} \binom{q_1}{j_1} \cdots \binom{q_d}{j_d} (-1)^{\langle \mathbf{1},\boldsymbol{j}\rangle}(1-e^{-\langle \boldsymbol{j},\boldsymbol{s}\rangle})$$
$$= (-1)^d \sum_{j=1}^{d}(1-e^{-s_j})^{q_j} +$$
$$(-1)^{d-1} \sum_{\substack{j_1,j_2\in\{1,\dots,d\}\\ j_1<j_2}} (1-e^{-s_{j_1}})^{q_{j_1}}(1-e^{-s_{j_2}})^{q_{j_2}} + \cdots$$
$$\cdots + (-1)(1-e^{-s_{i_1}})^{q_{i_1}}\cdots(1-e^{-s_{i_d}})^{q_{i_d}}$$

So we have that (4.8) becomes

$$e^{-\triangle_{t_0-\varepsilon}^{t_0}\psi_t(\boldsymbol{r})}\Bigg\{1+\sum_{i=1}^{d}\sum_{j=1}^{q_i}\binom{q_i}{j}(-1)^j$$
$$-\varepsilon\int_{(\mathbb{R}^+)^d} e^{-\langle \boldsymbol{r},\boldsymbol{s}\rangle}\sum_{i=1}^{d}\sum_{j=1}^{q_i}\binom{q_i}{j}(-1)^j(1-e^{-j_1 s_1})\eta'_{t_0}(\boldsymbol{s})\mathrm{d}\boldsymbol{s}$$
$$+\sum_{\substack{i_1,i_2\in\{1,\dots,d\}\\ i_1<i_2}}\sum_{j_1=1}^{q_{i_1}}\sum_{j_2=1}^{q_{i_2}}\binom{q_{i_1}}{j_1}\binom{q_{i_2}}{j_2}(-1)^{j_1+j_2}$$
$$-\varepsilon\int_{(\mathbb{R}^+)^d} e^{-\langle \boldsymbol{r},\boldsymbol{s}\rangle}\sum_{\substack{i_1,i_2\in\{1,\dots,d\}\\ i_1<i_2}}\sum_{j_1=1}^{q_{i_1}}\sum_{j_2=1}^{q_{i_2}}\binom{q_{i_1}}{j_1}\binom{q_{i_2}}{j_2}(-1)^{j_1+j_2}$$
$$\times(1-e^{-j_1 s_{i_1}-j_2 s_{i_2}})\eta'_{t_0}(\boldsymbol{s})\mathrm{d}\boldsymbol{s}+\cdots+\sum_{j_1=1}^{q_1}\cdots\sum_{j_d=1}^{q_d}\binom{q_1}{j_1}\cdots\binom{q_d}{j_d}(-1)^{\langle \mathbf{1},\boldsymbol{j}\rangle}$$
$$-\varepsilon\int_{(\mathbb{R}^+)^d} e^{-\langle \boldsymbol{r},\boldsymbol{s}\rangle}\sum_{j_1=1}^{q_1}\cdots\sum_{j_d=1}^{q_d}\binom{q_1}{j_1}\cdots\binom{q_d}{j_d}(-1)^{\langle \mathbf{1},\boldsymbol{j}\rangle}(1-e^{-\langle \boldsymbol{j},\boldsymbol{s}\rangle})\eta'_{t_0}(\boldsymbol{s})\mathrm{d}\boldsymbol{s}+o(\varepsilon)\Bigg\}$$

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

□

Similarly to the proof of Theorem 1 we use the next Lemma to simplify the calculations in the proof of the Theorem at hand.

Lemma 5. *In the setting of Theorem 1 suppose the censored to the right data $\mathscr{D}$ is comprised of a sole observation t_1 with frequencies $\boldsymbol{n}^e = (n_1^e,\ldots,n_d^e)$ and $\boldsymbol{n}^c = (n_1^c,\ldots,n_d^c)$. Let $t < t_1$; then*

$$
\mathbb{E}\left[e^{-\lambda_1\mu_1(0,t]-\ldots-\lambda_d\mu_d(0,t]}|\mathscr{D}_d\right] = \lim_{\varepsilon\to 0}\frac{\mathbb{E}\left[e^{-(\lambda+n^c+n^e)\mu(0,t]}\right]}{\mathbb{E}\left[e^{-(n^c+n^e)\mu(0,t]}\right]}
$$

Proof.

$$
\begin{aligned}
&\mathbb{E}\left[e^{-\lambda_1\mu_1(0,t]-\ldots-\lambda_d\mu_d(0,t]}|\mathscr{D}_d\right]\\
&= \lim_{\varepsilon\to 0}\frac{\mathbb{E}\left[\prod_{i=1}^d e^{-\lambda_i\mu_i(0,t]}e^{-n_i^c\mu_i(0,t_1]}\left(e^{-\mu_i(0,t_1-\varepsilon]}-e^{-\mu_i(0,t_1]}\right)^{n_i^e}\right]}{\mathbb{E}\left[\prod_{i=1}^d e^{-n_i^c\mu_i(0,t_1]}\left(e^{-\mu_i(0,t_1-\varepsilon]}-e^{-\mu_i(0,t_1]}\right)^{n_i^e}\right]}\\
&= \lim_{\varepsilon\to 0}\left\{\frac{\mathbb{E}\left[\prod_{i=1}^d e^{-\lambda_i\mu_i(0,t]-n_i^c\mu_i(0,t]-n_i^e\mu_i(0,t]}\right]}{\mathbb{E}\left[e^{-n_i^c\mu_i(0,t]-n_i^e\mu_i(0,t]}\right]}\right.\\
&\qquad\left.\times\frac{\mathbb{E}\left[\prod_{i=1}^d e^{-n_i^c\mu_i(t,t_1]}\left(e^{-\mu_i(t,t_1-\varepsilon]}-e^{-\mu_i(t,t_1]}\right)^{n_i^e}\right]}{\mathbb{E}\left[\prod_{i=1}^d e^{-n_i^c\mu_i(t,t_1]}\left(e^{-\mu_i(t,t_1-\varepsilon]}-e^{-\mu_i(t,t_1]}\right)^{n_i^e}\right]}\right\}\\
&= \frac{\mathbb{E}\left[\prod_{i=1}^d e^{-(\lambda_i+n_i^c+n_i^e)\mu_i(0,t]}\right]}{\mathbb{E}\left[\prod_{i=1}^d e^{-(n_i^c+n_i^e)\mu_i(0,t]}\right]} = e^{-(\psi_t(\boldsymbol{\lambda}+\boldsymbol{n}^c+\boldsymbol{n}^e)-\psi_t(\boldsymbol{n}^c+\boldsymbol{n}^e))}.
\end{aligned}
$$

□

We observe that if ψ_t is the Laplace exponent associated to the Lévy measure $\tilde{\nu}_d$ then $\psi_t^{(\boldsymbol{k})}(\boldsymbol{\lambda}) = \psi_t(\boldsymbol{\lambda}+\boldsymbol{k})-\psi_t(\boldsymbol{k})$ is the Laplace exponent associated to $e^{-\langle \boldsymbol{k},\boldsymbol{s}\rangle}\tilde{\nu}_d(\mathrm{d}\boldsymbol{s},\mathrm{d}x)$, with this in mind we have that, in the following, without loss of generality for calculations of the posterior Laplace exponent as in the previous Lemma it suffices to consider evaluation of the exponent in a time t which is greater than all the survival times in the survival data $\mathscr{D}_d$. Define

$$\Gamma_{\mathscr{D}_d,\varepsilon} = \bigcap_{i=1}^{d}\bigcap_{j=1}^{k}\Big\{((T_1^{(i)},J_1^{(i)},\ldots,T_{n_1}^{(i)},J_{n_1}^{(i)})\ :\ m_i^c\left(\{T_{(j)}\}\right)=n_{i,j}^c\ ,$$
$$m_i^e\left((T_{(j)}-\varepsilon,T_{(j)}]\right)=n_{i,j}^e\Big\}$$

so that

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

We observe that defining $T_{(0)}=0$ and $\bar{n}_{i,k+1}^e=0$ for $i\in\{1,\ldots,d\}$

$$\begin{aligned}
&\mathbb{E}\Big[e^{-\lambda_1\mu_1(0,t]-\cdots\lambda_d\mu_d(0,t]}\mathbb{1}_{\Gamma_{\boldsymbol{D},\varepsilon}}(\boldsymbol{D})\Big]\\
&=\mathbb{E}\left[\prod_{i=1}^{d}e^{-\lambda_i\mu_i(0,t]}\prod_{j=1}^{k}e^{-n_{i,j}^c\mu_i(0,T_{(j)}]-n_{i,j}^e\mu_i(0,T_{(j)}-\varepsilon]}\left(1-e^{-\mu_i(T_{(j)}-\varepsilon,T_{(j)}]}\right)^{n_{i,j}^e}\right]\\
&=\mathbb{E}\Bigg[\prod_{i=1}^{d}e^{-\lambda_i\mu_i(T_{(k)},t]}\prod_{j=1}^{k}\Big[e^{-\lambda_i\mu_i(T_{(j-1)},T_{(j)}-\varepsilon]-\lambda_i\mu_i(T_{(j)}-\varepsilon,T_{(j)}]}\\
&\quad\times e^{-n_{i,j}^c\sum_{r=1}^{j}\left(\mu_i(T_{(r)}-\varepsilon,T_{(r)}]+\mu_i(T_{(r-1)},T_{(r)}-\varepsilon]\right)-n_{i,j}^e\sum_{r=1}^{j}\mu_i(T_{(r-1)},T_{(r)}-\varepsilon]}\\
&\quad\times e^{-n_{i,j}^e\sum_{r=1}^{j-1}\mu_i(T_{(r)}-\varepsilon,T_{(r)}]}\left(1-e^{-\mu_i(T_{(j)}-\varepsilon,T_{(j)}]}\right)^{n_{i,j}^e}\Big]\Bigg]\\
&=\mathbb{E}\Bigg[\prod_{i=1}^{d}e^{-\lambda_i\mu_i(T_{(k)},t]-\sum_{j=1}^{k}n_{i,j}^c\sum_{r=1}^{j}\left(\mu_i(T_{(r)}-\varepsilon,T_{(r)}]+\mu_i(T_{(r-1)},T_{(r)}-\varepsilon]\right)}\\
&\quad\times e^{-\sum_{j=1}^{k}n_{i,j}^e\sum_{r=1}^{j}\mu_i(T_{(r-1)},T_{(r)}-\varepsilon]-\sum_{j=1}^{k}n_{i,j}^e\sum_{r=1}^{j-1}\mu_i(T_{(r)}-\varepsilon,T_{(r)}]}\\
&\quad\times\prod_{j=1}^{k}e^{-\lambda_i\mu_i(T_{(j-1)},T_{(j)}-\varepsilon]-\lambda_i\mu_i(T_{(j)}-\varepsilon,T_{(j)}]}\left(1-e^{-\mu_i(T_{(j)}-\varepsilon,T_{(j)}]}\right)^{n_{i,j}^e}\Bigg]\\
&=\mathbb{E}\Bigg[\prod_{i=1}^{d}e^{-\lambda_i\mu_i(T_{(k)},t]}\prod_{j=1}^{k}e^{-\left(\lambda_i+\bar{n}_{i,j}^c+\bar{n}_{i,j+1}^e\right)\mu_i(T_{(j)}-\varepsilon,T_{(j)}]}\left(1-e^{-\mu_i(T_{(j)}-\varepsilon,T_{(j)}]}\right)^{n_{i,j}^e}\\
&\quad\times e^{-\lambda_i\mu_i(T_{(j-1)},T_{(j)}-\varepsilon]-\bar{n}_{i,j}^c\mu_i(T_{(j-1)},T_{(j)}-\varepsilon]-\bar{n}_{i,j}^e\mu_i(T_{(j-1)},T_{(j)}-\varepsilon]}\Bigg]
\end{aligned}$$

So defining

$$I_{1,\varepsilon} = \prod_{j=1}^{k}\prod_{i=1}^{d}\left\{e^{-\left[\lambda_i+\bar{n}^c_{i,j}+\bar{n}^e_{i,j+1}\right]\mu_i(T_{(j)}-\varepsilon,T_{(j)}]}\left(1-e^{-\mu_i(T_{(j)}-\varepsilon,T_{(j)}]}\right)^{n^e_{i,j}}\right\}$$

$$I_{2,\varepsilon} = \prod_{i=1}^{d} e^{-\lambda_i\mu_i(T_{(k)},t]}\prod_{j=1}^{k}\left\{e^{-\lambda_i\mu_i(T_{(j-1)},T_{(j)}-\varepsilon]-(\bar{n}^c_{i,j}+\bar{n}^e_{i,j})\mu_i(T_{(j-1)},T_{(j)}-\varepsilon]}\right\}$$

We get from the independence property of CRM's that

$$\mathbb{E}\left[e^{-\lambda_1\mu_1(0,t]-\cdots-\lambda_d\mu_d(0,t]}\mathbb{1}_{\Gamma_{\boldsymbol{D},\varepsilon}}(\boldsymbol{D})\right] = \mathbb{E}[I_{1,\varepsilon}]\,\mathbb{E}[I_{2,\varepsilon}] \tag{4.10}$$

We observe that for $r_i = \lambda_i + \bar{n}^c_{i,j} + \bar{n}^e_{i,j+1}$, $i \in \{1,\dots,d\}$ we have that $\min\{r_1,\dots,r_d\} \geq 1$ and for $j \in \{1,\dots,k\}$ such that $T_{(j)}$ is an exact observation we have that $\max\{n_{1,j},\dots,n_{d,j}\} \geq 1$ so Lemma 4 can be applied yielding

$$\mathbb{E}\left[\prod_{i=1}^{d} e^{-\left(\lambda_i+\bar{n}^c_{i,j}+\bar{n}^e_{i,j+1}\right)\mu_i(T_{(j)}-\varepsilon,T_{(j)}]}\left(1-e^{-\mu_i(T_{(j)}-\varepsilon,T_{(j)}]}\right)^{n^e_{i,j}}\right]$$
$$= \varepsilon\int_{(\mathbb{R}^+)^d}\prod_{i=1}^{d}\left\{e^{-\left(\lambda_i+\bar{n}^c_{i,j}+\bar{n}^e_{i,j+1}\right)s_i}(1-e^{-s_i})^{n^e_{i,j}}\right\}\eta'_t T_{(j)}(\boldsymbol{s})\mathrm{d}\boldsymbol{s} + o(\varepsilon) \tag{4.11}$$

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

$$\lim_{\varepsilon\to 0}\mathbb{E}\left[\prod_{i=1}^{d} e^{-\left(\lambda_i+\bar{n}^c_{i,j}+\bar{n}^e_{i,j+1}\right)\mu_i(T_{(j)}-\varepsilon,T_{(j)}]}\left(1-e^{-\mu_i(T_{(j)}-\varepsilon,T_{(j)}]}\right)^{n^e_{i,j}}\right]$$
$$= \lim_{\varepsilon\to 0}\mathbb{E}\left[\prod_{i=1}^{d} e^{-\left(\lambda_i+\bar{n}^c_{i,j}+\bar{n}^e_{i,j+1}\right)\mu_i(T_{(j)}-\varepsilon,T_{(j)}]}\right] = 1 \tag{4.12}$$

From (4.11), (4.12) and the independence property of CRM's we obtain

$$\lim_{\varepsilon\to 0}\mathbb{E}[I_{1,\varepsilon}] =$$
$$\lim_{\varepsilon\to 0}\prod_{j\in I^{(e)}}\left\{\varepsilon\int_{(\mathbb{R}^+)^d}\prod_{i=1}^{d}\left\{e^{-\left(\lambda_i+\bar{n}^c_{i,j}+\bar{n}^e_{i,j+1}\right)s_i}(1-e^{-s_i})^{n^e_{i,j}}\right\}\eta'_{T_{(j)}}(\boldsymbol{s})\mathrm{d}\boldsymbol{s} + o(\varepsilon)\right\}$$

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

$$\lim_{\varepsilon\to 0}\mathbb{E}[I_{2,\varepsilon}] = e^{-\left(\psi_t(\boldsymbol{\lambda})-\psi_{T_{(k)}}(\boldsymbol{\lambda})\right)}$$
$$\times\prod_{j=1}^{k}\left\{e^{-\left(\psi_{T_{(j)}}(\boldsymbol{\lambda}+\bar{\boldsymbol{n}}_j^c+\bar{\boldsymbol{n}}_j^e)-\psi_{T_{(j-1)}}(\boldsymbol{\lambda}+\bar{\boldsymbol{n}}_j^c+\bar{\boldsymbol{n}}_j^e)\right)-\left(\psi_{T_{(j)}}(\bar{\boldsymbol{n}}_j^c+\bar{\boldsymbol{n}}_j^e)-\psi_{T_{(j-1)}}(\bar{\boldsymbol{n}}_j^c+\bar{\boldsymbol{n}}_j^e)\right)}\right\}$$

So by (4.10), (4.12) and (4.11) we get that

$$\lim_{\varepsilon\to 0}\mathbb{E}\left[e^{-\lambda_1\mu_1(0,t]-\cdots-\lambda_d\mu_d(0,t]}\mathbb{1}_{\Gamma_{\boldsymbol{D},\varepsilon}}(\mathscr{D}_d)\right] = e^{-\triangle_{T_{(k)}}^{t}\psi_t(\boldsymbol{\lambda})-\sum_{j=1}^{k}\triangle_{T_{(j-1)}}^{T_{(j)}}\psi_t(\boldsymbol{\lambda}+\bar{\boldsymbol{n}}_j^c+\bar{\boldsymbol{n}}_j^e)}$$
$$\times\prod_{j\in I^{(e)}}\lim_{\varepsilon\to 0}\left\{\varepsilon\int_{(\mathbb{R}^+)^d}\prod_{i=1}^{d}\left\{e^{-\left(\lambda_i+\bar{n}_{i,j}^c+\bar{n}_{i,j+1}^e\right)s_i}(1-e^{-s_i})^{n_{i,j}^e}\right\}\eta'_{T_{(j)}}(\boldsymbol{s})\mathrm{d}\boldsymbol{s}+o(\varepsilon)\right\}$$
$$\times e^{-\sum_{j=1}^{k}\triangle_{T_{(j)}}^{T_{(j-1)}}\psi_t(\bar{\boldsymbol{n}}_j^c+\bar{\boldsymbol{n}}_j^e)}$$

And similarly

$$\lim_{\varepsilon\to 0}\mathbb{P}\left[\mathscr{D}_d\in\Gamma_{\mathscr{D}_d,\varepsilon}\right] = e^{-\sum_{j=1}^{k}\triangle_{T_{(j-1)}}^{T_{(j)}}\psi_t\left(\bar{\boldsymbol{n}}_j^c+\bar{\boldsymbol{n}}_j^e\right)}$$
$$\times\prod_{j\in I^{(e)}}\lim_{\varepsilon\to 0}\left\{\varepsilon\int_{(\mathbb{R}^+)^d}\prod_{i=1}^{d}\left\{e^{-(\bar{n}_{i,j}^c+\bar{n}_{i,j+1}^e)s_i}(1-e^{-s_i})^{n_{i,j}^e}\right\}\eta_t'T_{(j)}(\boldsymbol{s})\mathrm{d}\boldsymbol{s}+o(\varepsilon)\right\} \tag{4.13}$$

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

$$\mathbb{E}\left[e^{-\lambda_1\mu_1(0,t]-\cdots-\lambda_d\mu_d(0,t]}|\mathscr{D}_d\right] = \lim_{\varepsilon\to 0}\frac{\mathbb{E}\left[e^{-\lambda_1\mu_1(0,t]-\cdots-\lambda_d\mu_d(0,t]}\mathbb{1}_{\Gamma_{\mathscr{D}_d,\varepsilon}}(\boldsymbol{D})\right]}{\mathbb{P}\left[\boldsymbol{D}\in\Gamma_{\mathscr{D}_d,\varepsilon}\right]}$$
$$= e^{-\sum_{j=1}^{k+1}\triangle_{T_{(j-1)}}^{T_{(j)}}\left(\psi_t(\boldsymbol{\lambda}+\bar{\boldsymbol{n}}_j^c+\bar{\boldsymbol{n}}_j^e)-\psi_t(\bar{\boldsymbol{n}}_j^c+\bar{\boldsymbol{n}}_j^e)\right)}$$
$$\prod_{j\in I^{(e)}}\lim_{\varepsilon\to 0}\left\{\frac{\varepsilon\int_{(\mathbb{R}^+)^d}\prod_{i=1}^{d}\left\{e^{-\left(\lambda_i+\bar{n}_{i,j}^c+\bar{n}_{i,j+1}^e\right)s_i}(1-e^{-s_i})^{n_{i,j}^e}\right\}\eta'_{T_{(j)}}(\boldsymbol{s})\mathrm{d}\boldsymbol{s}+o(\varepsilon)}{\varepsilon\int_{(\mathbb{R}^+)^d}\prod_{i=1}^{d}\left\{e^{-\left(\bar{n}_{i,j}^c+\bar{n}_{i,j+1}^e\right)s_i}(1-e^{-s_i})^{n_{i,j}^e}\right\}\eta'_{T_{(j)}}(\boldsymbol{s})\mathrm{d}\boldsymbol{s}+o(\varepsilon)}\right\}$$
$$= e^{-\sum_{j=1}^{k+1}\int_{(\mathbb{R}^+)^d\times(T_{(j-1)},T_{(j)}]}\left(1-e^{-\langle\boldsymbol{\lambda},\boldsymbol{s}\rangle}\right)e^{-\langle\bar{\boldsymbol{n}}_j^c+\bar{\boldsymbol{n}}_j^e,\boldsymbol{s}\rangle}\nu(\mathrm{d}\boldsymbol{s},\mathrm{d}u)}$$
$$\prod_{j\in I^{(e)}}\left\{\frac{\int_{(\mathbb{R}^+)^d}\prod_{i=1}^{d}\left\{e^{-\left(\lambda_i+\bar{n}_{i,j}^c+\bar{n}_{i,j+1}^e\right)s_i}(1-e^{-s_i})^{n_{i,j}^e}\right\}\eta'_{T_{(j)}}(\boldsymbol{s})\mathrm{d}\boldsymbol{s}}{\int_{(\mathbb{R}^+)^d}\prod_{i=1}^{d}\left\{e^{-\left(\bar{n}_{i,j}^c+\bar{n}_{i,j+1}^e\right)s_i}(1-e^{-s_i})^{n_{i,j}^e}\right\}\eta'_{T_{(j)}}(\boldsymbol{s})\mathrm{d}\boldsymbol{s}}\right\} \tag{4.14}$$

Proof of Corollary 4.

In equation (4.14) above take $\boldsymbol{\lambda} = (1,\dots,1)$ and using the discussion after Lemma 5 replace M with M_t to obtain the result.

Proof of Corollary 5.

From equation (4.13) in the proof of Theorem 12 we obtain the related likelihood.

Proof of Proposition 11.

In the setting of Theorem 12.

$$\begin{aligned}
\psi_t^{\circ j}(\boldsymbol{\lambda}) &= \int_{\left(t\wedge T_{(j-1)},t\wedge T_{(j)}\right)\times(\mathbb{R}^+)^d} (1-e^{-\langle\boldsymbol{\lambda},\boldsymbol{s}\rangle})e^{-(\bar{\boldsymbol{n}}_j^c+\bar{\boldsymbol{n}}_j^e)s}\tilde{\nu}_d(\boldsymbol{s},\mathrm{d}x)\mathrm{d}\boldsymbol{s} \\
&= \int_{\left(t\wedge T_{(j-1)},t\wedge T_{(j)}\right)\times(\mathbb{R}^+)^d} (1-e^{-\langle\boldsymbol{\lambda}+\bar{\boldsymbol{n}}_j^c+\bar{\boldsymbol{n}}_j^e,\boldsymbol{s}\rangle})\tilde{\nu}_d(\boldsymbol{s},\mathrm{d}x)\mathrm{d}\boldsymbol{s} \\
&\quad - \int_{\left(t\wedge T_{(j-1)},t\wedge T_{(j)}\right)\times(\mathbb{R}^+)^d} (1-e^{-\langle\bar{\boldsymbol{n}}_j^c+\bar{\boldsymbol{n}}_j^e,\boldsymbol{s}\rangle})\tilde{\nu}_d(\boldsymbol{s},\mathrm{d}x)\mathrm{d}\boldsymbol{s} \\
&= \left(\psi_{t\wedge T_{(j)}}(\boldsymbol{\lambda}+\bar{\boldsymbol{n}}_j^c+\bar{\boldsymbol{n}}_j^e) - \psi_{t\wedge T_{(j-1)}}(\boldsymbol{\lambda}+\bar{\boldsymbol{n}}_j^c+\bar{\boldsymbol{n}}_j^e)\right) \\
&\quad - \left(\psi_{t\wedge T_{(j)}}(\bar{\boldsymbol{n}}_j^c+\bar{\boldsymbol{n}}_j^e) - \psi_{t\wedge T_{(j-1)}}(\bar{\boldsymbol{n}}_j^c+\bar{\boldsymbol{n}}_j^e)\right).
\end{aligned}$$

Proof of Proposition 12.

We will use the next lemma.

Lemma 6. *Let $m \in \mathbb{N}$, $\boldsymbol{q}_1,\dots,\boldsymbol{q}_m \in (\mathbb{R}^+)^d$ distinct of $\mathbf{0} = (0,\dots,0)$, $\boldsymbol{q} \in (\mathbb{R}^+)^d$ and $\emptyset \neq I = \{i_1\dots,i_{|I|}\} \subset \{1,\dots,m\}$. We denote $I\setminus 1 = \{i_2,\dots,i_{|I|}\}$. Then*

$$\begin{aligned}
&\int_{(\mathbb{R}^+)^d\times(0,t]} e^{-\langle\boldsymbol{q},\boldsymbol{s}\rangle}\prod_{l\in I}\left(1-e^{-\langle\boldsymbol{q}_l,\boldsymbol{s}\rangle}\right)\tilde{\nu}_d(\mathrm{d}\boldsymbol{s},\mathrm{d}x) \\
&= \sum_{S\subset I\setminus 1}(-1)^{\#(S)}\left(\psi_t(\boldsymbol{q}_1+\sum_{l\in S}\boldsymbol{q}_l+\boldsymbol{q}) - \psi_t(\sum_{l\in S}\boldsymbol{q}_l+\boldsymbol{q})\right)
\end{aligned}$$

Proof.

$$
\begin{aligned}
&\int_{(\mathbb{R}^+)^d\times(0,t]} e^{-\langle \boldsymbol{q},\boldsymbol{s}\rangle}\prod_{l\in I}\left(1-e^{-\langle \boldsymbol{q}_l,\boldsymbol{s}\rangle}\right)\tilde{\nu}_d(\mathrm{d}\boldsymbol{s},\mathrm{d}x)\\
&=\int_{(\mathbb{R}^+)^d\times(0,t]} e^{-\langle \boldsymbol{q},\boldsymbol{s}\rangle}(1-e^{-\langle \boldsymbol{q}_1,\boldsymbol{s}\rangle})\prod_{l\in I\setminus 1}\left(1-e^{-\langle \boldsymbol{q}_l,\boldsymbol{s}\rangle}\right)\tilde{\nu}_d(\mathrm{d}\boldsymbol{s},\mathrm{d}x)\\
&=\int_{(\mathbb{R}^+)^d\times(0,t]} e^{-\langle \boldsymbol{q},\boldsymbol{s}\rangle}(1-e^{-\langle \boldsymbol{q}_1,\boldsymbol{s}\rangle})\sum_{S\subset I\setminus 1}(-1)^{\#S}e^{-\langle \sum_{l\in S}\boldsymbol{q}_l,\boldsymbol{s}\rangle}\tilde{\nu}_d(\mathrm{d}\boldsymbol{s},\mathrm{d}x)\\
&=\sum_{S\subset I\setminus 1}(-1)^{\#S}\int_{(\mathbb{R}^+)^d} e^{-\langle \boldsymbol{q}+\sum_{l\in S}\boldsymbol{q}_l,\boldsymbol{s}\rangle}(1-e^{-\langle \boldsymbol{q}_1,\boldsymbol{s}\rangle})\tilde{\nu}_d(\mathrm{d}\boldsymbol{s},\mathrm{d}x)\\
&=\sum_{S\subset I\setminus 1}(-1)^{\#(S)}\Big(\psi_t(\boldsymbol{q}_1+\sum_{l\in S}\boldsymbol{q}_l+\boldsymbol{q})-\psi_t(\sum_{l\in S}\boldsymbol{q}_l+\boldsymbol{q})\Big)
\end{aligned}
$$

□

The proposition follows with I as in the hypothesis and $\boldsymbol{q}_i=\boldsymbol{e}_i$ for $i\in\{1,\dots,d\}$ in the above lemma.

Chapter 5

Generalized Additive Neutral to the Right Regression

5.1 Survival regression

In survival analysis it is often the case that the events of interest we want to analyse involve information that can be suitably quantified in a control variable which we refer to as *covariate*. Usually we want to use the covariates for prognosis purposes, i.e. their effect on the related survival function. To analyse the dependence between the variable of interest and the associated covariates we require a regression model. Regression in a NTR setting regression was investigated by Kim and Lee (2003), where they considered a Cox regression approach for NTR distributions.

Definition 24. (Kim and Lee (2003)) Let μ be a CRM and $T \sim \mathrm{NTR}(\mu)$ so that the survival function of Y is $S(t) = e^{-\mu(0,t]}$; furthermore let there be a vector of covariates $\boldsymbol{X} \in \mathbb{R}^d$ for some $d \in \mathbb{N}$. We say that $Y|\boldsymbol{X}$ follows a Cox NTR distribution if

$$S_{\mathrm{Cox}}(t) = \mathbb{P}[T > t\,|\boldsymbol{X}] = S(t)^{e^{\langle \boldsymbol{\beta}, \boldsymbol{X} \rangle}} = e^{-e^{\langle \boldsymbol{\beta}, \boldsymbol{X} \rangle}\mu(0,t]}$$

for some $\boldsymbol{\beta} \in \mathbb{R}^d$.

The Cox regression approach can be too restrictive in a variety of settings even if we use the flexible NTR distribution; for example, a shortcoming of this model is that it induces proportional hazards when we vary the covariates and therefore it does not allow for the crossing of the survival functions related to r.v.'s with different covariates outside of the set $\{t \in \mathbb{R}^+ : S_{\mathrm{Cox}}(t) = 0 \text{ or } S_{\mathrm{Cox}}(t) = 1\}$. Indeed, if p_1, p_2 are indexes associated to survival times Y_{p_1}, Y_{p_2} with respective covariates $\boldsymbol{X}^{(p_1)} \neq \boldsymbol{X}^{(p_2)}$ and survival functions $S^{(p_1)}_{\mathrm{Cox}}$, $S^{(p_2)}_{\mathrm{Cox}}$

given as in the definition above for a NTR distribution, then

$$S_{\text{Cox}}^{(p_1)}(t) - S_{\text{Cox}}^{(p_2)}(t) = S_{\text{Cox}}^{(p_1)}(t)(1 - S(t)^r)$$

where $r = e^{\langle \beta, \boldsymbol{X}_{p_2} \rangle} - e^{\langle \beta, \boldsymbol{X}_{p_1} \rangle}$. The quantity above can only be zero if $t = 0$ or $t = 1$.

5.2 Generalized additive NTR regression model

We observe that the d-variate multiple-sample NTR model of Chapter 4, displayed in equation (4.2), can be thought of as a regression model where the covariates denote the label of membership to one of d different populations. We will focus on the use of VCRM's to build a flexible Bayesian nonparametric regression model that can recover the multiple-sample and Cox NTR settings.

Definition 25. Let $n, m, d, b \in \mathbb{N} \setminus \{0\}$ and $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_d)$ a d-variate VCRM such that

$$\lim_{t \to \infty} \mu_i(0, t] \overset{\text{a.s.}}{=} \infty$$

for any $i \in \{1, \ldots, d\}$. We say that a collection of random elements

$$\{Y_i, \boldsymbol{X}_i\}_{i=1}^n$$

with $Y_i \in \mathbb{R}^+$, $\boldsymbol{X}_i = (X_{i,1}, \ldots, X_{i,m}) \in \mathbb{R}^m$, follows a generalized additive NTR regression model if

$$S(\boldsymbol{t}) = \mathbb{P}[Y_1 > t_1, \ldots, Y_n > t_n \,|\, \boldsymbol{\mu}, \boldsymbol{\beta}, \boldsymbol{X}] = \prod_{i=1}^{n} e^{-f_1(\boldsymbol{\beta}, \boldsymbol{X}_i)\mu_1(0,t_i] - \ldots - f_d(\boldsymbol{\beta}, \boldsymbol{X}_i)\mu_d(0,t_i]} \tag{5.1}$$

where $(t_1, \ldots, t_n) \in (\mathbb{R}^+)^n$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_b) \in \mathbb{R}^b$ and $f_i : \mathbb{R}^b \times \mathbb{R}^m \to \mathbb{R}^+$ for $i \in \{1, \ldots, d\}$.

We observe that the generalized additive NTR regression model can be seen as NTR distribution conditionally on the covariates $\boldsymbol{X}_i$

$$Y_i | \boldsymbol{\beta}, \boldsymbol{X}_i, \boldsymbol{\mu} \overset{\text{ind}}{\sim} \text{NTR}\left(\sum_{j=1}^{d} f_j(\boldsymbol{\beta}, \boldsymbol{X}_i) \mu_j \right); \tag{5.2}$$

such remark is useful as it allows us to use NTR results for our model. The Cox NTR model of Kim and Lee (2003) is recovered if $d = 1$ and $f_1(\boldsymbol{\beta}, \boldsymbol{X}) = e^{\langle \boldsymbol{\beta}, \boldsymbol{X} \rangle}$. However, when considering $d > 1$ regressor functions f in our model the survival functions for different covariate values can cross each other at any point $t \in \mathbb{R}^+$. Indeed, let S_{p_1} and S_{p_2} be the survival functions of r.v.'s Y_{p_1}, Y_{p_2} in the generalized additive NTR regression model with respective covariates $\boldsymbol{X}_{p_1} \neq \boldsymbol{X_{p_2}}$, then

$$S_{p_1}(t) - S_{p_2}(t) = S_{p_1}(t)\left(1 - \prod_{i=1}^{d} e^{-r_i \mu_i(0,t]}\right)$$

with $r_i = f_i(\boldsymbol{\beta}, \boldsymbol{X}_{p_2}) - f_i(\boldsymbol{\beta}, \boldsymbol{X}_{p_1})$. For example, if $d = 2$, the survival functions cross if the curves given by $\{(e^{-\mu_1(0,t]}, e^{-\mu_2(0,t]}) \; : \; t \in \mathbb{R}^+\}$ and $\{(t, t^c) \; : \; t \in \mathbb{R}^+, \; c = -r_1/r_2\}$ cross. In Figure 5.1 we illustrate such case.

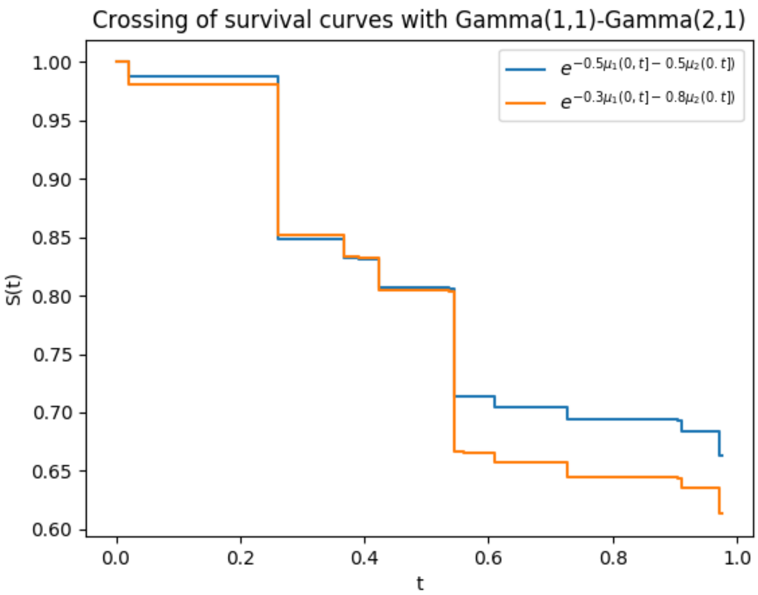


Fig. 5.1 Crossing of survival functions when considering a bidimensional CoRM with Gamma directing Lévy measure and independent Gamma(1,1) scores.

The multiple-sample NTR model, (4.3), can also be recovered by considering a covariate $\boldsymbol{X} \in \{0,1\}^m$ such that $X_i = 1$ if Y_i belongs to sample i and $X_i = 0$ otherwise, and choosing

$$f_i(\boldsymbol{\beta}, \boldsymbol{X}) = \mathbb{1}_{\{X_i = 1\}}. \tag{5.3}$$

In Section 5.3 we address the posterior consistency of the multiple-sample NTR model and highlight some interesting comparisons with our more general model (5.1). We have motivated our model as a generalization of the Cox regression and multiple-sample information models but it can also be viewed as a competing risks model. We assume d causes for the event of interest to happen and define the survival function for the i-th cause to be

$$\tilde{S}_i(t) = e^{-f_i(\boldsymbol{\beta},\boldsymbol{X})\mu_i(0,t]}$$

for $i \in \{1,\ldots,d\}$. The overall survival function is then given by the model in (5.2). This suggests a simulation scheme for our model. We sample $\tilde{Y}_i$ as survival times according to the survival function $\tilde{S}_i(t)$ and set

$$Y = \min\{\tilde{Y}_1,\ldots,\tilde{Y}_d\}. \tag{5.4}$$

The CoRM is an interesting choice for the underlying VCRM $\boldsymbol{\mu}$ in the generalized additive NTR regression model, (5.2). In this case the series representation of $\boldsymbol{\mu} = (\mu_1,\ldots,\mu_d)$ in (2.12) leads to the identity

$$T \sim \mathrm{NTR}\left(\sum_{j=1}^{\infty}\left(\sum_{i=1}^{d} f_i(\boldsymbol{\beta},\boldsymbol{X})m_{i,j}\right)w_j\delta_{u_j}\right) \tag{5.5}$$

which can be seen as a CoRM on $\mathbb{R}^m$ with directing Lévy measure $\nu^\star$ and score distribution given by the r.v.'s $\sum_{i=1}^{d} f_i(\boldsymbol{\beta},\boldsymbol{X})m_{i,j}$. In this form, each score is a random linear combination of the basis functions $f_1,\ldots,f_d$. For example, if we take $f_i(\boldsymbol{\beta},\boldsymbol{X}) = e^{\beta_i X_i}$ and the score distribution h, corresponding to the i.i.d. weights $\{(m_{1,j},\ldots,m_{d,j})\}_{j=1}^{\infty}$, to be a multivariate LogNormal, as in example (10), then we get a NTR distribution as follows

$$Y \sim \mathrm{NTR}\left(\sum_{j=1}^{\infty}\left(\sum_{i=1}^{d} e^{\beta_i X_i+\varepsilon_{i,j}}\right)w_j\delta_{u_j}\right)$$

where $\{(\varepsilon_{1,j},\ldots,\varepsilon_{d,j})\}_{j=1}^{\infty}$ are i.i.d. multivariate Normal r.v.'s. When we consider possibly censored to the right data, the posterior characterization of the generalized additive NTR model can be calculated. First, we fix the notation for the posterior calculation of (5.1) to be established in Theorem 13.

Definition 26. Let

$$\mathscr{D} = \{T_i,J_i\}_{i=1}^{n}$$

be survival data with possibly censored to the right observations. Given a sequence of $m-$dimensional covariates $\{\boldsymbol{X}_i\}_{i=1}^{n}$ and a $n \times m$ matrix $\hat{\boldsymbol{X}}$ with rows consisting of the covariate sequence, we say that

$$\mathscr{D}_{\hat{\boldsymbol{X}}} = \{T_i,J_i,\boldsymbol{X}_i\}_{i=1}^{n}$$

is survival data censored to the right with covariates.

For the sequence $T_1,\dots,T_n$ we define the $k \le n$ order statistics (without repetition) to be $T_{(1)} < \dots < T_{(k)}$. Let$T_{(0)} = 0$ and $T_{(k+1)} = \infty$. We define n_j^c and n_j^e to be the number of censored and exact observations at time $T_{(j)}$ respectively. The matrix $\hat{\boldsymbol{X}}$ is defined to have as rows the m-dimensional covariate vectors $\{\boldsymbol{X}_i\}_{i=1}^n$. We define sets

$$I_j^{(e)} = \{l : T_l = T_{(j)} \text{ and } J_l = 1\} \quad ; \quad I_j^{(c)} = \{l : T_l = T_{(j)} \text{ and } J_l = 0\},$$

with $j \in \{1,\dots,k\}$ and

$$I^{(e)} = \{l : T_l \text{ is an exact observation}\}.$$

We also define functions

$$h_{i,j}^{(e)}(\boldsymbol{z},\hat{\boldsymbol{Z}}) = \sum_{l \in I_j^{(e)}} f_i(\boldsymbol{z},\boldsymbol{Z}_l) \quad ; \quad h_{i,j}^{(c)}(\boldsymbol{z},\hat{\boldsymbol{Z}}) = \sum_{l \in I_j^{(c)}} f_i(\boldsymbol{z},\boldsymbol{Z}_l)$$

for $\boldsymbol{z} \in \mathbb{R}^b$, $b \in \mathbb{N} \setminus \{0\}$, and $\hat{\boldsymbol{Z}} = (Z_{l,i})$, a real valued matrix of dimension $n \times m$ with $j \in \{1,\dots,k\}$ and $i \in \{1,\dots,d\}$. We define the cumulative version of these functions by

$$\bar{h}_{i,j}^{(e)}(\boldsymbol{z},\hat{\boldsymbol{Z}}) = \sum_{r=j}^{k} h_{i,r}^{(e)}(\boldsymbol{z},\hat{\boldsymbol{Z}}) \quad ; \quad \bar{h}_{i,j}^{(c)}(\boldsymbol{z},\hat{\boldsymbol{Z}}) = \sum_{r=j}^{k} h_{i,r}^{(c)}(\boldsymbol{z},\hat{\boldsymbol{Z}})$$

with $j \in \{1,\dots,k\}$ and $\bar{h}_{i,k+1}^{(e)}(\boldsymbol{z},\hat{\boldsymbol{Z}}) = \bar{h}_{i,k+1}^{(c)}(\boldsymbol{z},\hat{\boldsymbol{Z}}) = 0$. We define vectors

$$\bar{\boldsymbol{h}}_j^{(e)}(\boldsymbol{z},\hat{\boldsymbol{Z}}) = \left(\bar{h}_{1,j}^{(e)}(\boldsymbol{z},\hat{\boldsymbol{Z}}),\dots,\bar{h}_{d,j}^{(e)}(\boldsymbol{z},\hat{\boldsymbol{Z}})\right)$$

and

$$\bar{\boldsymbol{h}}_j^{(c)}(\boldsymbol{z},\hat{\boldsymbol{Z}}) = \left(\bar{h}_{1,j}^{(c)}(\boldsymbol{z},\hat{\boldsymbol{Z}}),\dots,\bar{h}_{d,j}^{(c)}(\boldsymbol{z},\hat{\boldsymbol{Z}})\right).$$

With the above notation, the next theorem provides the posterior distribution of the generalized additive NTR regression model, (5.2), with a general VCRM and possibly censored data.

Theorem 13. *Let $\boldsymbol{\mu} = (\mu_1,\dots,\mu_d)$ be a d-variate VCRM with corresponding Lévy intensity of the form $\tilde{\nu}_d(\mathrm{d}\boldsymbol{s},\mathrm{d}x) = \tilde{\nu}_d(\boldsymbol{s},\mathrm{d}x)\mathrm{d}\boldsymbol{s}$ and $\mathscr{D}_{\hat{\boldsymbol{X}}}$ survival data with covariates following a generalized additive NTR regression with VCRM $\boldsymbol{\mu}$ and regressor functions $f_1,\dots f_m$. If $f_i > 0$ for at least one $i \in \{1,\dots,d\}$ and for $\eta_t(\boldsymbol{s}) = \tilde{\nu}_d(\boldsymbol{s},(0,t])$ the partial derivative $\eta'_{t_0}(\boldsymbol{s}) = \partial\eta_t(\boldsymbol{s})/\partial t\big|_{t=t_0}$ exists for arbitrary t_0 in $\mathbb{R}^+ \setminus \{0\}$, then the posterior distribution of*

$\boldsymbol{\mu}$ *given survival data* $\mathscr{D}_{\hat{\boldsymbol{X}}}$ *is the distribution of the random measure*

$$(\mu_1^\circ, \ldots, \mu_d^\circ) + \sum_{j \in I^{(e)}} (M_{1,j}\delta_{T_{(j)}}, \ldots, M_{d,j}\delta_{T_{(j)}})$$

where

i) $\boldsymbol{\mu}^\circ = (\mu_1^\circ, \ldots, \mu_d^\circ)$ *is a d-variate VCRM with Lévy intensity*

$$\nu_d^\circ(\mathrm{d}\boldsymbol{s}, \mathrm{d}x)\big|_{x \in [T_{(j-1)}, T_{(j)})} = e^{-\langle \bar{\boldsymbol{h}}_j^{(e)}(\boldsymbol{\beta}, \hat{\boldsymbol{X}}) + \bar{\boldsymbol{h}}_j^{(c)}(\boldsymbol{\beta}, \hat{\boldsymbol{X}}), \boldsymbol{s} \rangle} \tilde{\nu}_d(\mathrm{d}\boldsymbol{s}, \mathrm{d}x) \tag{5.6}$$

ii) The vectors of jumps $\{(M_{1,j}, \ldots, M_{d,j})\}_{j \in I^{(e)}}$ *are mutually independent and have, respectively, a d−variate probability density function given by*

$$g_j(\boldsymbol{s}) \propto \prod_{i=1}^{d} \left(e^{-\left(\bar{h}_{i,j+1}^{(e)}(\boldsymbol{\beta}, \hat{\boldsymbol{X}}) + \bar{h}_{i,j}^{(c)}(\boldsymbol{\beta}, \hat{\boldsymbol{X}})\right) s_i} \right) \prod_{l \in I_j^{(e)}} \left(1 - \prod_{i=1}^{d} e^{-f_i(\boldsymbol{\beta}, \boldsymbol{X}_l) s_i} \right) \eta'_{T_{(j)}}(\boldsymbol{s}) \tag{5.7}$$

iii) The random measure $\boldsymbol{\mu}^\circ$ *is independent of* $\{(M_{1,j}, \ldots, M_{d,j})\}_{j \in I^{(e)}}$.

We observe that we have the same posterior structure as for the multiple-sample model, a measure μ° with explicit Lévy intensity plus a series of δ measure with random weights and fixed location at the exact observation times; however the structure relating to the covariate structure is much richer now. The posterior characterization result is of special interest for its use in the inference scheme for model; we address this in the following corollaries.
The survival function $S(t) = \mathbb{P}[Y > t \,|\, \boldsymbol{\beta}, \boldsymbol{X}]$ conditional on the survival data $\mathscr{D}_{\hat{\boldsymbol{X}}}$ and fixed value $\boldsymbol{\beta}$ (which will often be a value sampled from the posterior distribution using a simulation algorithm) is presented in the next result.

Corollary 7. *In the setting of Theorem 13. For* $\emptyset \neq L \subset \{1, \ldots, d\}$ *denote*

$$I_t^{(e)} = \{l \,:\, T_{(l)} \text{ is an exact observation } \} \cap \{l \,:\, T_{(l)} \leq t\}.$$

Let $S^\star(t) = \mathbb{P}[Y^\star > t \,|\, \boldsymbol{\beta}, \boldsymbol{X}^\star]$ *be the survival function of a r.v.* $Y^\star$ *associated to a covariate vector* $\boldsymbol{X}^\star$ *and* $\boldsymbol{\beta}$ *a d−variate random vector; then*

$$\hat{S}^{\star}(t) = \mathbb{E}\left[\mathbb{E}[S^{\star}(t)|\boldsymbol{\mu}]\,|\mathscr{D}_{\hat{\boldsymbol{X}}},\boldsymbol{\beta},\boldsymbol{X}^{\star}\right] = e^{-\sum_{j=1}^{k+1}\left(\psi^{\circ}_{t\wedge T_{(j)}}(\boldsymbol{V}^{\star})-\psi^{\circ}_{t\wedge T_{(j-1)}}(\boldsymbol{V}^{\star})\right)\mathbb{1}_{\{T_{(j-1)}<t\}}}$$
$$\times \prod_{j\in I_t^{(e)}} \frac{\int_{(\mathbb{R}^+)^d}\prod_{i=1}^d \left(e^{-\left(V_i^{\star}+\bar{h}^{(e)}_{j+1,i}(\boldsymbol{\beta},\hat{\boldsymbol{X}})+\bar{h}^{(c)}_{j,i}(\boldsymbol{\beta},\hat{\boldsymbol{X}})\right)s_i}\right)\prod_{l\in I_j^{(e)}}\left(1-\prod_{i=1}^d e^{-f_i(\boldsymbol{\beta},\boldsymbol{X}_l)s_i}\right)\eta'_{T_{(j)}}(\boldsymbol{s})\mathrm{d}\boldsymbol{s}}{\int_{(\mathbb{R}^+)^d}\prod_{i=1}^d \left(e^{-\left(\bar{h}^{(e)}_{j+1,i}(\boldsymbol{\beta},\hat{\boldsymbol{X}})+\bar{h}^{(c)}_{j,i}(\boldsymbol{\beta},\hat{\boldsymbol{X}})\right)s_i}\right)\prod_{l\in I_j^{(e)}}\left(1-\prod_{i=1}^d e^{-f_i(\boldsymbol{\beta},\boldsymbol{X}_l)s_i}\right)\eta'_{T_{(j)}}(\boldsymbol{s})\mathrm{d}\boldsymbol{s}} \tag{5.8}$$

where $\boldsymbol{V}^{\star} = \left(V_1^{\star}\dots,V_d^{\star}\right) = (f_1(\boldsymbol{\beta},\boldsymbol{X}^{\star}),\dots,f_d(\boldsymbol{\beta},\boldsymbol{X}^{\star}))$ *and* ψ° *is the Laplace exponent of* μ°*.*

The underlying Lévy intensities with which we deal in our model usually have a parametric dependence in a vector of hyper-parameters $\boldsymbol{c}$. In the proof of Theorem 13 it is outlined how to derive the likelihood of the corresponding hyper-parameters when we consider censored to the right survival data with covariates, $\mathscr{D}_{\hat{\boldsymbol{X}}}$. This likelihood is essential for the inference scheme we present later on in this chapter.

Corollary 8. *In the setting of Theorem 12 with survival data* $\mathscr{D}_{\hat{\boldsymbol{X}}}$*, underlying Lévy intensity* $\tilde{\nu}_{d,\boldsymbol{c}}$*, partial derivative* $\eta'_{t,\boldsymbol{c}}$ *and* $\psi_{t,\boldsymbol{c}}$ *the associated Laplace exponent, for some real valued vector of hyper-parameters* $\boldsymbol{c}$*, we get the likelihood on* $\boldsymbol{c}$ *and* $\boldsymbol{\beta}$

$$l(\boldsymbol{\beta},\boldsymbol{c};\mathscr{D}_{\hat{\boldsymbol{X}}}) = e^{-\sum_{j=1}^{k}\left(\psi_{T_{(j)},\boldsymbol{c}}\left(\bar{\boldsymbol{h}}_j^{(c)}(\boldsymbol{\beta},\hat{\boldsymbol{X}})+\bar{\boldsymbol{h}}_j^{(e)}(\boldsymbol{\beta},\hat{\boldsymbol{X}})\right)-\psi_{T_{(j-1)},\boldsymbol{c}}\left(\bar{\boldsymbol{h}}_j^{(c)}(\boldsymbol{\beta},\hat{\boldsymbol{X}})+\bar{\boldsymbol{h}}_j^{(e)}(\boldsymbol{\beta},\hat{\boldsymbol{X}})\right)\right)}$$
$$\times \prod_{j\in I^{(e)}}\left\{\int_{(\mathbb{R}^+)^d}\prod_{i=1}^{d}\left(e^{-\left(\bar{h}^{(e)}_{j+1,i}(\boldsymbol{\beta},\hat{\boldsymbol{X}})+\bar{h}^{(c)}_{j,i}(\boldsymbol{\beta},\hat{\boldsymbol{X}})\right)s_i}\right)\prod_{l\in I_j^{(e)}}\left(1-\prod_{i=1}^{d}e^{-f_i(\boldsymbol{\beta},\boldsymbol{X}_l)s_i}\right)\eta'_{T_{(j)},\boldsymbol{c}}(\boldsymbol{s})\mathrm{d}\boldsymbol{s}\right\} \tag{5.9}$$

where for $\boldsymbol{\lambda}\in(\mathbb{R}^+)^d$*,* $t\in\mathbb{R}^+$

$$\psi_{t,\boldsymbol{c}}(\boldsymbol{\lambda}) = \int_{(\mathbb{R}^+)^d\times(0,t]}(1-e^{-\langle\boldsymbol{\lambda},\boldsymbol{s}\rangle})\tilde{\nu}_{d,\boldsymbol{c}}(\mathrm{d}\boldsymbol{s},\mathrm{d}x)$$

We can essentially use Proposition 11 from the last chapter to evaluate the Laplace exponent in the estimator that appears in Corollary 7. Next we give a Proposition to analytically calculate the integrals of the type involved in both the estimator of Corollary 7 and likelihood function of Corollary 8.

Proposition 13. *Let* $\boldsymbol{q}_1,\dots,\boldsymbol{q}_m \in (\mathbb{R}^+)^d$ *distinct of* $\mathbf{0} = (0,\dots,0)$, $\boldsymbol{q} \in (\mathbb{R}^+)^d$ *and* $\emptyset \neq I = \{i_1 \dots, i_{|I|}\} \subset \{1,\dots,m\}$. *We denote* $I \setminus 1 = \{i_2,\dots,i_{|I|}\}$. *Then*

$$\int_{(\mathbb{R}^+)^d \times (0t]} e^{-\langle \boldsymbol{q},\boldsymbol{s}\rangle} \prod_{l\in I}\left(1 - e^{-\langle \boldsymbol{q}_l,\boldsymbol{s}\rangle}\right) \tilde{\nu}_d(\mathrm{d}\boldsymbol{s},\mathrm{d}x)$$
$$= \sum_{S\subset I\setminus 1} (-1)^{\#(S)}\left(\psi_t(\boldsymbol{q}_1 + \sum_{l\in S}\boldsymbol{q}_l + \boldsymbol{q}) - \psi_t(\sum_{l\in S}\boldsymbol{q}_l + \boldsymbol{q})\right)$$

The preceding discussion highlight how the implementation of the inferential procedure, as for the multiple-sample information NTR model, depends on whether we can perform evaluations of the Laplace exponent or not.

5.3 Asymptotic results

In this section we use the main results of Kim and Lee (2001) and Kim and Lee (2004) to check posterior consistency and the Bernstein-von Mises theorem for the generalized additive NTR regression model. We use these results to provide guidelines about the choice of the VCRM and regressor functions $\{f_i\}_{i=1}^d$ in our model. For example in the CoRM case, this limits the choice of the score distribution and directing Lévy measure. These results also give insights into how these choices for the model affect the borrowing of information in a competing risks framework, (5.4), and how the model can be misspecified if the regressor functions are not flexible enough. The results presented in this section rely on the use of the cumulative hazard function for NTR distributions, see Proposition 4, by making use of the identity of (5.5) which relates our regression model to a NTR setting. In this section we consider possibly censored to the right survival data with covariates $\mathscr{D}_{\hat{\boldsymbol{X}}}^{(n)} = \{T_i, J_i, \boldsymbol{X}_i\}_{i=1}^n$, where special emphasis is placed on the number of observations n. When we say that a distribution has posterior consistency in this section it will be in the setting of Definition 9 with respect to a true underlying distribution with survival function S that generates survival data $\mathscr{D}_{\hat{\boldsymbol{X}}}^{(n)}$. In the context of (5.2), the following proposition establishes consistency if the entries of the VCRM are pairwise independent Gamma CRM's, see Examples 2 and 3 in Chapter 1.

Proposition 14. *Let* $d \in \mathbb{N}\setminus\{0\}$, $\boldsymbol{K} = (K_1,\dots,K_d)$ *be a random vector in* $(\mathbb{R}^+)^d$ *such that* $K_i \overset{a.s.}{\neq} 0$ *for* $i \in \{1,\dots,d\}$, *and let* $\boldsymbol{\mu} = (\mu_1,\dots,\mu_d)$ *be a vector of CRM's with independent entries such that marginally each entry is a Gamma CRM; then, given* $\boldsymbol{K}$, *the NTR*$(\sum_{i=1}^d K_i\mu_i)$ *distribution has a consistent posterior.*

On the other hand, if we consider complete identical dependence as discussed in Example 4 we get the next analogue result.

Proposition 15. *Let* $d \in \mathbb{N} \setminus \{0\}$, $\boldsymbol{K} = (K_1, \dots, K_d)$ *be a random vector in* $(\mathbb{R}^+)^d$ *such that* $\boldsymbol{K} \overset{a.s.}{\neq} \mathbf{0}$, μ *a CRM and* $\boldsymbol{\mu} = (\mu_1, \dots, \mu_d) = (\mu, \mu, \dots, \mu)$ *be a VCRM with Lévy intensity supported in* $\{(s_1, \dots, s_d) \,:\, s_1 = s_2 = \ldots = s_d\}$ *; then, given* $\boldsymbol{K}$, *the* $\sim NTR(\sum_{i=1}^d K_i \mu_i) = NTR\left((K_1 + \ldots K_d)\, \mu\right)$ *distribution has a consistent posterior if* μ *is a Gamma CRM.*

In terms of our model, (5.1), the previous two propositions give the next result.

Corollary 9. *Let* $\boldsymbol{\mu}$ *be a VCRM with Gamma CRM entries that are either mutually independent or a.s. equal. If* $f_i > 0$ *for every* $i \in \{1, \dots, d\}$ *then*

$$NTR\left(f_1(\boldsymbol{\beta}, \tilde{\boldsymbol{X}})\mu_1(0,t] + \ldots + f_d(\boldsymbol{\beta}, \tilde{\boldsymbol{X}})\mu_d(0,t] \middle| \mathscr{D}_{\hat{\boldsymbol{X}}}^{(n)}\right) \tag{5.10}$$

has a consistent posterior, with respect to the underlying true distribution with survival function S_0, *conditionally on* $\boldsymbol{\beta}$.

The next proposition can be seen as a generalization of Example 6 into a CoRM setting and also into the framework of our model.

Proposition 16. *Let* $\boldsymbol{K} = (K_1, \dots, K_d)$ *be a random vector with a probability density supported in* $(\mathbb{R}^+ \setminus \{0\})^d$ *and* $\boldsymbol{\mu} = (\mu_1, \dots, \mu_d)$ *a CoRM with Gamma directing Lévy measure and score distribution with density h supported in* $(\mathbb{R}^+)^d$. *Then the* $NTR(\sum_{i=1}^d K_i \mu_i)$ *distribution has a consistent posterior.*

The above result showcases a robustness for the posterior consistency of the NTR model with Gamma CRM when we consider the CoRM with directing Gamma Lévy measure as a perturbation. In terms of our model, (5.1), we get the next corollary for the proposition above

Corollary 10. *Let* $\boldsymbol{\mu}$ *be a CoRM with Gamma directing Lévy measure and score distribution with density h supported in* $(\mathbb{R}^+)^d$. *If* $f_i > 0$ *for every* $i \in \{1, \dots, d\}$ *then*

$$NTR\left(f_1(\boldsymbol{\beta}, \tilde{\boldsymbol{X}})\mu_1(0,t] + \ldots + f_d(\boldsymbol{\beta}, \tilde{\boldsymbol{X}})\mu_d(0,t] \middle| \mathscr{D}_{\hat{\boldsymbol{X}}}^{(n)}\right) \tag{5.11}$$

has a consistent posterior, with respect to the underlying true distribution with survival function S_0, *conditionally on* $\boldsymbol{\beta}$.

On a more general setting, if we consider the underlying VCRM in our generalized additive NTR regression model to be determined by an arbitrary absolutely continuous with respect to Lebesgue measure Lévy intensity $\tilde{\nu}_d$ then the conditions for posterior consistency in Proposition 5 are expressed as follows

Proposition 17. *Let $\boldsymbol{K}=(K_1,\ldots,K_d)$ be a random vector in $(\mathbb{R}^+)^d$ such that $K_i \overset{a.s.}{\neq} 0$ for $i\in\{1,\ldots,d\}$, and $\boldsymbol{\mu}=(\mu_1,\ldots,\mu_d)$ be a VCRM with absolutely continuous with respect to Lebesgue measure Lévy intensity $\tilde{\nu}_d(\boldsymbol{s},\mathrm{d}x)\mathrm{d}\boldsymbol{s}$. Then the distribution NTR$(\sum_{i=1}^d K_i\mu_i)$ has a consistent posterior if*

$$\xi(s,x)=\frac{\hat{\nu}(-\log(1-s),x)}{1-s}$$

satisfies the conditions of Proposition 5 where

$$\hat{\nu}(z)=\int_0^z\int_0^{z-y_d}\cdots\int_0^{z-y_d-\ldots-y_3}\frac{\tilde{\nu}_d\left(\frac{z-y_d-\ldots-y_2}{K_1},\frac{y_2}{K_2},\ldots,\frac{y_d}{K_d},x\right)}{K_1\cdots K_d}\mathrm{d}y_2\ldots\mathrm{d}y_d.$$

Theorem 14. *Let $b\in\mathbb{N}\setminus\{0\}$, $\boldsymbol{\beta}\in\mathbb{R}^b$, $Y^\star$ be an exact observation with covariate $\boldsymbol{X}^\star$, $S^\star(t)=\mathbb{P}[Y^\star>t\,|\boldsymbol{\beta},\boldsymbol{X}^\star]$ and $\boldsymbol{V}^\star=(f_1(\boldsymbol{\beta},\boldsymbol{X}^\star),\ldots,f_d(\boldsymbol{\beta},\boldsymbol{X}^\star))$. If*

$$NTR\left(f_1(\boldsymbol{\beta},\tilde{\boldsymbol{X}})\mu_1(0,t]+\ldots+f_d(\boldsymbol{\beta},\tilde{\boldsymbol{X}})\mu_d(0,t]\middle|\mathscr{D}_{\hat{\boldsymbol{X}}}^{(n)}\right) \tag{5.12}$$

has a consistent posterior, with respect to the underlying true distribution with survival function S_0, conditionally on $\boldsymbol{\beta}$; then

$$\lim_{n\to\infty}\mathbb{E}\left[S^\star(t)|\mathscr{D}_{\hat{\boldsymbol{X}}}^{(n)},\boldsymbol{\beta},\boldsymbol{X}^\star\right]=S_0(t)R(t;\hat{\boldsymbol{X}})$$

with

$$R(t;\hat{\boldsymbol{X}})=\lim_{n\to\infty}e^{-\sum_{j=1}^{k^{(n)}+1}\left(\psi^\circ_{t\wedge T_{(j)}}(\boldsymbol{V}^\star)-\psi^\circ_{t\wedge T_{(j)}}(\boldsymbol{1})+\psi^\circ_{t\wedge T_{(j-1)}}(\boldsymbol{1})-\psi^\circ_{t\wedge T_{(j-1)}}(\boldsymbol{V}^\star)\right)\mathbb{1}_{\{T_{(j-1)}<t\}}}\prod_{j\in I_t^{(e),(n)}}$$
$$\frac{\int_{(\mathbb{R}^+)^d}\prod_{i=1}^d\left(e^{-\left(V_i^\star+\bar{h}^{(e)}_{j+1,i}(\boldsymbol{\beta},\hat{\boldsymbol{X}}^{(n)})+\bar{h}^{(c)}_{j,i}(\boldsymbol{\beta},\hat{\boldsymbol{X}}^{(n)})\right)s_i}\right)\prod_{l\in I_j^{(e),(n)}}\left(1-\prod_{i=1}^d e^{-f_i(\boldsymbol{\beta},\boldsymbol{X}_l)s_i}\right)\nu(\boldsymbol{s})\mathrm{d}\boldsymbol{s}}{\int_{(\mathbb{R}^+)^d}\prod_{i=1}^d\left(e^{-\left(1+\bar{h}^{(e)}_{j+1,i}(\boldsymbol{\beta},\hat{\boldsymbol{X}}^{(n)})+\bar{h}^{(c)}_{j,i}(\boldsymbol{\beta},\hat{\boldsymbol{X}}^{(n)})\right)s_i}\right)\prod_{l\in I_j^{(e),(n)}}\left(1-\prod_{i=1}^d e^{-f_i(\boldsymbol{\beta},\boldsymbol{X}_l)s_i}\right)\nu(\boldsymbol{s})\mathrm{d}\boldsymbol{s}} \tag{5.13}$$

where dependence on n has been made explicit by a super-index (n).

We observe that if $f_i\equiv 1$ then $R\equiv 1$ so our regression model estimator is equivalent in the limit to the frequentist Kaplan-Meier estimator. The other asymptotic property we focus on in this chapter are the Bernstein-von Mises results, as in Definition 11, for the generalized additive NTR regression model. In such context we consider again a true underlying survival function S_0 which generates the survival data with covariates $\mathscr{D}_{\hat{\boldsymbol{X}}}^{(n)}$.

Proposition 18. *Let $\boldsymbol{K} = (K_1, \ldots, K_d)$ be a random vector with a probability density supported in $(\mathbb{R}^+ \setminus \{0\})^d$ such that $\mathbb{E}[K_i], \mathbb{E}[1/K_i] < \infty$ for any $i \in \{1, \ldots, d\}$; and $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_d)$ be a CoRM with Gamma directing Lévy measure and a score distribution with density h supported in $(\mathbb{R}^+)^d$ such that if $(U_1, \ldots, U_d) \sim \mathscr{L}(h)$ we have that $\mathbb{E}[U_i] < \infty$ for any $i \in \{1, \ldots, d\}$ and $\mathbb{E}\left[1/U_j\right] < \infty$ for at least one $j \in \{1, \ldots, d\}$. Then the distribution $NTR(\sum_{i=1}^d K_i \mu_i)$ attains the Bernstein-von Mises theorem.*

In the context of the generalized additive NTR regression model the next result follows.

Corollary 11. *Let $\boldsymbol{\mu}$ be a $d-$variate CoRM with Gamma directing Lévy measure and and a score distribution with density h supported in $(\mathbb{R}^+)^d$ such that if $(U_1, \ldots, U_d) \sim \mathscr{L}(h)$ we have that $\mathbb{E}[U_i] < \infty$ for any $i \in \{1, \ldots, d\}$ and $\mathbb{E}\left[1/U_j\right] < \infty$ for at least one $j \in \{1, \ldots, d\}$. If $\mathbb{E}\left[1/f_i(\boldsymbol{\beta}, \tilde{\boldsymbol{X}})\right], \mathbb{E}\left[f_i(\boldsymbol{\beta}, \tilde{\boldsymbol{X}})\right] < \infty$ for any $i \in \{1, \ldots, d\}$ then given survival data $\mathscr{D}_{\hat{\boldsymbol{X}}}^{(n)}$*

$$S(t; \tilde{\boldsymbol{X}}) = e^{-f_1(\boldsymbol{\beta}, \tilde{\boldsymbol{X}})\mu_1(0,t] - \ldots - f_d(\boldsymbol{\beta}, \tilde{\boldsymbol{X}})\mu_d(0,t]} \tag{5.14}$$

satisfies the Bernstein-von Mises theorem with respect to the Kaplan-Meier estimator of the possibly censored to the right observations in $\mathscr{D}_{\hat{\boldsymbol{X}}}^{(n)}$ for any $\tilde{\boldsymbol{X}} \in \mathbb{R}^d$.

Results as Theorem 14 are interesting as they show a case, namely when $f_i > 0$ for every $i \in \{1, , \ldots, d\}$, where the posterior mean estimator (5.8) collapses in the limit to the marginal survival associated to the survival data without covariates, $\mathscr{D}$, times a function R which accounts for the covariate dependence. If the regressor functions f_i are not chosen with enough flexibility we have that our regression model can become misspecified.

Example 17. Model misspecification
We draw samples from two different populations $\boldsymbol{Y}^{(1)} = \{Y_i^{(1)}\}_{i=1}^{\infty}$ and $\boldsymbol{Y}^{(2)} = \{Y_i^{(2)}\}_{i=1}^{\infty}$ given by

$$Y_i^{(1)} \overset{\text{i.i.d.}}{\sim} \text{Weibull}\left(\text{shape} = 2.1, \text{rate} = 0.5\right) \; ; \; Y_i^{(2)} \overset{\text{i.i.d.}}{\sim} \text{Weibull}\left(\text{shape} = 0.9, \text{rate} = 0.5\right).$$

Given $n \in \mathbb{N} \setminus \{0\}$ we consider $\{Y_i, Z_i\}_{i=1}^n$ such that no censoring is considered,

$$Y_i \overset{\text{i.i.d.}}{\sim} 0.5\text{Weibull}\left(\text{shape} = 2.1, \text{rate} = 0.5\right) + 0.5\text{Weibull}\left(\text{shape} = 0.9, \text{rate} = 0.5\right),$$

where we have used the usual notation for mixture distributions, and

$$Z_i = \begin{cases} 1, & \text{if } Y_i \sim \text{Weibull}\left(\text{shape} = 2.1, \text{rate} = 0.5\right). \\ 2, & \text{if } Y_i \sim \text{Weibull}\left(\text{shape} = 0.9, \text{rate} = 0.5\right). \end{cases}.$$

For illustration purposes we take a VCRM consisting of two independent homogeneous Gamma(α_i, β_i) CRM's with $\kappa(\mathrm{d}x) = \mathrm{d}x$, $i \in \{1,2\}$. If we consider the generalized additive NTR regression model with regressor functions f_1, f_2 given by

$$f_1(Z) = \delta_1(Z)\ ;\ f_2(Z) = \delta_2(Z),$$

then the model is equivalent to having independent NTR (Gamma(α_i, β_i)) distributions for each population, $\boldsymbol{Y}^{(i)}$, appearing in the survival data, $i \in \{1,2\}$. As we have Gamma CRM's there will be consistency for each population, see Figure 5.2.

Two populations fit with multiple-sample NTR model

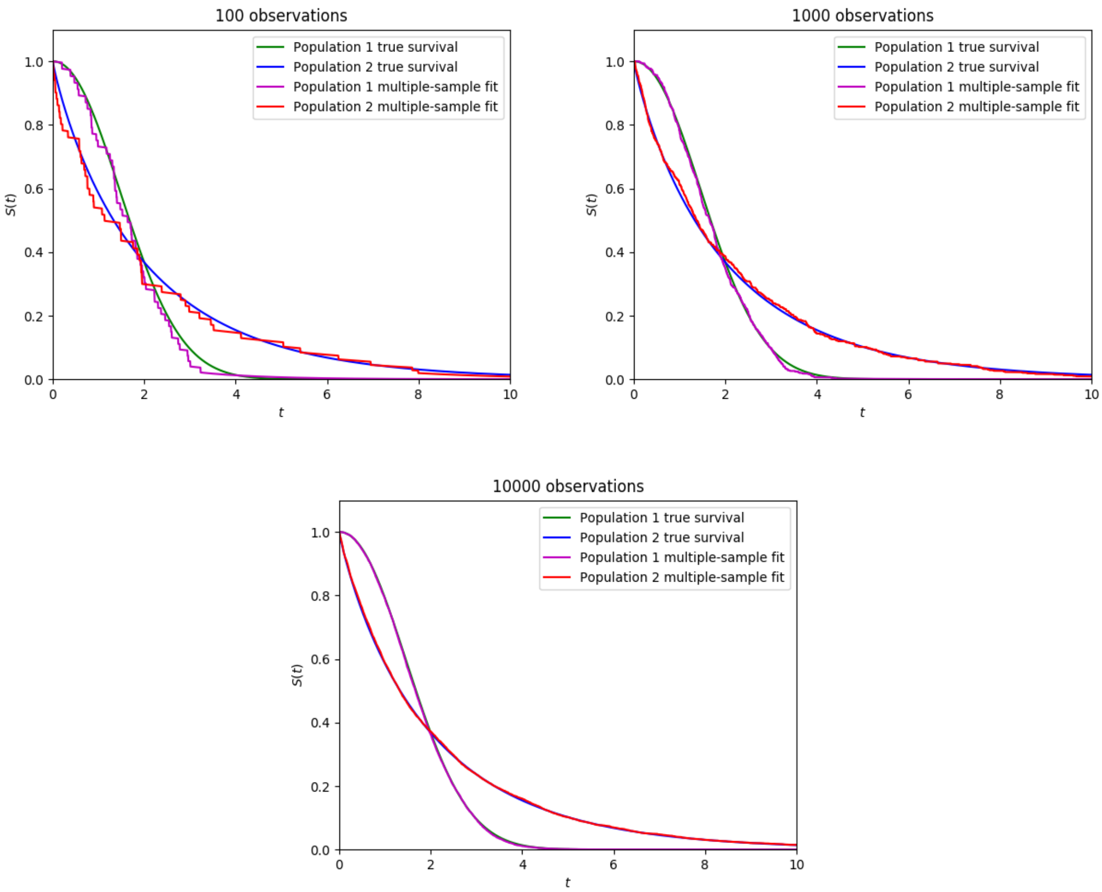


Fig. 5.2 Plot of NTR fits given by the estimator (5.8) for two Weibull distributed populations with independent Gamma$(1,1)$ CRM's. Draws from 100, 1000 and 10000 observations without censoring were considered.

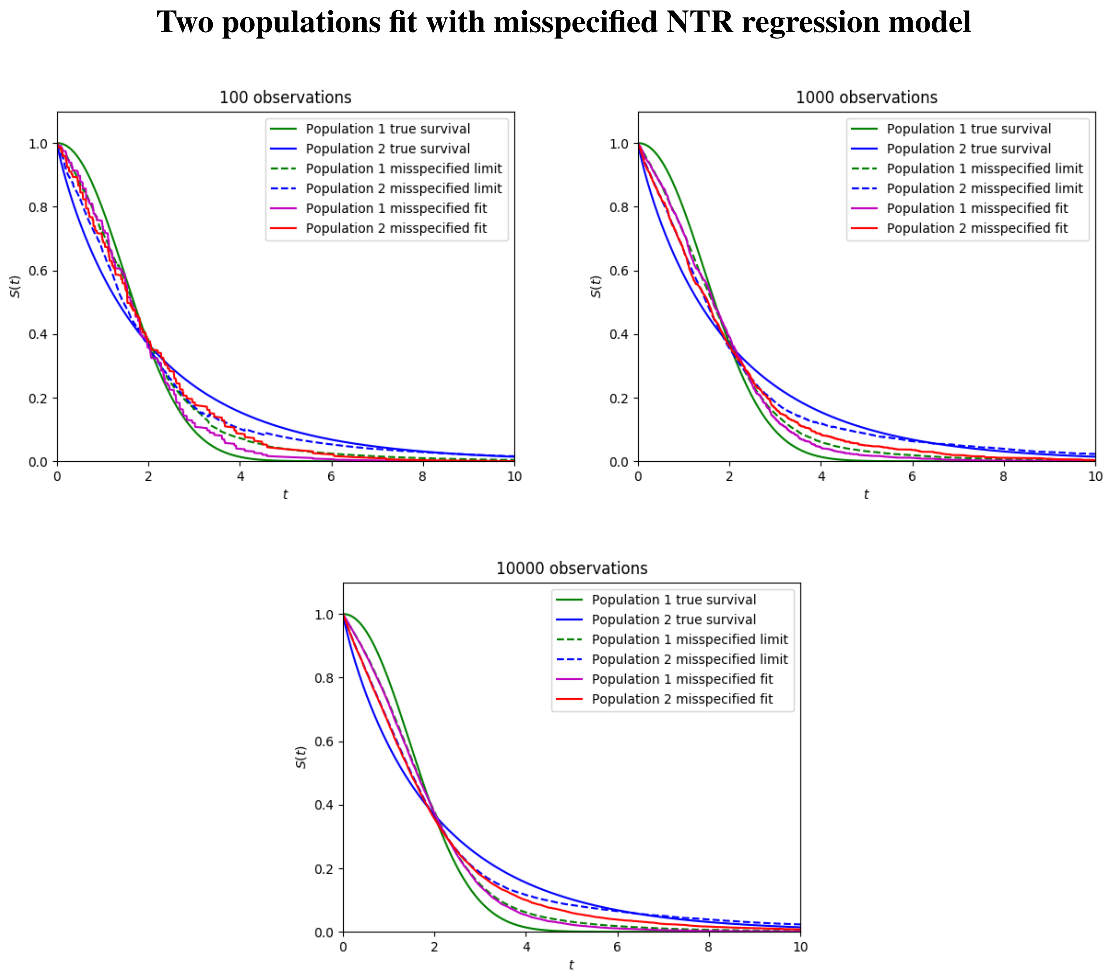


Fig. 5.3 Plot of NTR fits given by the estimator (5.8) for two Weibull distributed populations with independent Gamma$(1,1)$ CRM's. Draws from 100, 1000 and 10000 observations without censoring were considered. The misspecified limit $S_0(t)R(t;\hat{\boldsymbol{X}})$ is calculated with the true mixture of Weibull distribution for S_0 and an approximation of R by considering $\mathscr{D}^{(n)}_{\hat{\boldsymbol{X}}}$ in the limit in (5.13).

On the other hand, if we take the strictly positive regressor functions f_1, f_2 given by

$$f_1(Z) = 1.5\delta_1(Z) + 0.5\delta_0(Z)$$
$$f_2(Z) = 0.5\delta_1(Z) + 1.5\delta_0(Z),$$

then, still using independent Gamma CRM's for the VCRM, because of Theorem 14 the posterior distribution of individuals belonging to either population, with covariate $Z = 1$ or $Z = 2$, will collapse in the distribution given by $S_0(t)R(t;Z)$. Here S_0 is the survival function associated to the full set of observations $\{Y_i\}_{i=1}^n$, which in this case corresponds to the uniform mixture, weights 0.5 and 0.5 for each involved distribution, of the Weibull laws associated to each population $\boldsymbol{Y}^{(1)}, \boldsymbol{Y}^{(2)}$; and $R(\cdot;Z)$ is the function depending on the covariate Z that was given in Theorem 14 and which, by the choices of f_1, f_2 is different from 1 so the model will produce misspecification as showed in Figure 5.3

We see from the previous example that special care has to be taken in the choice of the regressor functions, so they are flexible enough to avoid misspecification as in the previous example.

5.4 Real data analyses

In this section we analyse real survival datasets with the generalized additive NTR regression model. For the underlying VCRM we will use a CoRM that can modulate between assigning mass on random weights in the axis of $(\mathbb{R}^+)^d$ and the identity line.

Definition 27. Let $\delta \in [0,1]$, we denote by $\delta-\text{LogNormal}$ the distribution given by the mixture

$$\frac{1}{d}\sum_{i=1}^{d} \text{LogNormal}\left((1-\delta)\boldsymbol{e}_i + \delta\mathbf{1}\,,\,\sigma I^{(d)}\right).$$

If $\delta = 0$ then the δ-LogNormal distribution accumulates mass near the axis in $(\mathbb{R}^+)^d$ and if $\delta = 1$ then the mass of the distribution is accumulated near the identity. As showcased in Example 11 such distribution can be used as the score distribution of a CoRM. In all the real data studies of this section we will use a $\delta-\text{LogNormal}-\text{Gamma}(\alpha,\beta)$ CoRM with $\sigma = 0.1$ in the $\delta-\text{LogNormal}$ score distribution and $\kappa(\mathrm{d}x) = \mathrm{d}x$ for location part in the homogeneous Gamma directing CRM. We observe that computation of the Laplace exponent of a $\delta-\text{LogNormal}-\text{Gamma}$ CoRM is not straightforward. However, as the Laplace exponent of the Gamma directing CRM is explicitly available and drawing samples from the score distribution is possible, we can use Theorem 6 to give a Monte-Carlo estimator for the Laplace exponent of the $\delta-\text{LogNormal}-\text{Gamma}$ CoRM. We propose a MCMC scheme for estimation of the mean posterior survival function in Corollary 7, equation (5.8), when assigning prior distributions on the vectors $\boldsymbol{\beta}, \boldsymbol{c}$, as in Corollary 8. For all the generalized additive NTR regression fits in this section we first perform a NTR maximum a posteriori estimate $\hat{\boldsymbol{c}}^{\text{maxpost}}$, see the example in Figure 1.5 and equation 1.24, for the vector of hyper-parameters $\boldsymbol{c} = (\alpha,\beta)$ for the directing Gamma Lévy measure associated to the underlying CoRM. So we fix $\boldsymbol{c} = \hat{\boldsymbol{c}}^{\text{maxpost}} = (\alpha^{\text{maxpost}}, \beta^{\text{maxpost}})$ in the rest of the inferential scheme. For the estimation of the parameters $\boldsymbol{\beta} \in \mathbb{R}^b$ and $\lambda \in [0,1]$ we use a pseudo-marginal Metropolis within Gibbs algorithm to draw

$$\boldsymbol{\beta}^{(i)}|\hat{\boldsymbol{c}}^{\text{maxpost}}, \mathscr{D}_{\hat{\boldsymbol{X}}}, \quad 1 \leq i \leq M;$$

and use the likelihood (5.9) to identify $\left(\hat{\boldsymbol{\beta}}^{\text{maxpost}}, \hat{\lambda}^{\text{maxpost}}\right)$ attaining the running maximum a posterior value along the MCMC chain above. We resort to the maximum a posteriori approaches instead of using averaged estimator along the MCMC chain as in (1.25), see the discussion pertaining Figure 1.6, because the Laplace exponent of the $\delta -$ LogNormal $-$ Gamma CoRM is not explicitly available so we recur to a Monte-Carlo approximation, which is computationally more expensive, instead. Even more, we have to recur to the pseudo-marginal approach as in each MCMC step the Monte-Carlo estimator of the $\delta -$ LogNormal $-$ Gamma CoRM's Laplace exponent is used.

Algorithm 6 Pseudo-marginal Metropolis within Gibbs for generalized additive NTR regression model

1: Draw $\lambda^{(i+1)}$ from a Metropolis-Hastings sampler with Uniorm$(0,1)$ proposal distribution and target distribution

$$l(\boldsymbol{\beta}^{(i)}, x; \mathscr{D}_{\hat{\boldsymbol{X}}})$$

with $\hat{\psi}^{\text{MonteCarlo}}$ instead of ψ.

2: Draw $\boldsymbol{\beta}^{(i+1)}$ from a Metropolis-Hastings sampler with suitable proposal distribution and target distribution

$$l(\boldsymbol{x}, \lambda^{(i+1)}; \mathscr{D}_{\hat{\boldsymbol{X}}}) p_{\boldsymbol{\beta}}(\boldsymbol{x}).$$

with $\hat{\psi}^{\text{MonteCarlo}}$ instead of ψ and $p_{\boldsymbol{\beta}}$ a suitable prior distribution in $\boldsymbol{\beta}$, the vector of parameters for the regressor functions $\{f_i\}_{i=1}^{\infty}$.

Anemia clinical trials

In Kalbfleisch and Prentice (2011), Table 1.2, two different treatments for patients with anemia were considered; one with cyclosporine and methotrexate, and the other with methotrexate alone. Each treatment study involved 64 patients. The Kaplan-Meier estimators of the two treatments cross each other; we see that we can recover this behaviour with the two-dimensional δ-LogNormal-Gamma CoRM and the multiple-sample model regressor functions

$$f_1(Z_i) = \mathbb{1}_{\{\text{Patient } Y_i \text{ belongs to treatment } 1.\}}$$
$$f_2(Z_i) = \mathbb{1}_{\{\text{Patient } Y_i \text{ belongs to treatment } 2.\}}.$$

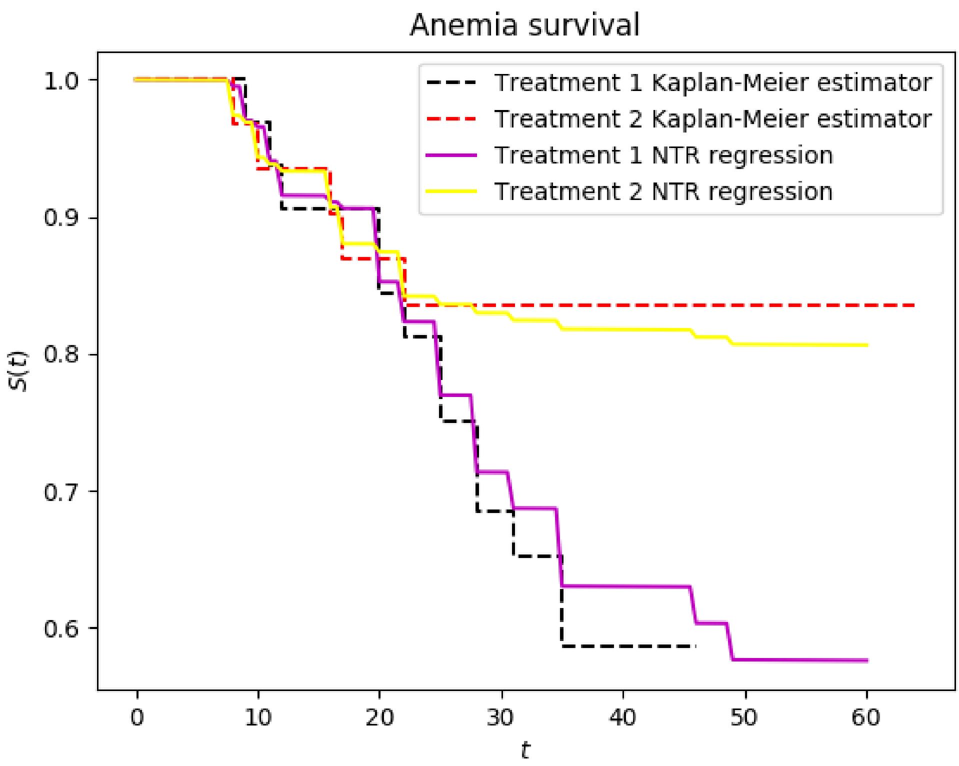


Fig. 5.4 δ-LogNormal-Gamma model fit for two anemia treatments with $\hat{\delta} = 0.13$. Kaplan-Meier estimators are given for comparison.

We observe that in Figure 5.4 the survival estimators in our model for the two treatments cross each other. As $\hat{\delta}^{\text{maxpost}} = 0.13$ is close to zero, there is a subtle sharing of information that can be observed when comparing with the Kaplan-Meier estimators.

Melanoma survival data

In Andersen et al. (2012), 205 patients with melanoma which had a tumour removed by surgery were considered. The thickness of the tumour was one of the covariables of interest as an increase in the tumour's thickness is thought to increase the chances of death. Again we use the two dimensional $\delta-\text{LogNormal}-\text{Gamma}$ CoRM and choose the regressor functions in a flexible way. Let $\boldsymbol{q} = (q_1, q_2, q_3 q_4, q_5)$ be the quantiles of the thickness covariates at 0.0, 0.25, 0.5, 0.75, 1.0, respectively, then given $\boldsymbol{\beta} \in [0,1]^5$ we set $\boldsymbol{k}(\boldsymbol{q}, \boldsymbol{\beta}) =$

$\{(q_1, \beta_1), \ldots (q_5, \beta_5)\}$ and regressor functions

$$f_1(z, \boldsymbol{\beta}) = \max\{0, \text{Spline}_{\boldsymbol{k}}^{(3)}(z)\}$$
$$f_2(z, \boldsymbol{\beta}) = \max\{0, \max(\boldsymbol{\beta}) - f_1(z)\}$$

where $\text{Spline}_{\boldsymbol{k}}^{(3)}$ is a spline of degree 3 with knots $\boldsymbol{k}$ and $\max(\boldsymbol{\beta}) = \max\{\beta_1, \ldots, \beta_5\}$. With such regressor functions we can approximately recover a multiple-sample model where the samples are given by the thickness covariate belonging to the supports of f_1, f_2. The maximum a posteriori regressor functions are given in Figure 5.5, where we observe that patients with tumour thickness between the 0.25 and 0.5 quantiles have disjoint supports of the regressor functions with respect to patients with tumour thickness above the 0.5 quantile; which we interpret as the two populations the model has fitted.

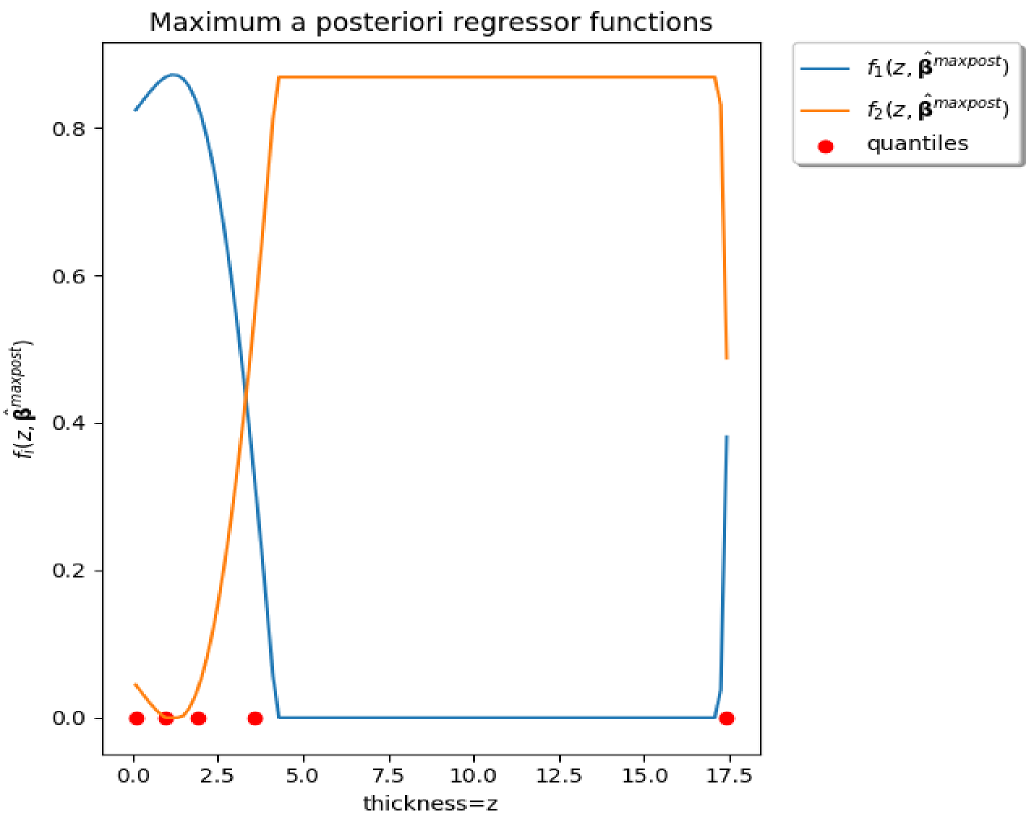


Fig. 5.5 Regressor functions f_1, f_2 for the melanoma real data, evaluated at $\hat{\boldsymbol{\beta}}^{\text{maxpost}}$.

The maximum a posteriori for δ heuristically quantifies how much sharing of information is the model gives to the pseudo-populations given by the disjoint supports of the regressor functions. In our case

$$\hat{\delta}^{\text{maxpost}} = 0.000371,$$

so there is very little sharing of information between the pseudo-populations.

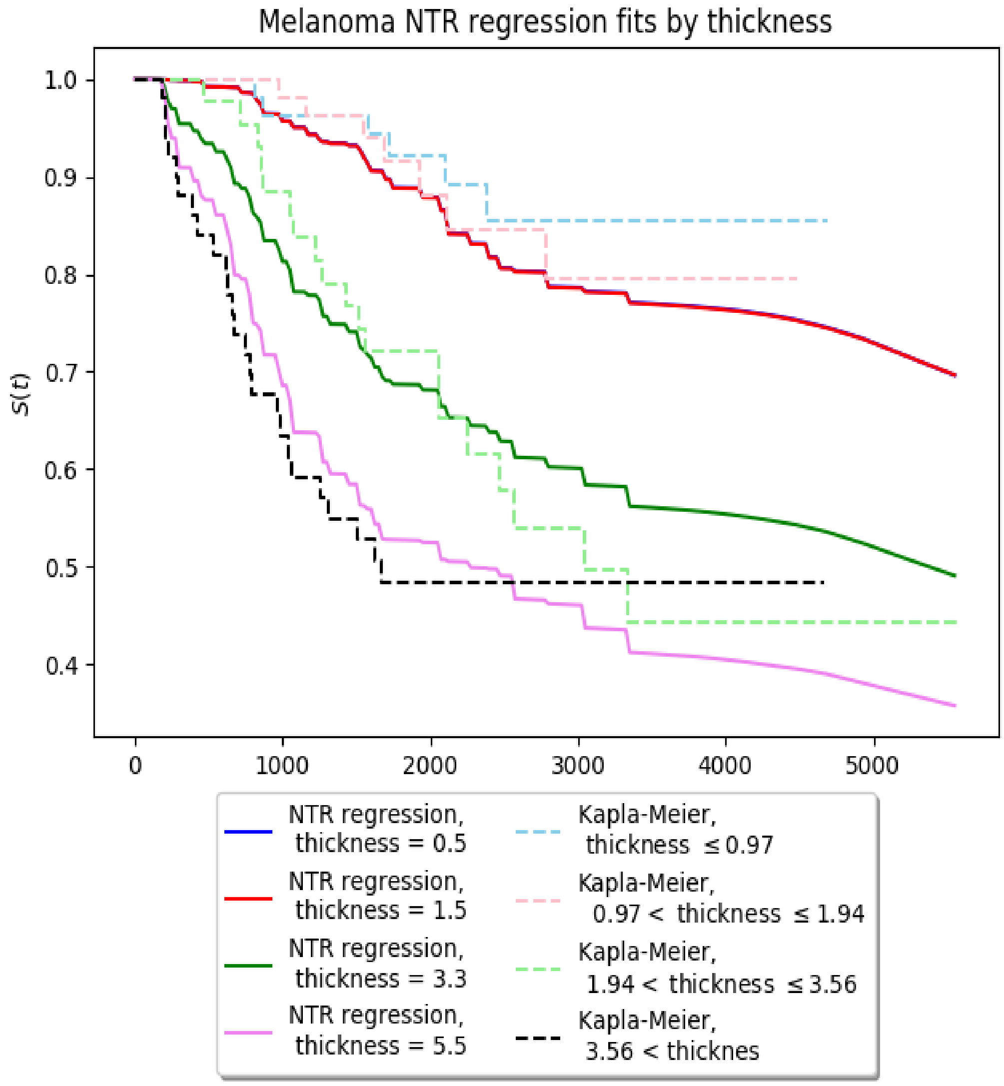


Fig. 5.6 δ – LogNormal – Gamma fits for thickness values 0.5, 1.5, 3.3 and 5.5. For comparison we present Kaplan-Meier fits of observations with thickness values between the quantiles at 0.0, 0.25, 0.5 and 0.75 and 1.0.

Kidney transplant data

We consider the Kidney transplant dataset from the survival analysis book of Klein and Moeschberger (2006) which is available in the R package "KMsurv" by Yan (2010). This dataset consists of 863 observations with an indicator covariate to specify if the observation was exact or censored, an indicator covariate to determine if the observation corresponded to a male or female patient, an indicator covariate to determine if the observation corresponded to a white or black patient, and an age covariate which we treat as a continuous variable.
As patients can be divided in populations 1) Male-White, 2) Male-Black, 3) Female-White and 4) Female-Black, we consider the 4-variate δ-LogNormal-Gamma CoRM. We consider regressor functions f_i, $i \in \{\mathrm{mw}, \mathrm{mb}, \mathrm{fw}, \mathrm{fb}\}$ for the label of each populations; which we define as

$$\begin{aligned}
f_{\mathrm{mw}}(z) &= e^{\beta_{0,\mathrm{mw}}+\beta_{1,\mathrm{mw}} z_{\mathrm{age}}} \mathbb{1}_{\{z_{\mathrm{gender}}=\mathrm{male},\, z_{\mathrm{race}}=white\}} \\
f_{\mathrm{mb}}(z) &= e^{\beta_{0,\mathrm{mb}}+\beta_{1,\mathrm{mb}} z_{\mathrm{age}}} \mathbb{1}_{\{z_{\mathrm{gender}}=\mathrm{male},\, z_{\mathrm{race}}=black\}} \\
f_{\mathrm{fw}}(z) &= e^{\beta_{0,\mathrm{fw}}+\beta_{1,\mathrm{fw}} z_{\mathrm{age}}} \mathbb{1}_{\{z_{\mathrm{gender}}=\mathrm{female},\, z_{\mathrm{race}}=white\}} \\
f_{\mathrm{fb}}(z) &= e^{\beta_{0,\mathrm{fb}}+\beta_{1,\mathrm{fb}} z_{\mathrm{age}}} \mathbb{1}_{\{z_{\mathrm{gender}}=\mathrm{female},\, z_{\mathrm{race}}=black\}}
\end{aligned}$$

The intercept coefficients $\beta_{0,\mathrm{mw}}, \beta_{0,\mathrm{mb}}, \beta_{0,\mathrm{fw}}, \beta_{0,\mathrm{fb}}$ account for the heterogeneity in the populations. The linear coefficients for the age $\beta_{1,\mathrm{mw}}, \beta_{1,\mathrm{mb}}, \beta_{1,\mathrm{fw}}, \beta_{1,\mathrm{fb}}$ account for decreasing survivals when the age augments. In the Male-White population there are 431 individuals, this is the biggest population in the sample. In Figure 5.7 the generalized additive NTR regression fit is presented for patients in this population with different age values. The Female-White populations has 278 individuals, making it the second largest; our model fits for this population are presented in Figure 5.8 with different age covariates. In contrast with the White populations, the Black-Male and Black-Female population contain fewer individual, 92 for Black-Male and 59 for Black-Female; for this reason we only present the fit for the age equal to 50 covariate, being close to the mean and mode of the age covariate for both populations. In Figures 5.9 and 5.10 we present the corresponding fits.

Kidney-transfer white male population NTR regression fits by age

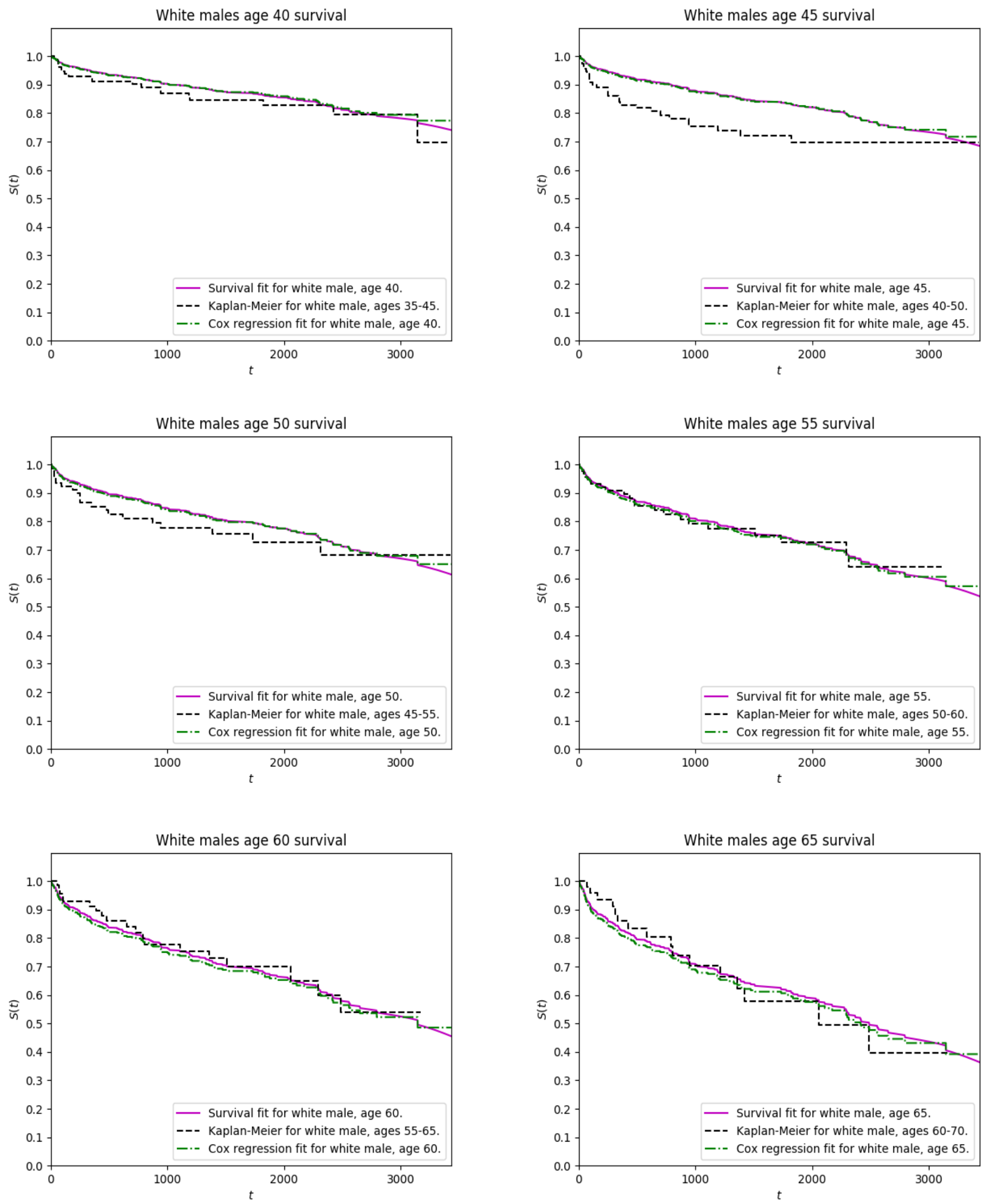


Fig. 5.7 δ – LogNormal – Gamma fits for the White-Male population in the kidney transfer data set for different age values. Kaplan-Meier estimators and the associated Cox regression estimator are presented for comparisson.

Kidney-transfer white female population NTR regression fits by age

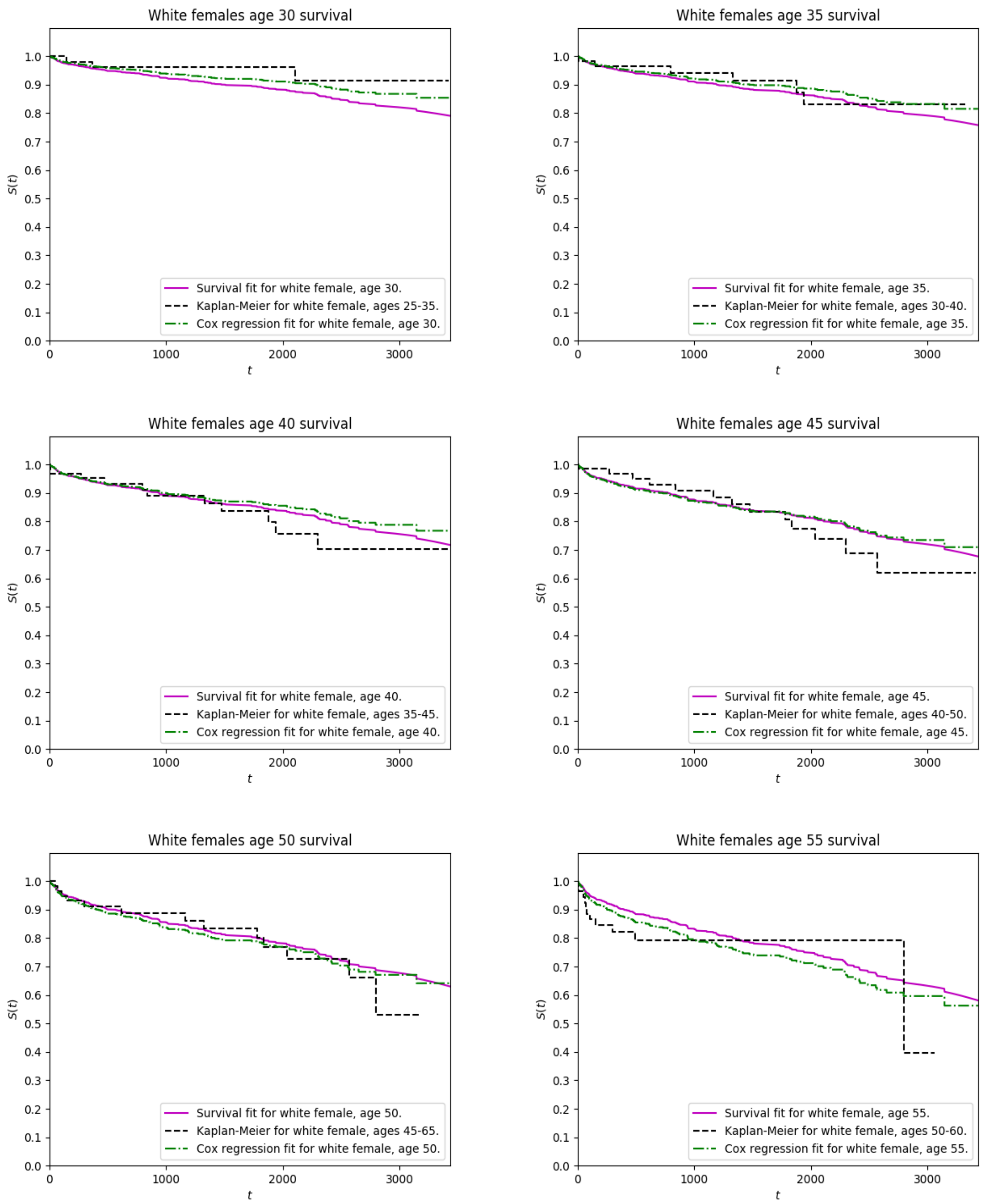


Fig. 5.8 $\delta-\text{LogNormal}-\text{Gamma}$ fits for the White-Female population in the kidney transfer data set for different age values. Kaplan-Meier estimators and the associated Cox regression estimator are presented for comparisson.

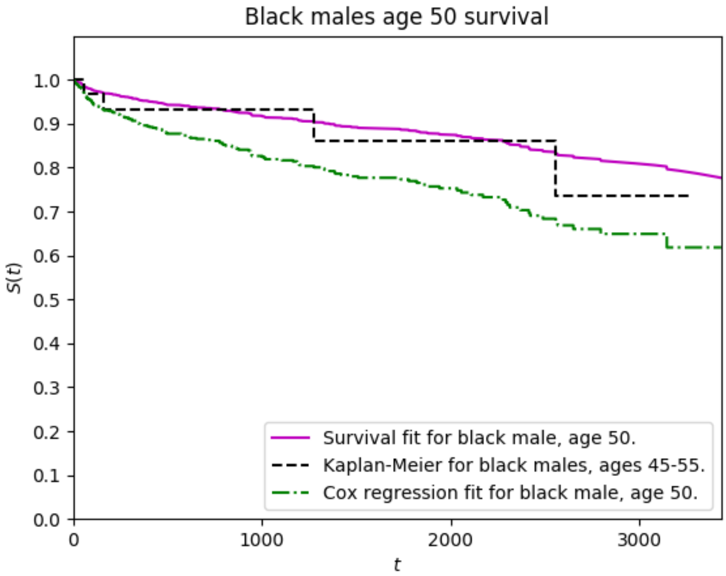


Fig. 5.9 δ – LogNormal – Gamma fits for the Black-Male population in the kidney transfer data set for different age values. Kaplan-Meier estimators and the associated Cox regression estimator are presented for comparisson.

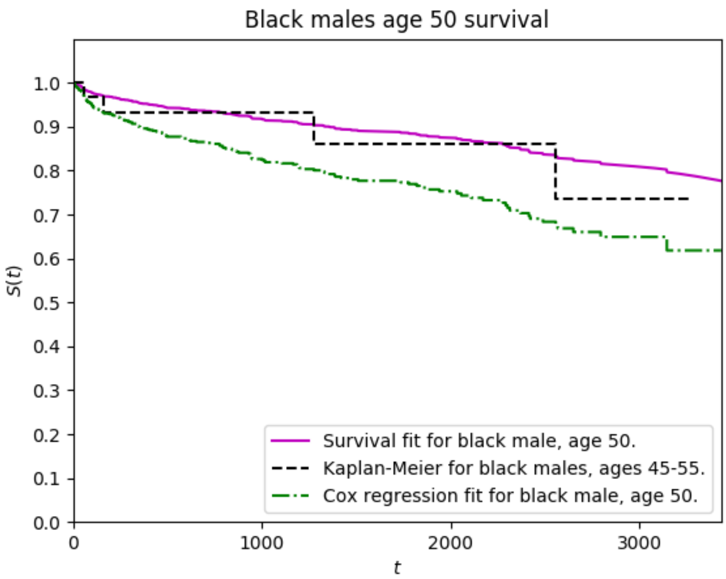


Fig. 5.10 δ – LogNormal – Gamma fits for the Black-Female population in the kidney transfer data set for different age values. Kaplan-Meier estimators and the associated Cox regression estimator are presented for comparisson.

5.5 Proof of generalized additive NTR regression model results

Proof of Theorem 13

First we will use the next lemma.

Lemma 7. *Let* $(\mu_1,\dots,\mu_d)$ *be a d-variate CRM such that* $\mu_1,\dots,\mu_d$ *are not independent and let the Levy intensity* $\tilde{\nu}_d(\boldsymbol{s},\mathrm{d}t)\mathrm{d}\boldsymbol{s}$ *of* $(\mu_1,\dots,\mu_d)$ *be such that* $\eta_t=\tilde{\nu}_d(\boldsymbol{s},(0,t])$ *is differentiable with respect to* $t\in\mathbb{R}^+$ *at some* $t_0\neq 0$ *and denote* $\eta'_{t_0}(\boldsymbol{s})=\partial\eta_t(\boldsymbol{s})\partial t\big|_{t=t_0}$*. Let it be* $\emptyset\neq I\subset\mathbb{N}\setminus\{0\}$*, if* $\forall l\in I$ $\boldsymbol{q}_l=(q_{l,1},\dots,q_{l,d})\in\mathbb{N}^d$ *are such that* $\max\{q_{l,1},\dots,q_{l,d}\}>0$*, and* $\boldsymbol{r}=(r_1,\dots,r_d)\in(\mathbb{R}^+)^d$ *are such that* $\min\{r_1,\dots,r_d\}>0$*, then*

$$\mathbb{E}\left[\prod_{i=1}^d\left(e^{-r_i\mu_i(A_\varepsilon)}\right)\prod_{l\in I}\left(1-\prod_{i=1}^d e^{-q_{l,i}\mu_i(A_\varepsilon)}\right)\right]$$
$$=\varepsilon\int_{(\mathbb{R}^+)^d}e^{-\langle\boldsymbol{r},\boldsymbol{s}\rangle}\prod_{l\in I}\left(1-e^{-\langle\boldsymbol{q}_l,\boldsymbol{s}\rangle}\right)\eta'_{t_0}(\boldsymbol{s})\mathrm{d}\boldsymbol{s}+o(\varepsilon)$$

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

Proof. Let it be $\emptyset\neq I\subset\mathbb{N}\setminus\{0\}$, then

$$\mathbb{E}\left[\prod_{i=1}^d\left(e^{-r_i\mu_i(t_0-\varepsilon,t_0]}\right)\prod_{l\in I}\left(1-\prod_{i=1}^d e^{-q_{l,i}\mu_i(t_0-\varepsilon,t_0]}\right)\right]$$
$$=\mathbb{E}\left[\prod_{i=1}^d\left(e^{-r_i\mu_i(t_0-\varepsilon,t_0]}\right)\sum_{S\subset I}(-1)^{\#(S)}\prod_{l\in S}\prod_{i=1}^d e^{-q_{l,i}\mu_i(t_0-\varepsilon,t_0]}\right]$$
$$=\sum_{S\subset I}(-1)^{\#(S)}\mathbb{E}\left[\prod_{i=1}^d e^{-(\sum_{l\in S}q_{l,i}+r_i)\mu_i(t_0-\varepsilon,t_0]}\right]$$
$$=\sum_{S\subset I}(-1)^{\#(S)}e^{-[\psi_{t_0}(r_1+\sum_{l\in S}q_{1,l},\dots,r_d+\sum_{l\in S}q_{d,l})-\psi_{t_0-\varepsilon}(r_1+\sum_{l\in S}q_{1,l},\dots,r_d+\sum_{l\in S}q_{d,l})]}$$
$$=e^{-\psi_{t_0}(r_1,\dots,r_d)+\psi_{t_0-\varepsilon}(r_1,\dots,r_d)}\sum_{S\subset I}(-1)^{\#(S)}e^{-\triangle_{t_0-\varepsilon}^{t_0}[\psi_t(r_1+\sum_{l\in S}q_{1,l},\dots,r_d+\sum_{l\in S}q_{d,l})-\psi_t(r_1,\dots,r_d)]}$$

With $\boldsymbol{j}=(j_1,\dots,j_d)\in(\mathbb{R}^+)^d$ such that $\min\{j_1,\dots,j_d\}$; we use the Taylor expansion

$$e^{-\triangle_{t_0-\varepsilon}^{t_0}[\psi_t(\boldsymbol{r}+\boldsymbol{j})-\psi_t(\boldsymbol{r})]}=e^{-\int_{(\mathbb{R}^+)^d}e^{-\langle\boldsymbol{r},\boldsymbol{x}\rangle}(1-e^{-\langle\boldsymbol{j},\boldsymbol{s}\rangle})\triangle_{t_0-\varepsilon}^{t_0}\eta_t(\boldsymbol{s})\mathrm{d}\boldsymbol{s}}$$
$$=1-\varepsilon\int_{(\mathbb{R}^+)^d}e^{-\langle\boldsymbol{r},\boldsymbol{s}\rangle}(1-e^{-\langle\boldsymbol{j},\boldsymbol{s}\rangle})\eta'_{t_0}(\boldsymbol{s})\mathrm{d}\boldsymbol{s}+o(\varepsilon) \tag{5.15}$$

So

$$
\begin{aligned}
&\mathbb{E}\left[\prod_{i=1}^{d}\left(e^{-r_i\mu_i(t_0-\varepsilon,t_0]}\right)\prod_{l\in I}\left(1-\prod_{i=1}^{d}e^{-q_{l,i}\mu_i(t_0-\varepsilon,t_0]}\right)\right]\\
&=e^{-\psi_{t_0}(r_1,\dots,r_d)+\psi_{t_0-\varepsilon}(r_1,\dots,r_d)}\left(1+\sum_{\emptyset\neq S\subset I}(-1)^{\#(S)}\{1\right.\\
&\quad\left.-\varepsilon\int_{(\mathbb{R}^+)^d}e^{-\langle \boldsymbol{r},\boldsymbol{s}\rangle}(1-e^{-\sum_{l\in S}\langle \boldsymbol{q}_l,\boldsymbol{s}\rangle})\eta'_{t_0}(\boldsymbol{s})\mathrm{d}\boldsymbol{s}+o(\varepsilon)\}\right)\\
&=-e^{-\psi_{t_0}(r_1,\dots,r_d)+\psi_{t_0-\varepsilon}(r_1,\dots,r_d)}\left(\sum_{\emptyset\neq S\subset I}(-1)^{\#(S)}\varepsilon\int_{(\mathbb{R}^+)^d}e^{-\langle \boldsymbol{r},\boldsymbol{s}\rangle}(1-e^{-\sum_{l\in S}\langle \boldsymbol{q}_l,\boldsymbol{s}\rangle})\eta'_{t_0}(\boldsymbol{s})\mathrm{d}\boldsymbol{s}\right.\\
&\quad\left.+o(\varepsilon)\right)\\
&=-e^{-\psi_{t_0}(r_1,\dots,r_d)+\psi_{t_0-\varepsilon}(r_1,\dots,r_d)}\left(\varepsilon\int_{(\mathbb{R}^+)^d}e^{-\langle \boldsymbol{r},\boldsymbol{s}\rangle}\sum_{\emptyset\neq S\subset I}(-1)^{\#(S)}(1-e^{-\sum_{l\in S}\langle \boldsymbol{q}_l,\boldsymbol{s}\rangle})\eta'_{t_0}(\boldsymbol{s})\mathrm{d}\boldsymbol{s}\right.\\
&\quad\left.+o(\varepsilon)\right)\\
&=-e^{-\psi_{t_0}(r_1,\dots,r_d)+\psi_{t_0-\varepsilon}(r_1,\dots,r_d)}\left(\varepsilon\int_{(\mathbb{R}^+)^d}e^{-\langle \boldsymbol{r},\boldsymbol{s}\rangle}(-1-\sum_{\emptyset\neq S\subset I}(-1)^{\#(S)}e^{-\sum_{l\in S}\langle \boldsymbol{q}_l,\boldsymbol{s}\rangle})\eta'_{t_0}(\boldsymbol{s})\mathrm{d}\boldsymbol{s}\right.\\
&\quad\left.+o(\varepsilon)\right)\\
&=(1+o(1))\left(\varepsilon\int_{(\mathbb{R}^+)^d}e^{-\langle \boldsymbol{r},\boldsymbol{s}\rangle}(1+\sum_{\emptyset\neq S\subset I}(-1)^{\#(S)}e^{-\sum_{l\in S}\langle \boldsymbol{q}_l,\boldsymbol{s}\rangle})\eta'_{t_0}(\boldsymbol{s})\mathrm{d}\boldsymbol{s}+o(\varepsilon)\right)\\
&=\varepsilon\int_{(\mathbb{R}^+)^d}e^{-\langle \boldsymbol{r},\boldsymbol{s}\rangle}(1+\sum_{\emptyset\neq S\subset I}(-1)^{\#(S)}e^{-\sum_{l\in S}\langle \boldsymbol{q}_l,\boldsymbol{s}\rangle})\eta'_{t_0}(\boldsymbol{s})\mathrm{d}\boldsymbol{s}+o(\varepsilon)\\
&=\varepsilon\int_{(\mathbb{R}^+)^d}e^{-\langle \boldsymbol{r},\boldsymbol{s}\rangle}(\sum_{S\subset I}(-1)^{\#(S)}e^{-\sum_{l\in S}\langle \boldsymbol{q}_l,\boldsymbol{s}\rangle})\eta'_{t_0}(\boldsymbol{s})\mathrm{d}\boldsymbol{s}+o(\varepsilon)\\
&=\varepsilon\int_{(\mathbb{R}^+)^d}e^{-\langle \boldsymbol{r},\boldsymbol{s}\rangle}\prod_{l\in I}\left(1-e^{-\langle \boldsymbol{q}_l,\boldsymbol{s}\rangle}\right)\eta'_{t_0}(\boldsymbol{s})\mathrm{d}\boldsymbol{s}+o(\varepsilon)
\end{aligned}
$$

□

Similarly to the proof of Theorem 12 we use the next Lemma to simplify the calculations in the proof of the Theorem at hand.

Lemma 8. *In the setting of Theorem 13 suppose the survival data with covariates* $\mathscr{D}_{\hat{\boldsymbol{X}}}$ *is comprised of a sole observation* t_1 *with associated covariates* $\boldsymbol{x}_l$ $l \in I$ *for some finite set*

$\emptyset \neq I \subset \mathbb{N}$. *Let* $t < t_1$*; then*

$$\mathbb{E}\left[e^{-\lambda_1\mu_1(0,t]-\ldots-\lambda_d\mu_d(0,t]}|\mathscr{D}_d\right] = \lim_{\varepsilon\to 0}\frac{\mathbb{E}\left[e^{-(\lambda+n^c+n^e)\mu(0,t]}\right]}{\mathbb{E}\left[e^{-(n^c+n^e)\mu(0,t]}\right]}$$

Proof.

$$\mathbb{E}\left[e^{-\lambda_1\mu_1(0,t]-\ldots-\lambda_d\mu_d(0,t]}|\mathscr{D}_{\hat{\boldsymbol{X}}}\right] =$$

$$\lim_{\varepsilon\to 0}\frac{\mathbb{E}\left[\prod_{i=1}^d\left(e^{-\lambda_i\mu_i(0,t]-h_i^{(c)}(\boldsymbol{\beta},\hat{\boldsymbol{X}})\mu_i(0,t_1]}\right)\prod_{l\in I}\left(\prod_{i=1}^d e^{-f_i(\boldsymbol{\beta x})\mu_i(0,t_1-\varepsilon]}-\prod_{i=1}^d e^{-f_i(\boldsymbol{\beta x})\mu_i(0,t_1]}\right)\right]}{\mathbb{E}\left[\prod_{i=1}^d\left(e^{-h_i^{(c)}(\boldsymbol{\beta},\hat{\boldsymbol{X}})\mu_i(0,t_1]}\right)\prod_{l\in I}\left(\prod_{i=1}^d e^{-f_i(\boldsymbol{\beta x})\mu_i(0,t_1-\varepsilon]}-\prod_{i=1}^d e^{-f_i(\boldsymbol{\beta x})\mu_i(0,t_1]}\right)\right]}$$

$$= \lim_{\varepsilon\to 0}\left\{\frac{\mathbb{E}\left[e^{-\lambda_i\mu_i(0,t]-h_i^{(c)}(\boldsymbol{\beta},\hat{\boldsymbol{X}})\mu_i(0,t]-h_i^{(e)}(\boldsymbol{\beta},\hat{\boldsymbol{X}})\mu_i(0,t]}\right]}{\mathbb{E}\left[e^{-h_i^{(c)}(\boldsymbol{\beta},\hat{\boldsymbol{X}})\mu_i(0,t]-h_i^{(e)}(\boldsymbol{\beta},\hat{\boldsymbol{X}})\mu_i(0,t]}\right]}\right.$$

$$\left.\times\frac{\mathbb{E}\left[\prod_{i=1}^d\left(e^{-h_i^{(c)}(\boldsymbol{\beta},\hat{\boldsymbol{X}})\mu_i(t,t_1]}\right)\prod_{l\in I}\left(\prod_{i=1}^d e^{-f_i(\boldsymbol{\beta x})\mu_i(t,t_1-\varepsilon]}-\prod_{i=1}^d e^{-f_i(\boldsymbol{\beta x})\mu_i(t,t_1]}\right)\right]}{\mathbb{E}\left[\prod_{i=1}^d\left(e^{-h_i^{(c)}(\boldsymbol{\beta},\hat{\boldsymbol{X}})\mu_i(t,t_1]}\right)\prod_{l\in I}\left(\prod_{i=1}^d e^{-f_i(\boldsymbol{\beta x})\mu_i(t,t_1-\varepsilon]}-\prod_{i=1}^d e^{-f_i(\boldsymbol{\beta x})\mu_i(t,t_1]}\right)\right]}\right\}$$

$$= \frac{\mathbb{E}\left[\prod_{i=1}^d e^{-(\lambda_i+h_i^{(c)}(\boldsymbol{\beta},\hat{\boldsymbol{X}})+h_i^{(e)}(\boldsymbol{\beta},\hat{\boldsymbol{X}}))\mu_i(0,t]}\right]}{\mathbb{E}\left[\prod_{i=1}^d e^{-(h_i^{(c)}(\boldsymbol{\beta},\hat{\boldsymbol{X}})+h_i^{(e)}(\boldsymbol{\beta},\hat{\boldsymbol{X}}))\mu_i(0,t]}\right]} = e^{-\left(\psi_t(\boldsymbol{\lambda}+\boldsymbol{h}^{(c)}(\boldsymbol{\beta},\hat{\boldsymbol{X}})+\boldsymbol{h}^{(e)}(\boldsymbol{\beta},\hat{\boldsymbol{X}}))-\psi_t\left(\boldsymbol{h}^{(c)}(\boldsymbol{\beta},\hat{\boldsymbol{X}})+\boldsymbol{h}^{(e)}(\boldsymbol{\beta},\hat{\boldsymbol{X}})\right)\right)}$$

□

We observe that if ψ_t is the Laplace exponent associated to the Lévy measure $\tilde{\nu}_d$ then $\psi_t^{(\boldsymbol{k})}(\boldsymbol{\lambda}) = \psi_t(\boldsymbol{\lambda}+\boldsymbol{k}) - \psi_t(\boldsymbol{k})$ is the Laplace exponent associated to $e^{-\langle \boldsymbol{k},\boldsymbol{s}\rangle}\tilde{\nu}_d(\mathrm{d}\boldsymbol{s},\mathrm{d}x)$, with this in mind we have that, in the following, without loss of generality for calculations of the posterior Laplace exponent as in the previous Lemma it suffices to consider evaluation of the exponent in a time t which is greater than all the survival times in the survival data $\mathscr{D}_{\hat{\boldsymbol{X}}}$.
Define

$$\Gamma_{\mathscr{D}_{\hat{\boldsymbol{X}}},\varepsilon} = \bigcap_{j=1}^k\{(T_1,J_1,\boldsymbol{X}_1,\ldots,T_n,J_n,\mathbb{P},bX_n)\}$$

so that

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

We observe that selecting ε sufficiently small such that $t \notin (T_{(j)}-\varepsilon, T_{(j)})$ for all $j \in \{1,\ldots,k\}$

$$
\begin{aligned}
&\mathbb{E}\Big[e^{-\lambda_1\mu_1(0,t]-\cdots\lambda_d\mu_d(0,t]}\mathbb{1}_{\Gamma_{\mathscr{D}_{\hat{\boldsymbol{X}}},\varepsilon}}\big(\mathscr{D}_{\hat{\boldsymbol{X}}}\big)\,|(\mu_1,\ldots,\mu_d)\Big]\\
&=\mathbb{E}\left[\left(\prod_{i=1}^{d}e^{-\lambda_i\mu_i(0,t]}\right)\prod_{j=1}^{k}\left\{\prod_{l\in I_j^{(e)}}\left(\prod_{i=1}^{d}e^{-f_i(\boldsymbol{\beta},\boldsymbol{X}_l)\mu_i(0,T_{(j)}-\varepsilon]}-\prod_{i=1}^{d}e^{-f_i(\boldsymbol{\beta},\boldsymbol{X}_l)\mu_i(0,T_{(j)}]}\right)\right.\right.\\
&\quad\left.\left.\times\prod_{i=1}^{d}\left(e^{-h_{j,i}^{(c)}(\boldsymbol{\beta},\hat{\boldsymbol{X}})\mu_i(0,T_{(j)}]}\right)\right\}\right]\\
&=\mathbb{E}\left[\left(\prod_{i=1}^{d}e^{-\lambda_i\mu_i(0,t]}\right)\prod_{j=1}^{k}\left\{\prod_{i=1}^{d}\left(e^{-h_{j,i}^{(e)}(\boldsymbol{\beta},\hat{\boldsymbol{X}})\mu_i(0,T_{(j)}-\varepsilon]}\right)\right.\right.\\
&\quad\left.\left.\times\prod_{l\in I_j^{(e)}}\left(1-\prod_{i=1}^{d}e^{-f_i(\boldsymbol{\beta},\boldsymbol{X}_l)\mu_i(T_{(j-1)}-\varepsilon,T_{(j)}]}\right)\prod_{i=1}^{d}\left(e^{-h_{j,i}^{(c)}(\boldsymbol{\beta},\hat{\boldsymbol{X}})\mu_i(0,T_{(j)}]}\right)\right\}\right]\\
&=\mathbb{E}\left[\left(\prod_{i=1}^{d}e^{-\lambda_i\left(\sum_{j=1}^{k}\{\mu_i(T_{(j-1)},T_{(j)}-\varepsilon]+\mu_i(T_{(j)}-\varepsilon,T_{(j)}]\}+\mu_i(T_{(k)},t]\right)}\right)\right.\\
&\quad\times\prod_{j=1}^{k}\left\{\prod_{i=1}^{d}\left(e^{-h_{j,i}^{(e)}(\boldsymbol{\beta},\hat{\boldsymbol{X}})\mu_i(0,T_{(j)}-\varepsilon]}\right)\right.\\
&\quad\left.\left.\times\prod_{l\in I_j^{(e)}}\left(1-\prod_{i=1}^{d}e^{-f_i(\boldsymbol{\beta},\boldsymbol{X}_l)\mu_i(T_{(j-1)}-\varepsilon,T_{(j)}]}\right)\prod_{i=1}^{d}\left(e^{-h_{j,i}^{(c)}(\boldsymbol{\beta},\hat{\boldsymbol{X}})\mu_i(0,T_{(j)}]}\right)\right\}\right]\\
&=\mathbb{E}\left[\left(\prod_{i=1}^{d}e^{-\lambda_i\left(\sum_{j=1}^{k}\{\mu_i(T_{(j-1)},T_{(j)}-\varepsilon]+\mu_i(T_{(j)}-\varepsilon,T_{(j)}]\}+\mu_i(T_{(k)},t]\right)}\right)\right.\\
&\quad\times\prod_{j=1}^{k}\left\{\prod_{i=1}^{d}\left(e^{-h_{j,i}^{(e)}(\boldsymbol{\beta},\hat{\boldsymbol{X}})\left(\sum_{r=1}^{j}\mu_i(T_{(r-1)},T_{(r)}-\varepsilon]+\sum_{r=1}^{j-1}\mu_i(T_{(r)}-\varepsilon,T_{(r)}]\right)}\right)\right.\\
&\quad\times\prod_{l\in I_j^{(e)}}\left(1-\prod_{i=1}^{d}e^{-f_i(\boldsymbol{\beta},\boldsymbol{X}_l)\mu_i(T_{(j-1)}-\varepsilon,T_{(j)}]}\right)\\
&\quad\left.\left.\times\prod_{i=1}^{d}\left(e^{-h_{j,i}^{(c)}(\boldsymbol{\beta},\hat{\boldsymbol{X}})\left(\sum_{r=1}^{j}\mu_i(T_{(r-1)},T_{(r)}-\varepsilon]+\sum_{r=1}^{j}\mu_i(T_{(r)}-\varepsilon,T_{(r)}]\right)}\right)\right\}\right]
\end{aligned}
$$

$$
\begin{aligned}
&= \mathbb{E}\left[\left(\prod_{i=1}^{d} e^{-\lambda_i \mu_i(T_{(k)},t]} \prod_{j=1}^{k} e^{-\lambda_i \mu_i(T_{(j-1)},T_{(j)}-\varepsilon]-\lambda_i \mu_i(T_{(j)}-\varepsilon,T_{(j)}]}\right)\right.\\
&\quad \times \prod_{i=1}^{d}\left(e^{-\sum_{j=1}^{k} h_{j,i}^{(e)}(\boldsymbol{\beta},\hat{\boldsymbol{X}})\left(\sum_{r=1}^{j}\mu_i(T_{(r-1)},T_{(r)}-\varepsilon]+\sum_{r=1}^{j-1}\mu_i(T_{(r)}-\varepsilon,T_{(r)}]\right)}\right)\\
&\quad \times \prod_{j=1}^{k}\left\{\prod_{l\in I_j^{(e)}}\left(1-\prod_{i=1}^{d} e^{-f_i(\boldsymbol{\beta},\boldsymbol{X}_l)\mu_i(T_{(j-1)}-\varepsilon,T_{(j)}]}\right)\right\}\\
&\quad \left.\times \prod_{i=1}^{d}\left(e^{-\sum_{j=1}^{k} h_{j,i}^{(c)}(\boldsymbol{\beta},\hat{\boldsymbol{X}})\left(\sum_{r=1}^{j}\mu_i(T_{(r-1)},T_{(r)}-\varepsilon]+\sum_{r=1}^{j}\mu_i(T_{(r)}-\varepsilon,T_{(r)}]\right)}\right)\right]\\
&= \mathbb{E}\left[\left(\prod_{i=1}^{d} e^{-\lambda_i \mu_i(T_{(k)},t]} \prod_{j=1}^{k} e^{-\lambda_i \mu_i(T_{(j-1)},T_{(j)}-\varepsilon]-\lambda_i \mu_i(T_{(j)}-\varepsilon,T_{(j)}]}\right)\right.\\
&\quad \times \prod_{i=1}^{d}\left(e^{-\sum_{r=1}^{k} \bar{h}_{r,i}^{(e)}(\boldsymbol{\beta},\hat{\boldsymbol{X}})\mu_i(T_{(r-1)},T_{(r)}-\varepsilon]-\sum_{r=1}^{k-1}\bar{h}_{r+1,i}^{(e)}(\boldsymbol{\beta},\hat{\boldsymbol{X}})\mu_i(T_{(r)}-\varepsilon,T_{(r)}]}\right)\\
&\quad \times \prod_{j=1}^{k}\left\{\prod_{l\in I_j^{(e)}}\left(1-\prod_{i=1}^{d} e^{-f_i(\boldsymbol{\beta},\boldsymbol{X}_l)\mu_i(T_{(j-1)}-\varepsilon,T_{(j)}]}\right)\right\}\\
&\quad \left.\times \prod_{i=1}^{d}\left(e^{-\sum_{r=1}^{k} \bar{h}_{r,i}^{(c)}(\boldsymbol{\beta},\hat{\boldsymbol{X}})\mu_i(T_{(r-1)},T_{(r)}-\varepsilon]-\sum_{r=1}^{k}\bar{h}_{r,i}^{(c)}(\boldsymbol{\beta},\hat{\boldsymbol{X}})\mu_i(T_{(r)}-\varepsilon,T_{(r)}]}\right)\right]\\
&= \mathbb{E}\left[\left(\prod_{i=1}^{d} e^{-\lambda_i \mu_i(T_{(k)},t]}\right)\prod_{j=1}^{k}\left\{\prod_{i=1}^{d}\left(e^{-\lambda_i \mu_i(T_{(j-1)},T_{(j)}-\varepsilon]-\lambda_i \mu_i(T_{(j)}-\varepsilon,T_{(j)}]}\right)\right.\right.\\
&\quad \times \prod_{i=1}^{d}\left(e^{-\bar{h}_{j,i}^{(e)}(\boldsymbol{\beta},\hat{\boldsymbol{X}})\mu_i(T_{(j-1)},T_{(j)}-\varepsilon]-\bar{h}_{j+1,i}^{(c)}(\boldsymbol{\beta},\hat{\boldsymbol{X}})\mu_i(T_{(j)}-\varepsilon,T_{(j)}]}\right)\prod_{l\in I_j^{(e)}}\left(1-\prod_{i=1}^{d} e^{-f_i(\boldsymbol{\beta},\boldsymbol{X}_l)\mu_i(T_{(j-1)}-\varepsilon,T_{(j)}]}\right)\\
&\quad \left.\left.\times \prod_{i=1}^{d}\left(e^{-\bar{h}_{j,i}^{(c)}(\boldsymbol{\beta},\hat{\boldsymbol{X}})\mu_i(T_{(j-1)},T_{(j)}-\varepsilon]-\bar{h}_{j,i}^{(c)}(\boldsymbol{\beta},\hat{\boldsymbol{X}})\mu_i(T_{(j)}-\varepsilon,T_{(j)}]}\right)\right\}\right]
\end{aligned}
$$

So defining

$$
\begin{aligned}
I_{1,\varepsilon} = \prod_{j=1}^{k} \Bigg\{ \prod_{i=1}^{d} \left(e^{-\left(\lambda_i + \bar{h}^{(e)}_{j+1,i}(\boldsymbol{\beta},\hat{\boldsymbol{X}}) + \bar{h}^{(c)}_{j,i}(\boldsymbol{\beta},\hat{\boldsymbol{X}})\right)\mu_i(T_{(j)}-\varepsilon,T_{(j)}]} \right) \\
\times \prod_{l\in I^{(e)}_j} \left(1 - \prod_{i=1}^{d} e^{-f_i(\boldsymbol{\beta},\boldsymbol{X}_l)\mu_i(T_{(j-1)}-\varepsilon,T_{(j)}]} \right) \Bigg\} \\
I_{2,\varepsilon} = \left(\prod_{i=1}^{d} e^{-\lambda_i\mu_i(T_{(k)},t]} \right) \prod_{j=1}^{k}\prod_{i=1}^{d} \Bigg(e^{-\lambda_i\mu_i(T_{(j-1)},\,T_{(j)}-\varepsilon]} \\
\times e^{-\left(\bar{h}^{(e)}_{j,i}(\boldsymbol{\beta},\hat{\boldsymbol{X}}) + \bar{h}^{(c)}_{j,i}(\boldsymbol{\beta},\hat{\boldsymbol{X}})\right)\mu_i(T_{(j-1)},T_{(j)}-\varepsilon]} \Bigg)
\end{aligned}
$$

We get from the independence property of CRM's that

$$
\mathbb{E}\left[e^{-\lambda_1\mu_1(0,t]-\cdots-\lambda_d\mu_d(0,t]} \mathbb{1}_{\Gamma_{\boldsymbol{D},\varepsilon}}\left(\mathscr{D}_{\hat{\boldsymbol{X}}}\right) \right] = \mathbb{E}[I_{1,\varepsilon}]\,\mathbb{E}[I_{2,\varepsilon}] \tag{5.16}
$$

We observe that for $r_i = \lambda_i + \bar{h}^{(e)}_{j+1,i}(\boldsymbol{\beta},\hat{\boldsymbol{X}}) + \bar{h}^{(c)}_{j,i}(\boldsymbol{\beta},\hat{\boldsymbol{X}}),\ i \in \{1,\ldots,d\}$, we set $q_{l,i} = f_i(\boldsymbol{\beta},\boldsymbol{X}_l)$ and observe that $\min\{r_1,\ldots,r_d\} > 0$ and $\max\{q_{1,i},\ldots,q_{l,i}\} > 0$ so Lemma 7 can be applied yielding

$$
\begin{aligned}
&\mathbb{E}\left[\prod_{i=1}^{d} \left(e^{-\left(\lambda_i + \bar{h}^{(e)}_{j+1,i}(\boldsymbol{\beta},\hat{\boldsymbol{X}}) + \bar{h}^{(c)}_{j,i}(\boldsymbol{\beta},\hat{\boldsymbol{X}})\right)\mu_i(T_{(j)}-\varepsilon,T_{(j)}]} \right) \prod_{l\in I^{(e)}_j} \left(1 - \prod_{i=1}^{d} e^{-f_i(\boldsymbol{\beta},\boldsymbol{X}_l)\mu_i(T_{(j-1)}-\varepsilon,T_{(j)}]} \right) \right] \\
&= \varepsilon \int_{(\mathbb{R}^+)^d} \prod_{i=1}^{d} \left(e^{-\left(\lambda_i + \bar{h}^{(e)}_{j+1,i}(\boldsymbol{\beta},\hat{\boldsymbol{X}}) + \bar{h}^{(c)}_{j,i}(\boldsymbol{\beta},\hat{\boldsymbol{X}})\right)s_i} \right) \prod_{l\in I^{(e)}_j} \left(1 - \prod_{i=1}^{d} e^{-f_i(\boldsymbol{\beta},\boldsymbol{X}_l)s_i} \right) \eta'_{T_{(j)}}(\boldsymbol{s})\mathrm{d}\boldsymbol{s} + o(\varepsilon)
\end{aligned} \tag{5.17}
$$

On the other hand, for $j \notin \mathscr{J} = \{j : T_{(j)} \text{ is an exact observation}\}$ we have $I^{(e)}_j = \emptyset$ so by the continuity of $\eta_t(\boldsymbol{s})$ in t we get that

$$
\lim_{\varepsilon\to 0} \mathbb{E}\left[\prod_{i=1}^{d} \left(e^{-\left(\lambda_i + \bar{h}^{(e)}_{j+1,i}(\boldsymbol{\beta},\hat{\boldsymbol{X}}) + \bar{h}^{(c)}_{j,i}(\boldsymbol{\beta},\hat{\boldsymbol{X}})\right)\mu_i(T_{(j)}-\varepsilon,T_{(j)}]} \right) \right] = 1 \tag{5.18}
$$

From (5.17), (5.18) and the independence property of CRM's we obtain

$$\lim_{\varepsilon\to 0}\mathbb{E}[I_{1,\varepsilon}] = \lim_{\varepsilon\to 0}\prod_{j\in I^{(e)}}\left\{\varepsilon\int_{(\mathbb{R}^+)^d}\prod_{i=1}^{d}\left(e^{-\left(\lambda_i+\bar{h}^{(e)}_{j+1,i}(\boldsymbol{\beta},\hat{\boldsymbol{X}})+\bar{h}^{(c)}_{j,i}(\boldsymbol{\beta},\hat{\boldsymbol{X}})\right)s_i}\right)\right.$$
$$\left.\times\prod_{l\in I^{(e)}_j}\left(1-\prod_{i=1}^{d}e^{-f_i(\boldsymbol{\beta},\boldsymbol{X}_l)s_i}\right)\eta'_{T^{(j)}}(\boldsymbol{s})\mathrm{d}\boldsymbol{s}+o(\varepsilon)\right\}$$

Also by continuity and independence we get

$$\lim_{\varepsilon\to 0}\mathbb{E}[I_{2,\varepsilon}] = e^{-[\psi_t(\boldsymbol{\lambda})-\psi_{T_{(k)}}(\boldsymbol{\lambda})]}\prod_{j=1}^{k}e^{-[\psi_{T_{(j)}}(\boldsymbol{\lambda}+\bar{\boldsymbol{h}}^{(e)}_j(\boldsymbol{\beta},\hat{\boldsymbol{X}})+\bar{\boldsymbol{h}}^{(c)}_j(\boldsymbol{\beta},\hat{\boldsymbol{X}}))-\psi_{T_{(j-1)}}(\boldsymbol{\lambda}+\bar{\boldsymbol{h}}^{(e)}_j(\boldsymbol{\beta},\hat{\boldsymbol{X}})+\bar{\boldsymbol{h}}^{(c)}_j(\boldsymbol{\beta},\hat{\boldsymbol{X}}))]}$$

So by (5.16) we have that

$$\lim_{\varepsilon\to 0}\mathbb{E}\left[e^{-\lambda_1\mu_1(0,t]-\cdots-\lambda_d\mu_d(0,t]}\mathbb{1}_{\Gamma_{\mathscr{D}_{\hat{\boldsymbol{X}}},\varepsilon}}\left(\mathscr{D}_{\hat{\boldsymbol{X}}}\right)\right]=$$
$$e^{-[\psi_t(\boldsymbol{\lambda})-\psi_{T_{(k)}}(\boldsymbol{\lambda})]}\prod_{j=1}^{k}e^{-[\psi_{T_{(j)}}(\boldsymbol{\lambda}+\bar{\boldsymbol{h}}^{(e)}_j(\boldsymbol{\beta},\hat{\boldsymbol{X}})+\bar{\boldsymbol{h}}^{(c)}_j(\boldsymbol{\beta},\hat{\boldsymbol{X}}))-\psi_{T_{(j-1)}}(\boldsymbol{\lambda}+\bar{\boldsymbol{h}}^{(e)}_j(\boldsymbol{\beta},\hat{\boldsymbol{X}})+\bar{\boldsymbol{h}}^{(c)}_j(\boldsymbol{\beta},\hat{\boldsymbol{X}}))]}$$
$$\times\prod_{j\in I^{(e)}}\lim_{\varepsilon\to 0}\left\{\varepsilon\int_{(\mathbb{R}^+)^d}\prod_{i=1}^{d}\left(e^{-\left(\lambda_i+\bar{h}^{(e)}_{j+1,i}(\boldsymbol{\beta},\hat{\boldsymbol{X}})+\bar{h}^{(c)}_{j,i}(\boldsymbol{\beta},\hat{\boldsymbol{X}})\right)s_i}\right)\right.$$
$$\left.\prod_{l\in I^{(e)}_j}\left(1-\prod_{i=1}^{d}e^{-f_i(\boldsymbol{\beta},\boldsymbol{X}_l)s_i}\right)\eta'_{T^{(j)}}(\boldsymbol{s})\mathrm{d}\boldsymbol{s}+o(\varepsilon)\right\}$$

And similarly

$$\lim_{\varepsilon\to 0}\mathbb{P}\left[\boldsymbol{D}\in\Gamma_{\mathscr{D}_{\hat{\boldsymbol{X}}},\varepsilon}\right]=$$
$$\prod_{j=1}^{k}e^{-[\psi_{T_{(j)}}(\bar{\boldsymbol{h}}^{(e)}_j(\boldsymbol{\beta},\hat{\boldsymbol{X}})+\bar{\boldsymbol{h}}^{(c)}_j(\boldsymbol{\beta},\hat{\boldsymbol{X}}))-\psi_{T_{(j-1)}}(\bar{\boldsymbol{h}}^{(e)}_j(\boldsymbol{\beta},\hat{\boldsymbol{X}})+\bar{\boldsymbol{h}}^{(c)}_j(\boldsymbol{\beta},\hat{\boldsymbol{X}}))]}$$
$$\times\lim_{\varepsilon\to 0}\varepsilon^{\#(I^{(e)})}\prod_{j\in I^{(e)}}\left\{\int_{(\mathbb{R}^+)^d}\prod_{i=1}^{d}\left(e^{-\left(\bar{h}^{(e)}_{j+1,i}(\boldsymbol{\beta},\hat{\boldsymbol{X}})+\bar{h}^{(c)}_{j,i}(\boldsymbol{\beta},\hat{\boldsymbol{X}})\right)s_i}\right)\right.$$
$$\left.\prod_{l\in I^{(e)}_j}\left(1-\prod_{i=1}^{d}e^{-f_i(\boldsymbol{\beta},\boldsymbol{X}_l)s_i}\right)\eta'_{T^{(j)}}(\boldsymbol{s})\mathrm{d}\boldsymbol{s}+o(\varepsilon)\right\}\tag{5.19}$$

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

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

$$= e^{-\sum_{j=1}^{k+1} \triangle_{T_{(j-1)}}^{T_{(j)}} \left(\psi_t(\boldsymbol{\lambda}+\bar{h}_j^{(c)}(\boldsymbol{\beta},\hat{\boldsymbol{X}})+\bar{h}_j^{(e)}(\boldsymbol{\beta},\hat{\boldsymbol{X}}))-\psi_t(\bar{h}_j^{(c)}(\boldsymbol{\beta},\hat{\boldsymbol{X}})+\bar{h}_j^{(e)}(\boldsymbol{\beta},\hat{\boldsymbol{X}}))\right)}$$

$$\prod_{j\in I^{(e)}} \frac{\int_{(\mathbb{R}^+)^d} \prod_{i=1}^d \left(e^{-\left(\lambda_i+\bar{h}_{j+1,i}^{(e)}(\boldsymbol{\beta},\hat{\boldsymbol{X}})+\bar{h}_{j,i}^{(c)}(\boldsymbol{\beta},\hat{\boldsymbol{X}})\right)s_i}\right) \prod_{l\in I_j^{(e)}} \left(1-\prod_{i=1}^d e^{-f_i(\boldsymbol{\beta},\boldsymbol{X}_l)s_i}\right) \eta'_{T(j)}(\boldsymbol{s})\mathrm{d}\boldsymbol{s}}{\int_{(\mathbb{R}^+)^d} \prod_{i=1}^d \left(e^{-\left(\bar{h}_{j+1,i}^{(e)}(\boldsymbol{\beta},\hat{\boldsymbol{X}})+\bar{h}_{j,i}^{(c)}(\boldsymbol{\beta},\hat{\boldsymbol{X}})\right)s_i}\right) \prod_{l\in I_j^{(e)}} \left(1-\prod_{i=1}^d e^{-f_i(\boldsymbol{\beta},\boldsymbol{X}_l)s_i}\right) \eta'_{T(j)}(\boldsymbol{s})\mathrm{d}\boldsymbol{s}} \tag{5.20}$$

Proof of Corollary 7

In equation (5.20) above take $\boldsymbol{\lambda} = (1,\ldots,1)$ and using the discussion after Lemma 8 replace $I^{(e)}$ with $I_t^{(e)}$ to obtain the result.

Proof of Corollary 8

From equation (5.19) in the proof of Theorem 13 we obtain the related likelihood.

Proof of Proposition 13

The result follows from a straightforward application of Lemma 6, in the proof of Proposition 12.

Proof of Proposition 14

Let $\boldsymbol{K}$ and $\boldsymbol{\mu}$ be as in the hypothesis. We use the next Lemma to deal with the effect of $\boldsymbol{K}$ in the Lévy intensity.

Lemma 9. *Let $S \sim NTR(K\mu)$ such that μ is an homogeneous gamma CRM with Lévy intensity $\nu(\mathrm{d}s,\mathrm{d}x) = \rho(\mathrm{d}s;\alpha,\beta)\kappa(\mathrm{d}x)$ with κ bounded and strictly positive and K an independent random variable with probability density supported in $\mathbb{R}^+$, then the posterior distribution of $S|K$ is consistent with respect to S_0.*

Proof. The Lévy intensity of $K\mu|K$ is given by

$$\nu_K(A,B) = \int_A \frac{\nu\left(\frac{z}{K},B\right)}{K}\mathrm{d}z = \kappa(B)\int_A \frac{\beta e^{-\frac{\alpha z}{K}}}{z}\mathrm{d}z$$

for $A \in \mathbb{R}^+$ so the hazard rate of S has absolutely continuous w.r.t. Lebesgue measure Lévy intensity given by

$$\xi_K(s,x) = \frac{\nu\left(\frac{-\log(1-s)}{K}, x\right)}{K(1-s)} = \frac{\kappa(x)\beta(1-s)^{\frac{\alpha}{K}}}{-\log(1-s)(1-s)}.$$

As K has probability zero of being zero, the conditions of Proposition 5 follow from Example 6. □

Furthermore we have the next Lemma to deal with the independent entries of the VCRM in Proposition 14.

Lemma 10. *Let $S \sim NTR(\sum_{i=1}^d \mu_i)$. If $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_d)$ has mutually independent entries and for each $i \in \{1, \ldots, d\}$ μ_i has Lévy intensity ν_i such that*

$$\xi_i(s,x) = \frac{\nu_i(-\log(1-s), x)}{1-s}$$

satisfies Proposition 1, then S is consistent.

Proof. Let $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_d)$ be a VCRM with independent entries such that each μ_i has a Lévy intensity $\nu_i(s, \mathrm{d}x)\mathrm{d}s$ then $\mu_+ = \sum_{i=1}^d \mu_i$ has a Lévy intensity ρ_+ given by

$$\nu_+(A,B) = \nu_1(A,B) + \ldots + \nu_d(A,B).$$

It follows that $\nu_+(\mathrm{d}s, \mathrm{d}x) = \nu_+(s, \mathrm{d}x)\mathrm{d}s = (\nu_1(s, \mathrm{d}x) + \ldots + \nu_d(s, \mathrm{d}x))\,\mathrm{d}s$. An application of Proposition 5 implies that $S \sim \mathrm{NTR}(\sum_{i=1}^d \mu_i)$ is consistent. □

So from Lemma 9 we have that $\boldsymbol{\mu_K} = (K_1\mu_1, \ldots, K_d\mu_d)$ has marginals which satisfy the conditions of Lemma 10, from which the consistency of $S|\boldsymbol{K} \sim \mathrm{NTR}(\sum_{i=1}^d K_i\mu_i)$ follows.

Proof of Proposition 15

We have that $S|\boldsymbol{K} \sim \mathrm{NTR}((K_1 + \ldots + K_d)\mu)$ where μ is a Gamma CRM and $K = K_1 + \ldots K_d$ is a random variable with probability density function supported in $\mathbb{R}^+$ so the result follows from Lemma 9.

Proof of Corollary 9

As each f_i regressor function is strictly positive, the result follows from Propositions 14 and 15 in view of identity (5.2).

Proof of Proposition 16

We make use of the next Lemma for this proof Furthermore for fixed $t \in \mathbb{R}^+$ the r.v. $S(t;\tilde{\boldsymbol{X}})$ is bounded, so it is uniformly integrable, and converges in probability to $S_0(t)$ hence it converges in the mean as well.

Lemma 11. *Let $\mu = \sum_{i=1}^{\infty} w_i \delta_{u_i}$ be an homogeneous Gamma(α,β) CRM in $\mathbb{X}$ for some $\kappa(x)$ and $\{v_i\}_{i=1}^{\infty}$ be an i.i.d. sequence of random variables with common probability density g supported in $\mathbb{R}^+$, then $S \sim NTR\left(\sum_{i=1}^{\infty} v_i w_i \delta_{u_i}\right)$ is consistent.*

Proof. Using Proposition 1 we can check that the Lévy intensity of $\sum_{i=1}^{\infty} v_i w_i \delta_{u_i}$ is given by

$$\nu_{\mathscr{L}(V)}(A,B) = \mathbb{E}\left[\int_A \frac{\rho(\frac{z}{V};\alpha,\beta)\kappa(B)}{V}\mathrm{d}z\right]$$

for A in $\mathbb{R}^+$ and V a random variable given by the probability density g. So the hazard rate of S has Lévy intensity given by

$$\xi_{\mathscr{L}(V)}(s,\mathrm{d}x) = \mathbb{E}\left[\frac{\rho\left(\frac{-\log(1-s)}{V};\alpha,\beta\right)\kappa(\mathrm{d}x)}{V(1-s)}\right] = \beta\kappa(\mathrm{d}x)\mathbb{E}\left[\frac{(1-s)^{\frac{\alpha}{V}}}{-\log(1-s)(1-s)}\right].$$

Using Lebesgue dominated convergence theorem the conditions of Proposition 5 follow analogously to Example 6, ☐

Let $\sum_{i=1}^{\infty} w_i \delta_{u_i}$ be a series representation of the directing Gamma (α,β) Lévy measure of the CoRM $\boldsymbol{\mu}$ then

$$\sum_{j=1}^{d} K_j \mu_j = \sum_{i=1}^{\infty} (K_1 J_{1,i} + \ldots + K_d J_{d,i}) w_i \delta_{u_i}$$

where $\{(J_{1,i},\ldots,J_{d,i})\}_{i=1}^{\infty}$ are i.i.d. with distribution given by the score distribution h. It follows form the previous Lemma that $S|\boldsymbol{K} \sim \mathrm{NTR}\left(\sum_{i=1}^{d} K_d \mu_d\right)$ is consistent.

Proof of Corollary 10

As the regressor functions f_i are strictly positive the result follows from the Proposition 16 in view of identity (5.2).

Proof of Proposition 17

From Proposition 1 we get that the Lévy intensity of $\mu_1 + \ldots + \mu_d$ is given by $\hat{\nu}(\mathrm{d}z, \mathrm{d}x) = \hat{\nu}(z, \mathrm{d}x)\mathrm{d}z$ where

$$\hat{\nu}(z, \mathrm{d}x) = \int_0^z \int_0^{z-y_d} \cdots \int_0^{z-y_d-\ldots-y_3} \nu(z - y_2 - \ldots - y_d, y_2, \ldots, y_d)\mathrm{d}y_2 \cdots \mathrm{d}y_d. \quad (5.21)$$

Using again Proposition 1, the Lévy intensity of $(K_1\mu_1, \ldots, K_d\mu_d)$ given $\boldsymbol{K}$ is $\tilde{\nu}(\mathrm{d}\boldsymbol{s}, \mathrm{d}x) = \tilde{\nu}(\boldsymbol{s}, \mathrm{d}x)\mathrm{d}\boldsymbol{s}$ with

$$\tilde{\nu}(\boldsymbol{s}, \mathrm{d}x) = \frac{\nu\left(\frac{s_1}{K_1}, \ldots, \frac{s_d}{K_d}, \mathrm{d}x\right)}{K_1 \cdots K_d}. \quad (5.22)$$

The affirmation of Proposition 17 follows from Proposition 6 applied to the Lévy intensity obtained by applying (5.21) together with (5.22) and Proposition 4.

Proof of Theorem 14

If $t = 0$ the result is trivial so let $t > 0$. By hypothesis we have that

$$S_0(t) = \lim_{n\to\infty} e^{-\sum_{j=1}^{k(n)+1}\left(\psi^{\circ}_{t\wedge T_{(j)}}(\mathbf{1}) - \psi^{\circ}_{t\wedge T_{(j-1)}}(\mathbf{1})\right)\mathbb{1}_{\{T_{(j-1)}<t\}}} \prod_{j\in I_t^{(e),(n)}}$$
$$\frac{\int_{(\mathbb{R}^+)^d}\prod_{i=1}^d\left(e^{-\left(1++\bar{h}^{(e)}_{j+1,i}(\boldsymbol{\beta},\hat{\boldsymbol{X}}^{(n)})+\bar{h}^{(c)}_{j,i}(\boldsymbol{\beta},\hat{\boldsymbol{X}}^{(n)})\right)s_i}\right)\prod_{l\in I_j^{(e),\,(n)}}\left(1-\prod_{i=1}^d e^{-f_i(\boldsymbol{\beta},\boldsymbol{X}_l)s_i}\right)\nu(\boldsymbol{s})\mathrm{d}\boldsymbol{s}}{\int_{(\mathbb{R}^+)^d}\prod_{i=1}^d\left(e^{-\left(\bar{h}^{(e)}_{j+1,i}(\boldsymbol{\beta},\hat{\boldsymbol{X}}^{(n)})+\bar{h}^{(c)}_{j,i}(\boldsymbol{\beta},\hat{\boldsymbol{X}}^{(n)})\right)s_i}\right)\prod_{l\in I_j^{(e),(n)}}\left(1-\prod_{i=1}^d e^{-f_i(\boldsymbol{\beta},\boldsymbol{X}_l)s_i}\right)\nu(\boldsymbol{s})\mathrm{d}\boldsymbol{s}}$$

due to convergence in probability and uniform integrability. $S_0(t)$ and $\lim_{n\to\infty}\mathbb{E}\left[S^{\star}(t)|\mathscr{D}^{(n)}_{\hat{\boldsymbol{X}}}, \boldsymbol{\beta}, \boldsymbol{X}^{\star}\right]$ exist by hypothesis so R is well defined. From the definition of the estimator (5.8) the result follows.

Proof of Proposition18

We use the next Lemma for the proof.

Lemma 12. *Let $\mu = \sum_{i=1}^{\infty} w_i\delta_{u_i}$ be an homogeneous Gamma(α, β) CRM in $\mathbb{X}$ with positive and bounded κ and $\{V_i\}_{i=1}^{\infty}$ be an i.i.d. sequence of random variables with common probability density g supported in $\mathbb{R}^+$, then $S \sim$ NTR$(\sum_{i=1}^{\infty} v_iw_i\delta_{u_i})$ attains the Bernstein-von Mises conditions of Proposition 6 if $\mathbb{E}[V] < \infty$ and $\mathbb{E}\left[\frac{1}{V}\right] < \infty$ for V with distribution given by g.*

We recall from the proof of Proposition 16 that the associated hazard rate for the case at hand takes the form

$$\xi(s,x) = \mathbb{E}\left[\frac{\beta(1-s)^{\frac{\alpha}{V}}\kappa(x)}{-\log(1-s)(1-s)}\right].$$

where the distribution of V is given by g. We use $\log(x) \leq x-1$ and the Tonelli theorem to see that

$$0 < \lambda(x) = \int_0^1 \mathbb{E}\left[\frac{s\beta(1-s)^{\frac{\alpha}{V}}\kappa(x)}{-log(1-s)(1-s)}\mathrm{d}s\right] \leq \kappa(x)\mathbb{E}\left[\beta\int_0^1 (1-s)^{\alpha-1}\mathrm{d}s\right] = \frac{\kappa(x)\beta\mathbb{E}[V]}{\alpha} < \infty.$$

So the first condition of Proposition 6 follows. Similarly, taking $\lambda(x) = c\kappa(x)$ for some $c \in (0,\infty)$ we observe that

$$\gamma_k(s) = \mathbb{E}\left[\frac{\beta s(1-s)^k}{-log(1-s)c}\right] \leq \frac{\beta}{\lambda} < \infty$$

for any $k > 0$, so the second condition of in Proposition 6 follows. Finally

$$\begin{aligned} g'(s) &= \mathbb{E}\left[-\frac{\beta(1-s)^{\frac{\alpha}{V}-1}}{c\log(1-s)} + \frac{\beta(\frac{\alpha}{V}-1)s(1-s)^{\frac{\alpha}{V}-2}}{c\log(1-s)} - \frac{\beta s(1-s)^{\frac{\alpha}{V}-2}}{c(\log(1-s))^2}\right] \\ &= \mathbb{E}\left[\frac{\beta(\frac{\alpha}{V}-1)s(1-s)^{\frac{\alpha}{V}-2}}{c\log(1-s)}\right] + \frac{\beta\mathbb{E}\left[(1-s)^{\frac{\alpha}{V}-1}\right]}{c}\left(\frac{-\log(1-s)-s(1-s)^{-1}}{(\log(1-s))^2}\right) \end{aligned}$$

Using L'Hopital's rule and the Dominated convergence Theorem we see that

$$\begin{aligned} \lim_{s\to 0} g'(s) &= -\frac{\beta(\mathbb{E}\left[\frac{\alpha}{V}\right]-1)}{c} + \frac{\beta}{c}\lim_{s\to 0}\frac{(1-s)^{-1}-(1-s)^{-1}-s(1-s)^{-2}}{-2\log(1-s)(1-s)^{-1}} \\ &= -\frac{\beta(\mathbb{E}\left[\frac{\alpha}{V}\right]-1)}{c} - \frac{\beta}{2c} < \infty. \end{aligned}$$

So we conclude the condition for the Bernstein-von Mises result. Proceeding as in the proof of Proposition 16, we observe that we have a series representation of the form

$$\sum_{i=1}^{\infty}(K_1J_{1,i} + \ldots + K_dJ_{d,i})w_i\delta_{u_i}$$

where $K_1, \ldots, K_d$ are as in the hypothesis of the proposition to proof and $\left(J_{1,i}, \ldots, J_{d,i}\right) \overset{\text{i.i.d.}}{\sim} \mathscr{L}(h)$ for any $i \in \{1,2,\ldots\}$. It follows that for $i \in \{1,2,\ldots\}$

$$\mathbb{E}\left[K_1 J_{1,i} + \ldots + K_d J_{d,i}\right] < \infty$$

and for some $j \in \{1,\ldots,d\}$

$$\mathbb{E}\left[1/\left(K_1 J_{1,i} + \ldots + K_d J_{d,i}\right)\right] < \mathbb{E}\left[1/(K_j J_{j,i})\right] < \infty.$$

So an application of Lemma 12 when conditioning on $\boldsymbol{K}$ concludes the proof.

Proof of Corollary 11

As the regressor functions f_i are strictly positive the result follows from the Proposition 18 in view of identity (5.2).

References

Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (2012). *Statistical models based on counting processes*. Springer Science & Business Media.

Caron, F. and Fox, E. B. (2017). Sparse graphs using exchangeable random measures. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(5):1295–1366.

Caron, F. and Rousseau, J. (2017). On sparsity and power-law properties of graphs based on exchangeable point processes. *arXiv preprint arXiv:1708.03120*.

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

Daley, D. J. and Vere-Jones, D. (2007). *An introduction to the theory of point processes: volume II: general theory and structure*. Springer Science & Business Media.

de Finetti, B. (1938). Sur la condition de quivalence partielle. *Act. Scient. Ind.*, pages 5–18.

de Finetti, B. (1980). On the condition of partial exchangeability. *Studies in inductive logic and probability*, 2:193–205.

Dey, J., Erickson, R., and Ramamoorthi, R. (2003). Some aspects of neutral to right priors. *International statistical review*, 71(2):383–401.

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

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

Esmaeili, H. and Klüppelberg, C. (2010). Parameter estimation of a bivariate compound Poisson process. *Insurance: mathematics and economics*, 47(2):224–233.

Ferguson, T. S. and Klass, M. J. (1972). A representation of independent increment processes without gaussian components. *The Annals of Mathematical Statistics*, 43.5:1634–1643.

Griffin, J. and Leisen, F. (2018). Modelling and computation using NCoRM mixtures for density regression. *Bayesian Analysis*, 13(3):897–916.

Griffin, J. E. and Leisen, F. (2017). Compound random measures and their use in bayesian non-parametrics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(2):525–545.

Grothe, O. and Nicklas, S. (2013). Vine constructions of Lévy copulas. *Journal of Multivariate Analysis*, 119:1–15.

Kalbfleisch, J. D. and Prentice, R. L. (2011). *The statistical analysis of failure time data*, volume 360. John Wiley & Sons.

Kallenberg, O. (2006). *Probabilistic symmetries and invariance principles*. Springer Science & Business Media.

Kallsen, J. and Tankov, P. (2006). Characterization of dependence of multidimensional Lévy processes using Lévy copulas. *Journal of Multivariate Analysis*, 97(7):1551–1572.

Kim, Y. and Lee, J. (2001). On posterior consistency of survival models. *Annals of Statistics*, pages 666–686.

Kim, Y. and Lee, J. (2003). Bayesian analysis of proportional hazard models. *The Annals of Statistics*, 31(2):493–511.

Kim, Y. and Lee, J. (2004). A Bernstein-von Mises theorem in the nonparametric right-censoring model. *The Annals of Statistics*, 32(4):1492–1512.

Kingman, J. (1967). Completely random measures. *Pacific Journal of Mathematics*, 21(1):59–78.

Kingman, J. (2005). Poisson processes. *Encyclopedia of biostatistics*, 6.

Klein, J. P. and Moeschberger, M. L. (2006). *Survival analysis: techniques for censored and truncated data*. Springer Science & Business Media.

Kyprianou, A. E. (2006). *Introductory lectures on fluctuations of Lévy processes with applications*. Springer Science & Business Media.

Leisen, F. and Lijoi, A. (2011). Vectors of two-parameter poisson–Dirichlet processes. *Journal of Multivariate Analysis*, 102(3):482–495.

Leisen, F., Lijoi, A., and Spanó, D. (2013). A vector of Dirichlet processes. *Electronic Journal of Statistics*, 7:62–90.

Nelsen, R. B. (2007). *An introduction to copulas*. Springer Science & Business Media.

Riva-Palacio, A. and Leisen, F. (2018). Bayesian nonparametric estimation of survival functions with multiple-samples information. *Electronic Journal of Statistics*, 12(1):1330–1357.

Robert, C. P. and Casella, G. (2010). *Introducing monte carlo methods with R*, volume 18. Springer.

Rosiński, J. (2001). Series representations of Lévy processes from the perspective of point processes. In *Lévy processes*, pages 401–415. Springer.

Ryzhik and Gradshteyn (1965). *Table of integrals, series, and products*. Academic press New York.

Sato, K.-i. (1999). *Lévy processes and infinitely divisible distributions*. Cambridge university press.

Todeschini, A., Miscouridou, X., and Caron, F. (2016). Exchangeable random measures for sparse and modular graphs with overlapping communities. *arXiv preprint arXiv:1602.02114*.

Wolpert, R. L. and Ickstadt, K. (1998). Simulation of Lévy random fields. In *Practical nonparametric and semiparametric Bayesian statistics*, pages 227–242. Springer.

Yan, J. (2010). KMsurv: Data sets from Klein and Moeschberger (1997), survival analysis.

Zhu, W. and Leisen, F. (2015). A multivariate extension of a vector of two-parameter Poisson–Dirichlet processes. *Journal of Nonparametric Statistics*, 27(1):89–105.