Collegio Carlo Alberto

Nonparametric priors for vectors of survival functions

Ilenia Epifani
Antonio Lijoi

No. 132
December 2009

Carlo Alberto Notebooks

www.carloalberto.org/working_papers

NONPARAMETRIC PRIORS FOR VECTORS OF SURVIVAL FUNCTIONS

Ilenia Epifani and Antonio Lijoi

Politecnico di Milano and Università di Pavia

Abstract: The paper proposes a new nonparametric prior for two–dimensional vectors of survival functions (S_1, S_2). The definition we introduce is based on the notion of Lévy copula and it will be used to model, in a nonparametric Bayesian framework, two–sample survival data. Such an application will yield a natural extension of the more familiar neutral to the right process of Doksum (1974) adopted for drawing inferences on single survival functions. We, then, obtain a description of the posterior distribution of (S_1, S_2), conditionally on possibly right–censored data. As a by–product of our analysis, we find out that the marginal distribution of a pair of observations from the two samples coincides with the Marshall–Olkin or the Weibull distribution according to specific choices of the marginal Lévy measures.

Key words and phrases: Bayesian nonparametrics, Completely random measures, Dependent stable processes, Lévy copulas, Posterior distribution, Right–censored data, Survival function

1. Introduction

A typical approach to the definition of nonparametric priors is based on the use of completely random measures, namely random measures inducing independent random variables when evaluated on pairwise disjoint measurable sets. The Dirichlet process introduced by Ferguson (1974) is a noteworthy example being generated, in distribution, by the normalization of a gamma random measure. Other well–known examples appear in the survival analysis literature. In Doksum (1974), a prior for the survival function is defined by

$$S(t|\mu) = \mathbb{P}\,[Y > t \mid \mu] = \exp\{-\mu(0,t]\} \qquad \forall t \geq 0 \tag{1.1}$$

where μ is a completely random measure defined on some probability space $(\Omega, \mathscr{F}, \mathbb{P})$ such that $\mathbb{P}[\lim_{t\to\infty} \mu((0,t]) = \infty] = 1$. As shown in Doksum (1974) Equation (1.1) defines a *neutral to the right* (NTR) prior, namely a random probability measure such that the random variables

$$1 - S(t_1|\mu),\ 1 - \frac{S(t_2|\mu)}{S(t_1|\mu)},\ \dots\ ,\ 1 - \frac{S(t_n|\mu)}{S(t_{n-1}|\mu)}$$

are mutually independent for any choice of $0 < t_1 < t_2 < \cdots < t_n < \infty$. When referring to model (1.1) for a survival time Y, we will henceforth use the notation $Y|\mu \sim \mathrm{NTR}(\mu)$. According to an alternative approach established by Hjort (1990), a beta completely random measure is used to define a prior for the cumulative hazard function

$$\Lambda(t|\mu) = \int_0^t \mathbb{P}[s \le Y \le s + \mathrm{d}s | Y \ge s,\ \mu] = \mu(0,t]. \tag{1.2}$$

These two constructions are equivalent. As shown in Hjort (1990), a prior for the survival function is NTR if and only if its corresponding cumulative hazard is a completely random measure. Moreover, if $Y_1, \ldots, Y_n$ are the first n elements of a sequence of exchangeable survival times, one can explicitly evaluate the posterior distribution of the survival function and of the cumulative hazard as defined in (1.1) and in (1.2). The former can be found in Ferguson (1974) and in Ferguson and Phadia (1979) and the latter was achieved by Hjort (1990).

In the present paper we look for bivariate extensions of the previous definitions. We wish to introduce priors for vectors of dependent survival (S_1, S_2) or cumulative hazard (Λ_1, Λ_2) functions. This will be accomplished by resorting to vectors of completely random measures (μ_1, μ_2), with fixed margins, such that μ_i gives rise to a univariate NTR prior. The dependence between μ_1 and μ_2 will be devised in such a way that the vector measure (μ_1, μ_2) is completely random, that is for any pair of disjoint measurable sets A and B the vectors $(\mu_1(A), \mu_2(A))$ and $(\mu_1(B), \mu_2(B))$ are independent. An appropriate tool to achieve this goal is represented by Lévy copulas. See Tankov (2003), Cont and Tankov (2004) and Kallsen and Tankov (2006). A typical application where this model is useful concerns survival, or failure, times related to statistical units drawn from two separate groups such as, e.g., in the analysis of time-to-response outcomes in group-randomized intervention trials. Suppose, for example, that statistical units are patients suffering from a certain illness and they are split into two groups according to the treatment they are subject to. Let $Y_1^{(1)}, \ldots, Y_{n_1}^{(1)}$ and $Y_1^{(2)}, \ldots, Y_{n_2}^{(2)}$ be the survival times related to n_1 and n_2 units drawn from the first and the second group, respectively. Then, one can assume that

$$S(u,v) = \mathbb{P}\left[Y_i^{(1)} > u,\ Y_j^{(2)} > v \,\middle|\, (\mu_1, \mu_2) \right] = \exp\{-\mu_1(0,u] - \mu_2(0,v]\} \tag{1.3}$$

$$\mathbb{P}\left[Y_1^{(i)} > t_1, \ldots, Y_n^{(i)} > t_n \,\middle|\, (\mu_1, \mu_2) \right] = \prod_{j=1}^{n} \exp\{-\mu_i(0,t_j]\} \qquad i = 1,2 \tag{1.4}$$

for any u, v, $t_1, \ldots, t_n$ positive. According to (1.3) and (1.4), we assume exchangeability in each group and this seems natural since patients sharing the same treatment can be thought of as homogeneous. On the other hand, given the marginal random survival functions, the lifetimes, or times-to-event, are independent among the two groups. This is similar to frailty models where, conditional on the frailty, the two survival times are independent. The dependence among the data, which is reasonable since people from the two groups share the same kind of illness, is induced indirectly by the dependence between the two marginal survival functions. It will be seen that this approach has some interesting advantages: (i) it leads to a representation of the posterior distribution of (S_1, S_2), or of (Λ_1, Λ_2), which is an extension of the univariate case; (ii) the resulting representation of the Laplace functional of the bivariate process suggests the definition of a new measure of dependence between survival functions; (iii) for appropriate choices of μ_1 and μ_2, the marginal distribution of $(Y^{(1)}, Y^{(2)})$ coincides with some well–known bivariate survival functions such as the Marshall–Olkin and the Weibull distributions. Recently, Ishwaran and Zarepour (2008) have faced a similar issue and provide a definition of vectors of completely random measures based on series representations which are named bivariate G–measures.

Even beyond applications to survival analysis, our results connect to a very active area of research in Bayesian nonparametric statistics. Indeed, exchangeable models commonly used in Bayesian inference are not well suited for dealing with regression problems and a lot of effort has been recently put to define new priors which incorporate covariates information. These are referred to as *dependent processes*, the most popular example being the dependent Dirichlet process introduced in a few pioneering contributions by MacEachern (1999, 2000, 2001). Later developments on dependent Dirichlet processes can be found in De Iorio, Müller, Rosner and MacEachern (2004), Griffin and Steel (2006), Rodriguez, Dunson and Gelfand (2008), Dunson, Xue and Carin (2008) and Dunson and Park (2008). The idea, in these papers, is to construct a family $\{\tilde{P}_z : z \in Z\}$ of random probability measures indexed by a covariate (or vector of covariates) z taking values in some set Z. Hence, one defines $\tilde{P}_z$ as a discrete random probability measure $\sum_i \pi_i(z)\, \delta_{X_i(z)}$ with both random masses π_i and atoms X_i depending on the z values, and the π_i's determined through a stick–breaking procedure. The nonparametric prior we propose here can be seen as a dependent process with Z consisting of two points $\{z_1, z_2\}$: the dependence structure between $\tilde{P}_{z_1}$ and $\tilde{P}_{z_2}$ is determined by a Lévy copula. The main advantage of our model is the possibility of deriving closed form expressions for Bayesian estimators which, at least to our knowledge, cannot be achieved when resorting to dependent stick–breaking processes. Another prior which fits into this framework is the bivariate Dirichlet process defined in Walker and Muliere (2003).

The structure of the paper is as follows. In Section 2 we recall some elementary facts concerning completely random measures. In Section 3 we describe the notion of Lévy copula. Section 4 illustrates the new prior we introduce and some relevant properties it features. In Section 5 a description of the posterior distribution is provided. Section 6 concisely describes the connection of our work with the analysis of cumulative hazards. Section 7 illustrates an application with a dataset of right-censored samples. Finally, Section 8 contains some concluding remarks. All proofs are deferred to the Appendix.

2. Some preliminaries

In this section we briefly recall the notion of completely random measure (CRM) which is the main ingredient in the definition of various commonly used priors in Bayesian nonparametrics. A *completely random measure* μ on a complete and separable metric space $\mathbb{X}$ is a measurable function defined on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and taking values in the space of all measures on $\mathbb{X}$ such that for any choice of sets $A_1, \ldots, A_n$ in the σ–field $\mathscr{X}$ of Borel subsets of $\mathbb{X}$ such that $A_i \cap A_j = \varnothing$ for any $i \neq j$, the random variables $\mu(A_1), \ldots, \mu(A_n)$ are mutually independent. It is well–known that $\mu = \mu_c + \sum_{i=1}^q J_i\, \delta_{x_i}$ where μ_c is a CRM such that, for some measure $\tilde{\nu}$ on $\mathbb{X} \times \mathbb{R}^+$,

$$\mathrm{E}\left[\mathrm{e}^{-\lambda\, \mu_c(A)}\right] = \mathrm{e}^{-\int_{A\times\mathbb{R}^+}\left(1-\mathrm{e}^{-\lambda x}\right)\tilde{\nu}(\mathrm{d}s,\mathrm{d}x)} \qquad \forall A \in \mathscr{X} \quad \forall \lambda > 0, \tag{2.1}$$

$x_1, \ldots, x_q$ are fixed points of discontinuity in $\mathbb{X}$ and the jumps $J_1, \ldots, J_q$ are independent and non–negative random variables being also independent from μ_c. With no loss of generality we can omit the consideration of the fixed jump points and in the sequel suppose $\mu = \mu_c$. The measure $\tilde{\nu}$ in (2.1) takes on the name of *Lévy measure*. See Kingman (1993) for an elegant and deep account on CRMs. As anticipated in the previous section, when $\mathbb{X} = \mathbb{R}^+$ a NTR process is defined as a random probability measure whose distribution function $\{F(t) : t \geq 0\}$ has the same distribution as $\{1 - \mathrm{e}^{-\mu(0,t]} : t \geq 0\}$.

If we wish to make use of (1.3) and (1.4), it would be desirable that the probability distribution of (μ_1, μ_2) is characterized by

$$\mathrm{E}\left[\mathrm{e}^{-\lambda_1\, \mu_1(0,t] - \lambda_2\, \mu_2(0,t]}\right] = \mathrm{e}^{-\int_{(0,t]\times(\mathbb{R}^+)^2}\left[1-\mathrm{e}^{-\lambda_1 x_1 - \lambda_2 x_2}\right]\tilde{\nu}(\mathrm{d}s,\mathrm{d}x_1,\mathrm{d}x_2)}$$

for any $t \geq 0$ and $\lambda_1, \lambda_2 > 0$. Hence the vector (μ_1, μ_2) has independent increments and the measure $\tilde{\nu}$ is the associated Lévy measure. Given its importance in later discussion, for the sake of simplicity in notation we let

$$\psi_t(\lambda_1, \lambda_2) := \int_{(0,t]\times(\mathbb{R}^+)^2} \left[1 - \mathrm{e}^{-\lambda_1 x_1 - \lambda_2 x_2}\right] \tilde{\nu}(\mathrm{d}s, \mathrm{d}x_1, \mathrm{d}x_2) \qquad \forall \lambda_1, \lambda_2 > 0 \tag{2.2}$$

denote the Laplace exponent of the (vector) random measure (μ_1, μ_2). Introduce the function $h_{t_1,t_2}(\lambda_1, \lambda_2) = \psi_{t_1 \wedge t_2}(\lambda_1, \lambda_2) - \psi_{t_1\wedge t_2}(\lambda_1, 0) - \psi_{t_1 \wedge t_2}(0, \lambda_2)$, with $a \wedge b := \min\{a, b\}$ for any $a, b \in \mathbb{R}$. Note that using the independence of the increments one has, for any $t_1 > 0, t_2 > 0$,

$$\mathrm{E}\left[\mathrm{e}^{-\lambda_1 \mu_1(0,t_1] - \lambda_2 \mu_2(0,t_2]}\right] = \mathrm{e}^{-\psi_{1,t_1}(\lambda_1) - \psi_{2,t_2}(\lambda_2) - h_{t_1,t_2}(\lambda_1,\lambda_2)} \tag{2.3}$$

where

$$\psi_{i,t}(\lambda) := \int_{(0,t]\times\mathbb{R}^+} [1 - \mathrm{e}^{-\lambda x}]\, \tilde{\nu}_i(\mathrm{d}s, \mathrm{d}x) = \int_{\mathbb{R}^+} [1 - \mathrm{e}^{-\lambda x}]\, \tilde{\nu}_{i,t}(\mathrm{d}x)$$

and $\tilde{\nu}_i$ is the (marginal) Lévy measure of μ_i and $\tilde{\nu}_{i,t}(\mathrm{d}x) := \tilde{\nu}_i((0,t] \times \mathrm{d}x)$, for $i \in \{1, 2\}$. Note that the marginal Lévy measures $\tilde{\nu}_1$ and $\tilde{\nu}_2$ can be deduced from ν since, for example,

$\tilde{\nu}_1(\mathrm{d}s, \mathrm{d}x) = \tilde{\nu}(\mathrm{d}s \times \mathrm{d}x \times \mathbb{R}^+)$. Consequently, one has $\psi_{1,t}(\lambda) = \psi_t(\lambda, 0)$ and $\psi_{2,t} = \psi_t(0, \lambda)$. It is further assumed that

$$\tilde{\nu}_t(\mathrm{d}x_1, \mathrm{d}x_2) := \tilde{\nu}((0,t] \times \mathrm{d}x_1 \times \mathrm{d}x_2) = \gamma(t)\, \nu(x_1, x_2)\, \mathrm{d}x_1\, \mathrm{d}x_2 \tag{2.4}$$

for some increasing and non–negative function $\gamma : \mathbb{R}^+ \to \mathbb{R}^+$ such that $\lim_{t\to\infty} \gamma(t) = \infty$: in this case we say that the vector measure (μ_1, μ_2) is *homogeneous* according to the terminology in Ferguson and Phadia (1979) and for simplicity we refer to ν in (2.4) as the corresponding bivariate Lévy density. It is immediate to check that, in this case, $\psi_t = \gamma(t)\, \psi$. Whenever $\tilde{\nu}_t$ is not representable as in (2.4), i.e. it cannot be expressed as a product of a factor depending only on t and another depending only on (x_1, x_2), we say that (μ_1, μ_2) is *non–homogeneous*. Finally, in the sequel we write $(\mu_1, \mu_2) \sim \mathscr{M}_2(\nu; \gamma)$ to denote a homogeneous vector of completely random measures characterized by (2.3) with Lévy intensity representable as in (2.4).

3. Lévy copulae

The notion of Lévy copula parallels the concept of distribution copulas and enables one to define a vector of completely random measures (μ_1, μ_2) on $(\mathbb{R}^+)^2$ starting from two marginal CRMs μ_1 and μ_2 with respective Lévy intensities $\{\tilde{\nu}_{1,t} : t \geq 0\}$ and $\{\tilde{\nu}_{2,t} : t \geq 0\}$. Here below we explicitly consider the case where the Lévy measure can be represented as follows

$$\tilde{\nu}_{i,t}(\mathrm{d}x) = \gamma(t)\, \nu_i(x)\, \mathrm{d}x \qquad i = 1, 2 \tag{3.1}$$

for any $t \geq 0$, where $t \mapsto \gamma(t)$ is a non–negative, increasing and differentiable function on $[0, \infty)$ such that $\lim_{t\to\infty} \gamma(t) = \infty$ and $\gamma(0) \equiv 0$. The function $\nu_i : \mathbb{R}^+ \to \mathbb{R}^+$ takes on the name of Lévy density and it is such that $\int_0^\infty (x \wedge 1)\, \nu_i(x)\, \mathrm{d}x < \infty$. Correspondingly one has $\psi_{i,t} = \gamma(t)\, \psi_i$ where $\psi_i(\lambda) = \int_0^\infty [1 - \mathrm{e}^{-\lambda x}] \nu_i(x)\, \mathrm{d}x$ for $i = 1, 2$. Moreover, the function

$$x \mapsto U_i(x) = \int_x^\infty \nu_i(s)\, \mathrm{d}s$$

defines the tail–integral corresponding to ν_i, $i \in \{1, 2\}$, which is continuous and monotone decreasing on $\mathbb{R}^+$. If the bivariate Lévy density ν, as displayed in (2.4), is such that $\int_0^\infty \nu(x_1, x_2) \mathrm{d}x_i = \nu_j(x_j)$, for any $i \in \{1, 2\}$ and $j \neq i$, then ν is the Lévy density of the bivariate random measure (μ_1, μ_2). The problem we now face consists in applying a procedure which enables to establish ν, given the marginals ν_1 and ν_2 have been assigned. In order to do so, we use the notion of Lévy copula, recently introduced by Tankov (2003) for Lévy processes with positive jumps and later extended in Kallsen and Tankov (2006) to encompass Lévy processes with jumps of any sign. A full and exhaustive account on Lévy copulas, with applications to financial modelling, can be found in Cont and Tankov (2004). Let us first recall the definition of Lévy copula.

DEFINITION 1. A *positive Lévy copula* is a function $C : [0,\infty]^2 \to [0,\infty]$ such that

(i) $C(x_1, 0) = C(0, x_2) = 0$;
(ii) for all $x_1 < y_1$ and $x_2 < y_2$ $C(x_1, x_2) + C(y_1, y_2) - C(x_1, y_2) - C(y_1, x_2) \geq 0$;
(iii) C has uniform margins, *i.e.* $C(x_1, \infty) = x_1$ and $C(\infty, x_2) = x_2$.

There are some examples of Lévy copulas whose form is reminiscent of copulas for distributions. As a first case, consider a vector (μ_1, μ_2) of CRMs with μ_1 and μ_2 independent. By virtue of Proposition 5.3 in Cont and Tankov (2004) one has $\nu(A) = \nu_1(A_1) + \nu_2(A_2)$ where $A_1 = \{x_1 : (x_1, 0) \in A\}$ and $A_2 = \{x_2 : (0, x_2) \in A\}$. The corresponding copula turns out to be $C_\perp(x_1, x_2) = x_1 \mathbb{1}_{x_2=\infty} + x_2 \mathbb{1}_{x_1=\infty}$. This is the independence copula. The case of complete dependence arises when, for any positive s and t, one has either $\mu_i(0, s] - \mu_i(0, s-] < \mu_i(0, t] - \mu_i(0, t-]$, for any $i = 1, 2$, or $\mu_i(0, s] - \mu_i(0, s-] > \mu_i(0, t] - \mu_i(0, t-]$, for any $i = 1, 2$. A copula yielding a completely dependent bivariate process with independent increments is $C_\parallel(x_1, x_2) = x_1 \wedge x_2$. Apart from these two extreme cases, there are other forms of copulas which capture intermediate cases of dependence. An example is the *Clayton copula* defined by

$$C_\theta(x_1, x_2) = \left\{ x_1^{-\theta} + x_2^{-\theta} \right\}^{-\frac{1}{\theta}}, \qquad \theta > 0 \tag{3.2}$$

where, as we shall see, the parameter θ regulates the degree of dependence between μ_1 and μ_2.

When the copula C and the tail integrals are sufficiently smooth the bivariate Lévy density ν, with fixed marginals ν_1 and ν_2, can be recovered from

$$\nu(x_1, x_2) = \frac{\partial^2}{\partial u\, \partial v} C(u, v) \bigg|_{u=U_1(x_1),\, v=U_2(x_2)} \nu_1(x_1)\, \nu_2(x_2). \tag{3.3}$$

Combining (3.3) with the Clayton copula C_θ in (3.2) one can show that the following holds true.

PROPOSITION 1. *Let ν_1 and ν_2 be two univariate Lévy densities such that, if $\nu(\,\cdot\,, \cdot\,; \theta)$ is obtained from* (3.3) *with $C = C_\theta$ given in* (3.2)*, one has that the integrability condition $\int_{\|\boldsymbol{x}\| \leq 1} \|\boldsymbol{x}\|\, \nu(x_1, x_2; \theta)\, \mathrm{d}x_1\, \mathrm{d}x_2 < \infty$ holds true. Then*

$$\begin{aligned} \psi(\lambda_1, \lambda_2; \theta) &= \int_{(\mathbb{R}^+)^2} [1 - \mathrm{e}^{-\lambda_1 x_1 - \lambda_2 x_2}]\, \nu(x_1, x_2; \theta)\, \mathrm{d}x_1 \mathrm{d}x_2 \\ &= \psi_\perp(\lambda_1, \lambda_2) - \lambda_1 \lambda_2 \int_{(\mathbb{R}^+)^2} \mathrm{e}^{-\lambda_1 x_1 - \lambda_2 x_2}\, C_\theta(U_1(x_1), U_2(x_2))\, \mathrm{d}x_1\, \mathrm{d}x_2 \end{aligned} \tag{3.4}$$

where $\psi_\perp(\lambda_1, \lambda_2) = \psi_1(\lambda_1) + \psi_2(\lambda_2)$ is the Laplace exponent corresponding to the independence case.

According to (3.4) in Proposition 1, the term responsible of the dependence is given by the quantity

$$\kappa(\theta;\lambda_1,\lambda_2) := \lambda_1\lambda_2 \int_{(\mathbb{R}^+)^2} \mathrm{e}^{-\lambda_1 x_1-\lambda_2 x_2}\, C_\theta(U_1(x_1),U_2(x_2))\,\mathrm{d}x_1\,\mathrm{d}x_2$$

and this will be used to introduce a novel measure of association between μ_1 and μ_2. In Tankov (2003) it is shown that, as $\theta \to 0$, one approaches the situation of independence, $\nu(x_1,x_2) = \nu_1(x_1)\delta_{\{0\}}(x_2) + \nu_2(x_2)\delta_{\{0\}}(x_1)$ and the corresponding Laplace exponent reduces to $\psi(\lambda_1,\lambda_2) = \psi_\perp(\lambda_1,\lambda_2)$. On the other hand, as $\theta \to \infty$, the limiting two–dimensional Lévy measure is concentrated on the set $\{(x_1,x_2) : U_1(x_1) = U_2(x_2)\}$. In this case the limiting Lévy measure does not have a density with respect to the Lebesgue measure on $\mathbb{R}^2$, but it is still of finite variation. See Section A2 in Appendix for a proof of this fact. The structure achieved through this limiting process is that of complete dependence. When the two marginals coincide, *i.e.* $\psi_1 = \psi_2 = \psi^*$, the Laplace exponent with complete dependence coincides with $\psi(\lambda_1,\lambda_2) = \psi^*(\lambda_1) + \psi^*(\lambda_2) - \psi^*(\lambda_1+\lambda_2)$.

Many common measures of association depend monotonically on θ through the function $\kappa(\theta) := \kappa(\theta;1,1)$. This will become apparent in the next section. Here we confine ourselves to pointing out a few properties of the function $\kappa(\theta)$.

Proposition 2. *Let ν_1 and ν_2 be two Lévy densities such that if ν is obtained from* (3.3) *with $C = C_\theta$, one has $\int_{\|\boldsymbol{x}\|\le 1} \|\boldsymbol{x}\|\,\nu(x_1,x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 < \infty$. Then*

(i) $\lim_{\theta\to 0}\kappa(\theta) = 0$;

(ii) $\lim_{\theta\to\infty}\kappa(\theta) = \int_{(\mathbb{R}^+)^2} \mathrm{e}^{-x_1-x_2}\,\min\{U_1(x_1),U_2(x_2)\}\,\mathrm{d}x_1\mathrm{d}x_2$;

(iii) $\theta \mapsto \kappa(\theta)$ *is a non decreasing function.*

One can thus note that, setting $\kappa(\infty) := \lim_{\theta\to\infty}\kappa(\theta)$, then

$$\bar\kappa(\theta) = \frac{\kappa(\theta)}{\kappa(\infty)} \in (0,1). \tag{3.5}$$

Values of $\bar\kappa(\theta)$ close to 0 suggest a weak dependence between the two CRMs μ_1 and μ_2. On the other hand, values of $\bar\kappa(\theta)$ close to 1 provide indication of the presence of a strong dependence among the jumps of the underlying random measures. If μ_1 and μ_2 are used to define NTR priors according to (1.3) and (1.4), the dependence between survival functions can be measured through $\bar\kappa$ and it does not depend on the point t at which the survival functions S_1 and S_2 can be evaluated: this is a straightforward consequence of the homogeneity of (μ_1,μ_2).

4. Priors for dependent survival functions

The model we are going to consider can be described as follows. Suppose there are two distinct groups of individuals or statistical units and denote with $Y^{(1)}$ and $Y^{(2)}$ the survival time for any individual in the first group and in the second group, respectively.

It is further assumed that

$$
\begin{aligned}
Y_j^{(i)} \,|\, (\mu_1, \mu_2) &\overset{\text{ind}}{\sim} \mathrm{NTR}(\mu_i) \qquad i = 1, 2 \\
(\mu_1, \mu_2) &\sim \mathscr{M}_2(\nu; \gamma)
\end{aligned} \tag{4.1}
$$

Hence, each sequence $(Y_j^{(i)})_{j\ge 1}$ is exchangeable and governed by a $\mathrm{NTR}(\mu_i)$ prior, with $i \in \{1, 2\}$. Given (μ_1, μ_2), any two observations $Y_j^{(1)}$ and $Y_l^{(2)}$ are independent. Nonetheless, they are marginally dependent in the sense that dependence arises when integrating out the vector (μ_1, μ_2) and it is generated via a Lévy copula. It is worth noting that, by virtue of Proposition 3 in Dey, Erickson and Ramamoorthi (2003), if each marginal Lévy measure in (3.1) is such that $\gamma(t) > 0$ for any $t > 0$ and ν_i is supported by $\mathbb{R}^+$, then the support of $t \mapsto S_i(t) = 1 - \exp\{-\mu_i(0, t]\}$, with respect to the topology of weak convergence, coincides with the whole space $\mathcal{S}$ of survival functions on $\mathbb{R}^+$. Hence, the support of the vector (S_1, S_2), with respect to the usual product topology, coincides with the space $\mathcal{S}^2$ of bivariate vectors of survival functions.

An interesting consequence of the proposed model concerns the form of such a marginal distribution for the vector of survival times $(Y^{(1)}, Y^{(2)})$. Indeed one obtains an expression which encompasses some well–known bivariate distributions used in survival analysis such as the Marshall–Olkin and the Weibull model.

Proposition 3. *Suppose the vector $Y^{(1)}$ and $Y^{(2)}$ are survival times modeled as in (4.1). Then*

$$
\mathbb{P}\left[Y^{(1)} > s,\, Y^{(2)} > t\right] = \exp\{-\gamma(s)\,\xi_1 - \gamma(t)\,\xi_2 - \gamma(s \vee t)\xi_{1,2}\} \tag{4.2}
$$

where $a \vee b = \max\{a, b\}$, $\xi_1 = \psi(1, 1) - \psi(0, 1) > 0$, $\xi_2 = \psi(1, 1) - \psi(1, 0) > 0$ and $\xi_{1,2} = \psi(1, 0) + \psi(0, 1) - \psi(1, 1) > 0$.

Note that the expression on the right of (4.2) is a typical representation for a bivariate survival distribution $\mathbb{P}[Y^{(1)} > s, Y^{(2)} > t]$: in fact $\gamma(s)\xi_1$ and $\gamma(t)\,\xi_2$ have the meaning of marginal cumulative hazard functions, whereas $\gamma(s \vee t)\xi_{1,2}$ defines the association structure. If $\gamma(t) \equiv t$, then the above survival function reduces to the Marshall–Olkin model. When $\gamma(t) \equiv t^\alpha$, one obtains a bivariate Weibull distribution. When we exploit the Clayton copula, in (4.2) one has

$$
\xi_{1,2} = \int_{(\mathbb{R}^+)^2} \mathrm{e}^{-x_1 - x_2}\, C_\theta(U_1(x_1), U_2(x_2))\, \mathrm{d}x_1 \mathrm{d}x_2 = \kappa(\theta).
$$

The random probability distribution arising from the specification in (4.1) can also be described in terms of random partitions in the same spirit of the characterization of the univariate NTR priors given in Doksum (1974).

Proposition 4. *Let F be a bivariate random distribution function on $(\mathbb{R}^+)^2$ and $\mu_{i,t} =$*

$\mu_i(0,t]$, for $i \in \{1,2\}$ and $t > 0$. Then $F(s,t)$ has the same distribution as $\{1 - \mathrm{e}^{-\mu_{1,s}}\}\{1-\mathrm{e}^{-\mu_{2,t}}\}$, for some bivariate completely random measure (μ_1,μ_2), if and only if for any choice of $k \geq 1$ and $0 < t_1 < \cdots < t_k$ there exist k independent random vectors $(V_{1,1},V_{2,1}),\ldots,(V_{1,k},V_{2,k})$ such that

$$\begin{aligned}(4.3)\quad \Big(F(t_1,t_1),F(t_2,t_2),\ldots,F(t_k,t_k)\Big) &\stackrel{d}{=} \\ &\stackrel{d}{=} \Big(V_{1,1}V_{2,1},[1-\bar V_{1,1}\bar V_{1,2}][1-\bar V_{2,1}\bar V_{2,2}],\ldots,[1-\prod_{j=1}^{k}\bar V_{1,j}][1-\prod_{j=1}^{k}\bar V_{2,j}]\Big)\end{aligned}$$

where $\bar V_{i,j} = 1 - V_{i,j}$ for any i and j.

One can use $\bar k(\theta)$ as a measure of dependence between μ_1 and μ_2, *i.e.* between the two random marginal survival functions. The statistical meaning of the association measure $\bar\kappa(\theta)$ becomes apparent if we compare it with the traditional correlation $\rho_\theta(t)$ between the marginal NTR survival functions $S_1(t) = \mathbb{P}\left[Y^{(1)} > t \,|\, \mu_1\right]$ and $S_2(t) = \mathbb{P}\left[Y^{(2)} > t \,|\, \mu_2\right]$. Indeed, one can show that the following holds true

Proposition 5. *Let $\kappa_i := \int_0^\infty (1-\mathrm{e}^{-x})^2 \nu_i(x)\,\mathrm{d}x$, for each $i \in \{1,2\}$. Then*

$$\rho_\theta(t) = \frac{\mathrm{e}^{\gamma(t)\kappa(\theta)} - 1}{\sqrt{[\mathrm{e}^{\gamma(t)\kappa_1}-1]\,[\mathrm{e}^{\gamma(t)\kappa_2}-1]}} \tag{4.4}$$

for any $t > 0$ and $\theta > 0$. Moreover, if $\nu_1 = \nu_2 = \nu^*$ then $\kappa(\infty) = \int_0^\infty (1-\mathrm{e}^{-x})^2 \nu^*(x)\,\mathrm{d}x$ and $\rho_\theta(t) < \bar\kappa(\theta)$ for any $t > 0$ and $\theta > 0$.

Hence, when the two marginals coincide, $\bar\kappa(\theta)$ is an upper bound for $\rho_\theta(t)$, for any $t > 0$. The merit of resorting to the approach of Lévy copulas, with the Clayton family $\{C_\theta : \theta > 0\}$, is that it enables one to specify and compare situations of complete dependence with the actual structure of dependence between the marginal random survival functions.

Turning attention to the concordance between survival times $Y^{(1)}$ and $Y^{(2)}$ from the two samples, one can prove the following interesting fact.

Proposition 6. *If $\rho_\theta(Y^{(1)},Y^{(2)})$ is the correlation coefficient between survival times $Y^{(1)}$ and $Y^{(2)}$ one has that for any $\theta > 0$*

$$\rho_\theta(Y^{(1)},Y^{(2)}) = \frac{\int_0^\infty\int_0^\infty \mathrm{e}^{-\gamma(t)\,\psi_1(1)-\gamma(s)\,\psi_2(1)}\ \left\{\mathrm{e}^{\gamma(s\wedge t)\,\kappa(\theta)}-1\right\}\,\mathrm{d}s\,\mathrm{d}t}{\prod_{i=1}^2\sqrt{2\int_0^\infty t\,\mathrm{e}^{-\gamma(t)\psi_i(1)}\,\mathrm{d}t - \left(\int_0^\infty \mathrm{e}^{-\gamma(t)\psi_i(1)}\,\mathrm{d}t\right)^2}} \tag{4.5}$$

where we recall that $\psi_i(\lambda) = \int_0^\infty [1-\mathrm{e}^{-\lambda x}]\,\nu_i(x)\,\mathrm{d}x$ for any $i \in \{1,2\}$.

In the special case where $\gamma(t) \equiv t$, it is immediate to deduce from (4.5) that $\rho_\theta(Y^{(1)},Y^{(2)}) = \kappa(\theta)/[\psi_1(1)+\psi_2(1)-\kappa(\theta)]$ for any $\theta > 0$. Hence, one can express the correlation between $Y^{(1)}$ and $Y^{(2)}$ in terms of the quantity $\kappa(\theta)$ which contributes to measuring the dependence between the random measures μ_1 and μ_2. Moreover, as expected, $\theta \mapsto \rho_\theta(Y^{(1)}, Y^{(2)})$ is an increasing function.

We close the present section with an example of prior for nonparametric inference that will also be employed in the illustrative section.

Example 1. (*Stable processes*). Let μ_1 and μ_2 be α_1–stable and α_2–stable random measures, respectively. This means that μ_i is characterized by the Lévy density $\nu_i(x) = Ax^{-1-\alpha_i}/\Gamma(1-\alpha_i)$, where α_i is a parameter in $(0,1)$, for $i \in \{1,2\}$, and $A > 0$ is a constant. The i–th tail integral is $U_i(x) = Ax^{-\alpha_i}/[\alpha_i\Gamma(1-\alpha_i)]$ for any $x > 0$. Using the copula C_θ described in (3.2), one can determine the following two–dimensional Lévy density on $\mathbb{R}^+ \times \mathbb{R}^+$:

$$\nu(x_1, x_2; \theta) = A(1+\theta)(\alpha_1\alpha_2)^{\theta+1}(\Gamma(1-\alpha_1)\Gamma(1-\alpha_2))^\theta \times \\ \times \frac{x_1^{\alpha_1\theta-1}x_2^{\alpha_2\theta-1}}{\left\{\alpha_1^\theta\Gamma^\theta(1-\alpha_1)x_1^{\alpha_1\theta} + \alpha_2^\theta\Gamma^\theta(1-\alpha_2)x_2^{\alpha_2\theta}\right\}^{1/\theta+2}} \,. \tag{4.6}$$

If the two marginal Lévy densities coincide, *i.e.* $\alpha_1 = \alpha_2 = \alpha$, (4.6) reduces to

$$\nu(x_1, x_2; \theta) = \frac{A(1+\theta)\alpha}{\Gamma(1-\alpha)} \times \frac{(x_1x_2)^{\alpha\theta-1}}{\left\{x_1^{\alpha\theta} + x_2^{\alpha\theta}\right\}^{1/\theta+2}} \,. \tag{4.7}$$

According to the discussion above, the correspondence between the triplet (ν_1, ν_2, C_θ) and ν is one–to–one. It is easy to find that the two–dimensional Lévy density on $(\mathbb{R}^+)^2$ given in (4.6) is of finite variation. Indeed, using the polar coordinates transformation, the integral $\int_{\|\boldsymbol{x}\|\le 1} \|\boldsymbol{x}\|\, \nu(x_1, x_2)\, \mathrm{d}x_1 \mathrm{d}x_2$ is proportional to

$$\int_0^\infty \mathrm{d}\rho \int_0^{\pi/2} \mathrm{d}u\, \frac{\rho^{\alpha_1\theta+\alpha_2\theta}\cos(u)^{\alpha_1\theta-1}\sin(u)^{\alpha_2\theta-1}}{\left[\alpha_1^\theta\Gamma^\theta(1-\alpha_1)(\rho\cos(u))^{\alpha_1\theta} + \alpha_2^\theta\Gamma^\theta(1-\alpha_2)(\rho\sin(u))^{\alpha_2\theta}\right]^{1/\theta+2}}$$

which is finite for any $\theta > 0$. As for the Laplace exponent corresponding to ν in (4.6), one finds out that $\psi(\lambda_1, \lambda_2; \theta)/A$ coincides with

$$\frac{\lambda_1^{\alpha_1}}{\alpha_1} + \frac{\lambda_2^{\alpha_2}}{\alpha_2} - \lambda_1\lambda_2 \int_{(\mathbb{R}^+)^2} \frac{\mathrm{e}^{-\lambda_1x_1-\lambda_2x_2}}{\left(\alpha_1^\theta\Gamma^\theta(1-\alpha_1)x_1^{\alpha_1\theta} + \alpha_2^\theta\Gamma^\theta(1-\alpha_2)^\theta x_2^{\alpha_2\theta}\right)^{1/\theta}}\, \mathrm{d}x_1\, \mathrm{d}x_2 \,.$$

Hence, in this case

$$\kappa(\theta) = A \int_{(\mathbb{R}^+)^2} \mathrm{e}^{-x_1-x_2} \left(\alpha_1^\theta\Gamma^\theta(1-\alpha_1)x_1^{\alpha_1\theta} + \alpha_2^\theta\Gamma^\theta(1-\alpha_2)^\theta x_2^{\alpha_2\theta}\right)^{-1/\theta} \mathrm{d}x_1\, \mathrm{d}x_2$$

and this expression can only be evaluated numerically or via some suitable simulation scheme. As for the Laplace exponent $\psi(\lambda_1, \lambda_2; \theta)$, letting $\theta \to \infty$ one finds that $\psi(\lambda_1, \lambda_2; \infty)/A$

coincides with

$$\frac{\lambda_1^{\alpha_1}}{\alpha_1}+\frac{\lambda_2^{\alpha_2}}{\alpha_2}-\frac{\lambda_2}{\alpha_2\Gamma(1-\alpha_2)}\int_{\mathbb{R}^+}\mathrm{e}^{-\lambda_1\left(\frac{\alpha_2\Gamma(1-\alpha_2)}{\alpha_1\Gamma(1-\alpha_1)}\right)^{1/\alpha_1}x^{\alpha_2/\alpha_1}-\lambda_2 x}\,x^{-\alpha_2}\,\mathrm{d}x-$$
$$-\frac{\lambda_1}{\alpha_1\Gamma(1-\alpha_1)}\int_{\mathbb{R}^+}\mathrm{e}^{-\lambda_2\left(\frac{\alpha_1\Gamma(1-\alpha_1)}{\alpha_2\Gamma(1-\alpha_2)}\right)^{1/\alpha_2}x^{\alpha_1/\alpha_2}-\lambda_1 x}\,x^{-\alpha_1}\,\mathrm{d}x.$$

If one further assumes that $\alpha_1=\alpha_2=\alpha$, then $\psi(\lambda_1,\lambda_2;\infty)=A\{\lambda_1^\alpha+\lambda_2^\alpha-(\lambda_1+\lambda_2)^\alpha\}/\alpha$. Here below we depict the behavior of the correlation coefficient $t\mapsto\rho_\theta(t)$ for different values of $\theta>0$ and for $\alpha_1=\alpha_2=0.5$. In Figure 4.1 one notices an ordering of the curves describing the correlations between the marginal survival functions: the line at the top corresponds to the largest value of θ being considered and the lowest line is associated to the smallest value for θ.

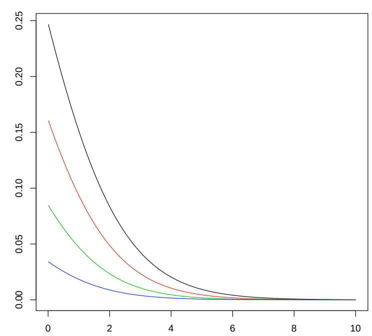

Figure 4.1: Correlation coefficient $\rho_\theta(t)$ corresponding to $\alpha=0.5$ and $\theta=10$ (first line from the top), $\theta=1$ (second line), $\theta=0.5$ (third line), $\theta=0.3$ (fourth line).

Some simplification for the above expressions of $\kappa(\theta)$ and $\psi(\lambda_1,\lambda_2;\theta)$ (with $\theta<\infty$) arises when $\alpha_1=\alpha_2=\alpha$ and $\theta=1/\alpha$. In this case one has, for $\lambda_1\neq\lambda_2$,

$$\frac{\kappa(1/\alpha;\lambda_1,\lambda_2)}{A}=\frac{\lambda_1\lambda_2}{\alpha\Gamma(1-\alpha)\Gamma(\alpha)}\int_0^\infty\frac{u^{\alpha-1}}{(\lambda_1+u)(\lambda_2+u)}\,\mathrm{d}u$$
$$=\frac{\lambda_1\lambda_2}{\alpha\Gamma(1-\alpha)\Gamma(\alpha)}\,\pi\mathrm{cosec}(\alpha\pi)\,\frac{\lambda_1^{\alpha-1}-\lambda_2^{\alpha-1}}{\lambda_2-\lambda_1}=\frac{\lambda_1\lambda_2\,[\lambda_1^{\alpha-1}-\lambda_2^{\alpha-1}]}{\alpha[\lambda_2-\lambda_1]}$$

since $\pi\mathrm{cosec}(\alpha\pi)=\Gamma(1-\alpha)\Gamma(\alpha)$. On the other hand, if $\lambda_1=\lambda_2=\lambda>0$, then $\kappa(1/\alpha;\lambda,\lambda)=A\alpha^{-1}(1-\alpha)\lambda^\alpha$ and $\psi(\lambda,\lambda;1/\alpha)=A\alpha^{-1}(1+\alpha)\lambda^\alpha$.
When $\alpha_1=\alpha_2=\alpha$ and $\theta=1/\alpha$ one can also deduce from Proposition 5 the (prior)

correlation between $S_1(t)$ and $S_2(t)$ which takes on the form

$$\rho_{1/\alpha}(t;A) = \frac{\mathrm{e}^{\frac{A(1-\alpha)}{\alpha}t} - 1}{\mathrm{e}^{\frac{A(2-2^\alpha)}{\alpha}t} - 1} \tag{4.8}$$

for any $t > 0$. It is worth pointing out a few properties of $\rho_{1/\alpha}$ in (4.8). Given $A > 0$ and $t > 0$, the function $\alpha \mapsto \rho_{1/\alpha}(t;A)$ is decreasing with

$$\lim_{\alpha\to 0} \rho_{1/\alpha}(t;A) = 1 \qquad\qquad \lim_{\alpha\to 1} \rho_{1/\alpha}(t;A) = \frac{1}{2\log 2} \ .$$

Hence, this prior specification leads to a sensible linear correlation between $S_1(t)$ and $S_2(t)$. Furthermore, one finds out that $A \mapsto \rho_{1/\alpha}(t;A)$ is decreasing for any t and α and

$$\lim_{A\to 0} \rho_{1/\alpha}(t;A) = \frac{1-\alpha}{2-2^\alpha} = \bar{\kappa}(1/\alpha) \ .$$

Hence, a prior opinion reflecting strong correlation between $S_1(t)$ and $S_2(t)$ should suggest using a low value of A.

5. Posterior analysis

Given the framework described in Section 4, we now tackle the issue of the determination of the posterior distribution of (μ_1, μ_2), given possibly right–censored data; this will also allow us to determine a Bayesian estimate of the survival functions S_1 and S_2 and to evaluate the change in the dependence structure determined by the observations.

The data consist of survival times from the two groups of individuals or generic statistical units $\{Y_j^{(1)}\}_{j=1}^{n_1}$ and $\{Y_j^{(2)}\}_{j=1}^{n_2}$. Next we let $\{c_j^{(1)}\}_{j=1}^{n_1}$ and $\{c_j^{(2)}\}_{j=1}^{n_2}$ represent the sets of censoring times corresponding to the first and second group of survival times, respectively. If $T_j^{(i)} = \min\{Y_j^{(i)}, c_j^{(i)}\}$ and $\Delta_j^{(i)} = \mathbb{1}_{(0,c_j^{(i)}]}(Y_j^{(i)})$ for $i \in \{1,2\}$, the actual data are given by $\boldsymbol{D} = \cup_{i=1}^2 \{(T_j^{(i)}, \Delta_j^{(i)})\}_{j=1}^{n_i}$. It is clear, from these definitions, that $\sum_{i=1}^2 \sum_{j=1}^{n_i} \Delta_j^{(i)} = n_e$ is the number of exact observations being recorded, whereas $n_c = n_1 + n_2 - n_e$ represents the number of censored ones regardless of the group they come from. Among the observations there might well be ties so that it is worth introducing $\{(T_j^{(1)*}, \Delta_j^{(1)*})\}_{j=1}^{k_1}$ and $\{(T_j^{(2)*}, \Delta_j^{(2)*})\}_{j=1}^{k_2}$ as the sets of distinct values of the observations relative to each group of survival data. Since some of the distinct and unique data might be shared by both groups, *i.e.* one can have that $\{(T_j^{(1)*}, \Delta_j^{(1)*})\}_{j=1}^{k_1} \cap \{(T_j^{(2)*}, \Delta_j^{(2)*})\}_{j=1}^{k_2} \neq \varnothing$, then the total number of distinct observations k in the whole sample might be less than $k_1 + k_2$. Obviously $k \le n_1 + n_2$.

For our purposes it is useful to consider the order statistic $(T_{(1)}, \ldots, T_{(k)})$, *i.e.* $0 < T_{(1)} < \cdots < T_{(k)}$, of the set of $k_1 + k_2$ observations $\cup_{i=1}^2 \{T_1^{(i)*}, \ldots, T_{k_i}^{(i)*}\}$ regardless of the group of survival times they come from. Moreover, we introduce the following set functions

$$A \mapsto \kappa_i(A) = \sum_{r=1}^{n_i} \Delta_r^{(i)}\, \mathbb{1}_A(T_r^{(i)}) \qquad A \mapsto \kappa_i^c(A) = \sum_{r=1}^{n_i} (1-\Delta_r^{(i)})\, \mathbb{1}_A(T_r^{(i)})$$

for $i \in \{1,2\}$. Their meaning is apparent: $\kappa_i(A)$ and $\kappa_i^c(A)$ are the number of exact and censored (respectively) observations from group i belonging to set A. By exploiting the functions κ_i and κ_i^c we define $\bar{N}_i(s) := \kappa_i((s,\infty))$, $\tilde{N}_i^c(s) := \kappa_i^c((s,\infty))$ and, for any $j \in \{1,\dots,k\}$ and $i \in \{1,2\}$,

$$n_{j,i} = \kappa_i(\{T_{(j)}\}) \geq 0 \qquad n_{j,i}^c = \kappa_i^c(\{T_{(j)}\}).$$

These two last quantities denote the number of exact and censored (respectively) observations from group i coinciding with $T_{(j)}$. For example, if $\max\{n_{j,1}, n_{j,2}\} = 0$, then it must be $\min\{n_{j,1}^c, n_{j,2}^c\} \geq 1$ and $T_{(j)}$ is a censored observation for group 1 or group 2 or for both groups. We also need to introduce cumulative frequencies such as $\bar{n}_{j,i} = \sum_{r=j}^k n_{r,i}$ and $\tilde{n}_{j,i}^c = \sum_{r=j}^k n_{r,i}^c$, for any $j \in \{1,\dots,k\}$. Complete these definitions by setting $\bar{n}_{k+1,i} \equiv 0$.

We are now able to provide a description of the posterior distribution of (μ_1,μ_2) given the actual data $\boldsymbol{D}$. Before stating the main result of the section we recall that henceforth $\nu_t(\mathrm{d}x_1,\mathrm{d}x_2) = \nu_t(x_1,x_2)\,\mathrm{d}x_1\mathrm{d}x_2$ for any $t>0$ and $x_1,x_2>0$. Moreover, when $t \mapsto \nu_t(x_1,x_2)$ is differentiable at $t=t_0$, we let $\nu_{t_0}'(x_1,x_2) = \dfrac{\partial \nu_t(x_1,x_2)}{\partial t}\Big|_{t=t_0}$.

Proposition 7. *Let (μ_1,μ_2) be a two–dimensional completely random measure whose Lévy intensity is such that $t \mapsto \nu_t(x_1,x_2)$ is differentiable on $\mathbb{R}^+$. Suppose that μ_1 and μ_2 are dependent. Then, the posterior distribution of (μ_1,μ_2), given data $\boldsymbol{D}$, coincides with the distribution of the random measure*

$$(\mu_1^*,\mu_2^*) + \sum_{\{r:\, \max\{\Delta_r^{(1)},\Delta_r^{(2)}\}=1\}} (J_{r,1}\delta_{T_{(r)}},\ J_{r,2}\delta_{T_{(r)}}) \tag{5.1}$$

where

(i) (μ_1^*,μ_2^*) *is a bivariate completely random measure with Lévy intensity given by*

$$\nu_t^*(x_1,x_2) = \left\{ \int_{(0,t]} \mathrm{e}^{-(\tilde{N}_1^c(s)+\bar{N}_1(s))x_1 - (\tilde{N}_2^c(s)+\bar{N}_2(s))x_2}\ \nu_s(x_1,x_2)\,\mathrm{d}s \right\}$$

(ii) *the vectors of jumps $(J_{r,1},J_{r,2})$, for $r \in \{i:\ \max\{\Delta_i^{(1)},\Delta_i^{(2)}\} = 1\}$, are mutually independent and the $r(j)$–th jump corresponding to the exact observation $y_{r(j)}^e = T_{(j)}$ has density function*

$$f_{r(j),j}(x_1,x_2) \propto \nu_{y_{r(j)}^e}'(x_1,x_2)\ \prod_{i=1}^{2} \mathrm{e}^{-(\tilde{n}_{j,i}^c+\bar{n}_{j+1,i})x_i}\ \left(1-\mathrm{e}^{-x_i}\right)^{n_{j,i}} \tag{5.2}$$

(iii) *the random measure* (μ_1^*, μ_2^*) *is independent from the jumps* $\{(J_{r,1}, J_{r,2}) : r = 1, \dots, k_e\}$, *with* k_e *denoting the total number of exact (distinct) observations in the sample.*

Proposition 7 implies a conjugacy property. Indeed, the bivariate survival function is still of the type (4.1) and it is induced by a vector of CRMs arising as the sum of: (i) a vector of CRMs with an updated Lévy intensity and without fixed jumps and (ii) a set of jumps corresponding to the exact observations. Hence, in our model we are able to preserve the conjugacy property which is known to hold true for univariate NTR priors. See Doksum (1974). Note that when ν_t is generated via a copula with marginals as in (3.1), in Proposition 7 one just needs $t \mapsto \gamma(t)$ to be differentiable and $\nu'_{t_0}(x_1, x_2) = \gamma'(t_0)\, \nu(x_1, x_2)$.

It is worth noting that the assumption of dependence between μ_1 and μ_2 can be removed. In this case, however, a slightly different representation of the posterior distribution of (μ_1, μ_2) holds true. Indeed, one has that, conditional on the observed data, μ_1 and μ_2 are still independent with

$$\mathbb{P}[\mu_1 \in A_1,\, \mu_2 \in A_2 | \boldsymbol{D}] = \mathbb{P}[\mu_1 \in A_1 \,|\, \boldsymbol{D}_1]\, \mathbb{P}[\mu_2 \in A_2 \,|\, \boldsymbol{D}_2]$$

where $\boldsymbol{D}_1 := \{(T_i^{(1)}, \Delta_i^{(1)})\}_{i=1}^{n_1}$, $\boldsymbol{D}_2 = \{(T_i^{(2)}, \Delta_i^{(2)})\}_{i=1}^{n_2}$ and one can easily verify that the representation of each marginal posterior coincides with the one provided in Doksum (1974). See also Ferguson (1974) and Ferguson and Phadia (1979).

6. Cumulative hazards

The approach on dependent survival functions we have undertaken in Sections 4 and 5 can be easily adapted to deal with vectors of cumulative hazards. A Bayesian nonparametric prior for a single cumulative hazard Λ has been first proposed by Hjort (1990), namely the celebrated beta process which is a process with independent increments. Moreover, as shown in Hjort (1990), a prior for the cumulative hazard coincides with an independent increments process if and only if the corresponding cumulative distribution function is neutral to the right. This correspondence holds true also when one considers vectors of survival or cumulative hazard functions. Firstly, following Basu (1971), define

$$\lambda(s,t) := \lim_{\Delta s \to 0 \Delta t \to 0} \mathbb{P}\left[s \le Y^{(1)} \le s + \Delta s,\ t \le Y^{(2)} \le t + \Delta t \,\middle|\, Y^{(1)} \ge s,\ Y^{(2)} \ge t \right]$$

as the hazard rate function of the vector $(Y^{(1)}, Y^{(2)})$ and, then, set

$$\Lambda(s,t) := \int_0^s \int_0^t \lambda(u,v)\, \mathrm{d}u\, \mathrm{d}v$$

as the cumulative hazard. By mimicking the construction highlighted in (1.3), one can assess a prior for Λ as follows

$$\Lambda(s,t\,|\,\mu_{1,H},\mu_{2,H})=\mu_{1,H}(0,s]\ \mu_{2,H}(0,t]$$

where $(\mu_{1,H},\mu_{2,H})$ is a vector of CRMs whose dependence is specified through a copula which gives a Lévy measure $\tilde{\nu}_H$. We will suppose that $\tilde{\nu}_H((0,t],\mathrm{d}x_1,\mathrm{d}x_2) = = \gamma(t)\,\nu_H(x_1,x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2$ where γ is a non decreasing and continuous function on $\mathbb{R}^+$. The corresponding bivariate survival function is given by

$$S(s,t\,|\,\mu_{1,H},\mu_{2,H})=\prod_{u\in(0,s]}\{1-\mu_{1,H}(\mathrm{d}u)\}\prod_{v\in(0,t]}\{1-\mu_{2,H}(\mathrm{d}v)\} \tag{6.1}$$

where $\prod_{u\in(a,b]}(1-\mu(\mathrm{d}u))$ is the usual notation for the integral product. See Gill and Johansen (1990). In order to establish the relationship between the definitions in (1.3) and in (6.1), suppose $s<t$ and set $\{u_{m,j}\}_{j=1}^{k_m}$ to be an arbitrary sequence of ordered points $0=u_{m,1}<u_{m,2}<\cdots<u_{m,k_m}=t$ such that $\lim_{m\to\infty}\max_{1\le j\le k_m-1}(u_{m,j+1}-u_{m,j})=0$. According to this notation, (6.1) leads to

$$S(s,t)=\lim_{m\to\infty}\prod_{\{j:\,u_{m,j}\in(0,s]\}}\{1-\mu_{1,H}(I_{m,j})\}\,\{1-\mu_{2,H}(I_{m,j})\}\ \times \\ \times\ \lim_{m\to\infty}\prod_{\{j:\,u_{m,j}\in(s,t]\}}\{1-\mu_{2,H}(I_{m,j})\}$$

where for simplicity of notation we have dropped the dependence of S on $(\mu_{1,H},\mu_{2,H})$ and $I_{m,j}=(u_{m,j-1},u_{m,j}]$. Given the independence of the increments of $(\mu_{1,H},\mu_{2,H})$, the evaluation of $\mathrm{E}[S^n(s,t)]$ can be accomplished if one determines moments of the type $\mathrm{E}[\{1-\mu_{1,H}(I_{m,j})\}^n\{1-\mu_{2,H}(I_{m,j})\}^n]$. The latter can be deduced from the Lévy–Khintchine representation of the Laplace transform of $(\mu_{1,H},\mu_{2,H})$ which yields

$$\mathrm{E}[\{1-\mu_{1,H}(I_{m,j})\}^n\{1-\mu_{2,H}(I_{m,j})\}^n]= \\ =\Delta_\gamma(I_{m,j})\int_{(0,1)^2}[1-(1-x_1)^n(1-x_2)^n]\,\nu_H(\mathrm{d}x_1,\mathrm{d}x_2)+o(\Delta_\gamma(I_{m,j}))$$

as $m\to\infty$, where $\Delta_\gamma(I_{m,j})=\gamma(u_{m,j})-\gamma(u_{m,j-1})$. Hence

$$\mathrm{E}\left[S^n(s,t)\right]=\exp\left\{-\gamma(s)\int_{(0,1)^2}[1-(1-x_1)^n(1-x_2)^n]\,\nu_H(\mathrm{d}x_1,\mathrm{d}x_2)\right\}\ \times \\ \times\ \exp\left\{-(\gamma(t)-\gamma(s))\int_{(0,1)}[1-(1-x_2)^n]\,\nu_{2,H}(\mathrm{d}x_2)\right\}$$

This coincides with the n–th moment of $S(s,t)$ defined according to (1.3) if and only if $\nu_H(\{(x_1,x_2)\in(0,1)^2:(-\log(1-x_1),-\log(1-x_2))\in A\})=\nu(A)$ for any measurable subset A of $(0,\infty)^2$, where ν is the Lévy intensity of the vector (μ_1,μ_2). Given this correspondence between priors for bivariate cdf's and priors for cumulative hazards, one naturally expects that the copula yielding ν from the marginals ν_1 and ν_2 coincides with the copula which gives rise to ν_H when starting from marginals $\nu_{1,H}$ and $\nu_{2,H}$. And one can show this is, indeed, the case.

Remark 1. It should be noted that an alternative model for the marginal cumulative hazards consists in the use of kernel mixtures of completely random measures. In other words, if $k_i : \mathbb{R}^+ \times \mathbb{X} \to \mathbb{R}^+$, for $i = 1,2$, are kernel functions, then one can set marginal cumulative hazards $\Lambda_i(t)$ as

$$\Lambda_i(t) = \int_0^t \int_{\mathbb{R}^+} k_i(x,s)\,\mu_i(\mathrm{d}s)\,\mathrm{d}x$$

where μ_i is a CRM with intensity measure of the form in (3.1). This yields the random survival function $S(t_1,t_2) = \exp\{-\Lambda_1(t_1) - \Lambda_2(t_2)\}$ where it is apparent that the bivariate process $\{(\Lambda_1(t),\Lambda_2(t)) :\ t \ge 0\}$ does not have independent increments. For the univariate case, this approach has been undertaken by Dykstra and Laud (1981) with a kernel $k(x,s) = \mathbb{1}_{[s,\infty)}(x)$ which yields monotone increasing hazard rates. A treatment for a general kernel has been provided by Lo and Weng (1989). In our setting, $(\int k(x,s)\,\mu_1(\mathrm{d}s),\ \int k(x,s)\,\mu_2(\mathrm{d}s))$ defines a prior for a vector of hazard rates which allows to draw inferences on the correspponding vector of survival functions. Note that if one uses the kernel in Dykstra and Laud (1981), one then has $\Lambda_i(t) = \int_0^t (t-s)\,\mu_i(\mathrm{d}s)$. Correspondingly

$$\mathrm{E}\left[S(t_1,t_2)\right] = \exp\left\{-\int_0^{t_1} \psi_1(t_1-s)\,\gamma'(s)\,\mathrm{d}s - \int_0^{t_2} \psi_2(t_2-s)\,\gamma'(s)\,\mathrm{d}s\right\} \times \exp\left\{-\int_0^{t_1} \zeta(t_1-s,t_2-s)\,\gamma'(s)\,\mathrm{d}s\right\}$$

where $\zeta(u,v) = \int_0^\infty \int_0^\infty (1-\mathrm{e}^{-ux_1})\,(1-\mathrm{e}^{-vx_2})\,\nu(x_1,x_2)\,\mathrm{d}x_1\mathrm{d}x_2$ for any $u,v > 0$. This model selects an absolutely continuous distribution for each component of the vector of survival functions, thus leading to smoother posterior estimates of the marginal survival functions. One can also deduce the posterior distribution of (μ_1,μ_2) given right–censored data, thus extending a result obtained in James (2005). It is however expected that, in a similar fashion as in the univariate case, one should resort to some simulation algorithm for obtaining a numerical evaluation of Bayesian estimates of quantities of interest. However we will not linger on this point and leave it as an issue to be dealt in future work.

7. Estimate of the survival functions

The results we have achieved so far easily yield a Bayesian estimate of the survival functions S_1 and S_2 and the correlation between them. The starting point is the Bayesian estimate of the survival function $S(t_1, t_2)$ defined in (1.3) that will be taken to coincide with the posterior mean of $\mathbb{P}[Y^{(1)} > t_1, Y^{(2)} > t_2 \,|\, (\mu_1, \mu_2)]$. This will enable us to estimate S_1 and S_2 and to evaluate the posterior correlation which requires the knowledge of the posterior second moment of $\mathbb{P}[Y^{(i)} > t \mid \mu_i]$. One has

Corollary 1. *Let $\mathcal{I}_t = \{j : \Delta_j^{(1)} \wedge \Delta_j^{(2)} = 1\} \cap \{j : T_{(j)} \leq t\}$ be the set of indices corresponding to the exact observations recorded up to time t and let $T_{(k+1)} = \infty$. For any $t > 0$, the posterior mean $\hat{S}(t,t)$ of $\mathbb{P}[Y^{(1)} > t, Y^{(2)} > t \,|\, (\mu_1, \mu_2)]$, given data $\boldsymbol{D}$, coincides with*

$$\exp\left\{ -\sum_{j=1}^{k+1} \left[\gamma(t \wedge T_{(j)}) - \gamma(T_{(j-1)})\right] \mathbb{1}_{[T_{(j-1)},\infty)}(t)\, \psi_j^*(1,1) \right\} \times \tag{7.1}$$

$$\times \prod_{j \in \mathcal{I}_t} \int_{(\mathbb{R}^+)^2} \mathrm{e}^{-x_1 - x_2} f_{r(j),j}(x_1, x_2)\, \mathrm{d}x_1\, \mathrm{d}x_2$$

where $\psi_j^*(\lambda_1, \lambda_2) = \int_{(\mathbb{R}^+)^2} \left[1 - \mathrm{e}^{-\lambda_1 x_1 - \lambda_2 x_2}\right] \mathrm{e}^{-\sum_{i=1}^2 (\tilde{n}_{j,i}^c + \bar{n}_{j,i}) x_i} \nu(x_1, x_2) \mathrm{d}x_1 \mathrm{d}x_2$ and the $f_{r(j),j}$ are the density functions of the jumps as described in (5.2).

The expression of $\hat{S}(t,t)$ provided in (7.1) is the building block for obtaining the estimate of S corresponding to any pair of points $(t_1, t_2) \in (\mathbb{R}^+)^2$. First of all notice that, for any $t > 0$, $\hat{S}(t, 0)$ and $\hat{S}(0, t)$ provide estimates of the marginal survival functions of $Y^{(1)}$ and $Y^{(2)}$, respectively. They both can be determined from (7.1).

Setting $\mathcal{I}_{1,t} = \{j : \Delta_j^{(1)} = 1\} \cap \{j : T_{(j)} \leq t\}$, one has

$$\hat{S}(t, 0) = \exp\left\{ -\sum_{j=1}^{k+1} \left[\gamma(t \wedge T_{(j)}) - \gamma(T_{(j-1)})\right] \mathbb{1}_{[T_{(j-1)},\infty)}(t)\, \psi_j^*(1,0) \right\} \times$$

$$\times \prod_{j \in \mathcal{I}_{1,t}} \frac{\int_{(\mathbb{R}^+)^2} \mathrm{e}^{-\sum_{i=1}^2 (1 + \tilde{n}_{j,i}^c + \bar{n}_{j+1,i}) x_i} (1 - \mathrm{e}^{-x_1})^{n_{j,1}} (1 - \mathrm{e}^{-x_2})^{n_{j,2}} \nu(x_1, x_2)\, \mathrm{d}x_1\, \mathrm{d}x_2}{\int_{(\mathbb{R}^+)^2} \mathrm{e}^{-\sum_{i=1}^2 [\tilde{n}_{j,i}^c + \bar{n}_{j+1,i}] x_i} (1 - \mathrm{e}^{-x_1})^{n_{j,1}} (1 - \mathrm{e}^{-x_2})^{n_{j,2}} \nu(x_1, x_2)\, \mathrm{d}x_1\, \mathrm{d}x_2}.$$

With the appropriate modifications, one determines an expression for $\hat{S}(0, t)$ as well. Finally, using the independence of the increments of the random measure in (5.1) and supposing that $s > t$, one has $\hat{S}(s, t) = \hat{S}(t, t)\, \hat{S}(s, 0)/\hat{S}(t, 0)$. A similar expression can be found for the case where $s < t$.

Furthermore, the posterior second moment of the marginal survival S_1 is given by

$$\hat{S}_{12}(t) = \exp\left\{ -\sum_{j=1}^{k+1} \left[\gamma(t \wedge T_{(j)}) - \gamma(T_{(j-1)})\right] \mathbb{1}_{[T_{(j-1)},\infty)}(t)\, \psi_j^*(2,0) \right\} \times$$

$$\times \prod_{j\in\mathcal{I}_{1,t}} \frac{\int_{(\mathbb{R}^+)^2} \mathrm{e}^{-\sum_{i=1}^2 (2+\tilde n^c_{j,i}+\bar n_{j+1,i})x_i}\,(1-\mathrm{e}^{-x_1})^{n_{j,1}}(1-\mathrm{e}^{-x_2})^{n_{j,2}}\,\nu(x_1,x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2}{\int_{(\mathbb{R}^+)^2} \mathrm{e}^{-\sum_{i=1}^2 [\tilde n^c_{j,i}+\bar n_{j+1,i}]x_i}\,(1-\mathrm{e}^{-x_1})^{n_{j,1}}(1-\mathrm{e}^{-x_2})^{n_{j,2}}\,\nu(x_1,x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2}\,.$$

It follows that an estimate of the correlation between S_1, S_2 can be obtained as

$$\hat\rho(S_1(t),\,S_2(t)) = \frac{\hat S(t,t)-\hat S(t,0)\hat S(0,t)}{\sqrt{(\hat S_{12}(t)-\hat S^2(t,0))(\hat S_{21}(t)-\hat S^2(0,t))}}\,. \tag{7.2}$$

From a computational point of view, one can usefully resort to the simple identity $\psi_j^*(\lambda_1,\lambda_2)=\psi(\lambda_1+\tilde n^c_{j,1}+\bar n_{j,1},\,\lambda_2+\tilde n^c_{j,2}+\bar n_{j,2})-\psi(\tilde n^c_{j,1}+\bar n_{j,1},\,\tilde n^c_{j,2}+\bar n_{j,2})$ and to

$$\begin{aligned}
&\int_{(\mathbb{R}^+)^2} \mathrm{e}^{-q_1x_1-q_2x_2}(1-\mathrm{e}^{-x_1})^{n_{j,1}}(1-\mathrm{e}^{-x_2})^{n_{j,2}}\,\nu(x_1,x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 =\\
&\quad= \mathbb{1}_{\{0\}^c}(n_{j,1})\sum_{k=1}^{n_{j,1}}\binom{n_{j,1}}{k}(-1)^{k+1}\left[\psi(k+q_1,q_2)-\psi(q_1,q_2)\right]\\
&\qquad+ \mathbb{1}_{\{0\}^c}(n_{j,2})\sum_{k=1}^{n_{j,2}}\binom{n_{j,2}}{k}(-1)^{k+1}\left[\psi(q_1,k+q_2)-\psi(q_1,q_2)\right]\\
&\qquad\quad- \mathbb{1}_{\{0\}^c}(n_{j,1})\mathbb{1}_{\{0\}^c}(n_{j,2})\sum_{k_1=1}^{n_{j,1}}\sum_{k_2=1}^{n_{j,2}}\binom{n_{j,1}}{k_1}\binom{n_{j,2}}{k_2}(-1)^{k_1+k_2}\times\\
&\qquad\qquad\times\left[\psi(k_1+q_1,k_2+q_2)-\psi(q_1,q_2)\right]\,.
\end{aligned}$$

These formulae make it clear that the only difficulty in evaluating posterior estimates, given the (possibly right–censored) data, lies in the evaluation of the bivariate Laplace exponent $\psi(\lambda_1,\lambda_2)$ for a set of non–negative integer values of (λ_1,λ_2). In particular, if the two–dimensional completely random measure (μ_1,μ_2) is constructed by means of a Clayton copula C_θ, Proposition 1 suggests that $\psi(\lambda_1,\lambda_2)$ can be easily evaluated either numerically or through some simulation scheme. Indeed, it is unlikely that one can obtain a closed analytic form for $\kappa(\theta;\lambda_1,\lambda_2)$ since it is hard to evaluate exactly the two–dimensional integral in that case. However, one can hope to evaluate $\kappa(\theta;\lambda_1,\lambda_2)$ through some numerical integration rule or via a simple Monte Carlo simulation scheme. As for the latter, one just needs to simulate a sample $\{(x_1^{(i)},x_2^{(i)})\}_{i=1}^M$ from the distribution of a vector of independent and exponentially distributed random variables with rate parameters λ_1 and λ_2, respectively. Hence, one obtains

$$\hat\kappa_M(\theta;\lambda_1,\lambda_2)=\frac{1}{M}\sum_{i=1}^M C_\theta(U_1(x_1^{(i)}),U_2(x_2^{(i)})) \tag{7.3}$$

as an approximate evaluation of $\kappa(\theta;\lambda_1,\lambda_2)$ which can be used to compute $\psi(\lambda_1,\lambda_2;\theta)$.

Considering the α_1 and α_2 marginal stable processes with a Clayton copula C_θ, as discussed in Example 1, one has

$$C_\theta(U_1(x_1),U_2(x_2)) = A\left\{[\alpha_1\Gamma(1-\alpha_1)x_1^{\alpha_1}]^\theta + [\alpha_2\Gamma(1-\alpha_2)x_2^{\alpha_2}]^\theta\right\}^{-1/\theta}$$

so that an approximated evaluation of $\psi(\lambda_1,\lambda_2;\theta)$ coincides with

$$\frac{A}{\alpha_1}\,\lambda_1^{\alpha_1} + \frac{A}{\alpha_2}\,\lambda_2^{\alpha_2} - \frac{A}{M}\sum_{i=1}^{M}\left\{\left[\alpha_1\Gamma(1-\alpha_1)(x_1^{(i)})^{\alpha_1}\right]^\theta + \left[\alpha_2\Gamma(1-\alpha_2)(x_2^{(i)})^{\alpha_2}\right]^\theta\right\}^{-1/\theta} .$$

These approximate evaluations can be replaced by exact computations when $\alpha_1 = \alpha_2 = \alpha = 1/\theta$ since, as highlighted in Example 1, one has

$$\psi(\lambda_1,\lambda_2;1/\alpha) = \frac{A}{\alpha}\left\{\frac{\lambda_2^{\alpha+1}-\lambda_1^{\alpha+1}}{\lambda_2-\lambda_1}\,\mathbb{1}_{\lambda_1\neq\lambda_2} + (1+\alpha)\,\lambda^\alpha\,\mathbb{1}_{\lambda_1=\lambda_2=\lambda}\right\}. \tag{7.4}$$

In the remaining part of the present section we deal with an illustrative example where we point out a possible MCMC sampling scheme to be implemented.

Example 2. (*Skin grafts data*). The dataset we are now going to examine has been already studied in the literature by Woolson and Lachenbruch (1980), Lin and Ying (1993) and Bulla, Muliere and Walker (2007). The data consist of survival times of closely matched and poorly matched skin grafts, with both grafts applied to the same burn patient. The strength of matching between donor and recipient has been evaluated in accordance with the HL-A transplantation antigen system. The data can, then, be split into two groups $\mathscr{Y}^{(1)}$ and $\mathscr{Y}^{(2)}$ including the days of survival of closely matched and poorly matched, respectively, skin grafts on burn patients. In this case one has $\mathscr{Y}^{(1)} = \{37,\ 19,\ 57^+,\ 93,\ 16,\ 22,\ 20,\ 18,\ 63,\ 29,\ 60^+\}$ and $\mathscr{Y}^{(2)} = \{29,\ 13,\ 15,\ 26,\ 11,\ 17,\ 26,\ 21,\ 43,\ 15,\ 40\}$, where data denoted as t^+ stand for right–censored times. For our purposes, we consider a model in which

$$\nu_1(x) = \nu_2(x) = \frac{A}{\Gamma(1-\alpha)}\,x^{-\alpha-1}$$

and Clayton copula $C_{1/\alpha}$ which can be easily seen to produce a bivariate Lévy intensity given by

$$\nu(x_1,x_2) = \frac{A(1+\alpha)}{\Gamma(1-\alpha)}\,(x_1+x_2)^{-\alpha-2}$$

According to Proposition 3, the prior guess at the shape of the survival function, conditional on a specific value of α, is

$$\mathrm{E}\left[\mathbb{P}\left[Y^{(1)} > s,\ Y^{(2)} > t \,|\, (\mu_1,\mu_2)\right]\right] = \mathrm{e}^{-A\left[s+t+\frac{(1-\alpha)}{\alpha}\,(t\vee s)\right]}$$

for any $s,t\geq 0$. If a prior for α is specified, one can adopt a Metropolis–Hastings algorithm to evaluate a posterior estimate of $S(t_1,t_2)$ given by

$$\hat{S}(t_1,t_2) = \int_{(0,1)} \mathrm{E}\left[\mathrm{e}^{-\mu_1(0,t_1]-\mu_2(0,t_2]} \,\middle|\, \boldsymbol{D},\alpha\right]\, \pi(\mathrm{d}\alpha \,|\, \boldsymbol{D})$$

$\pi(\,\cdot\,|\boldsymbol{D})$ denoting the posterior distribution of α given the data $\boldsymbol{D}$. Hence, one generates a sample $\{\alpha^{(1)},\dots,\alpha^{(M)}\}$ from the posterior distribution $\pi(\,\cdot\,|\boldsymbol{D})$ of α, given the data $\boldsymbol{D}$, and evaluate $\hat{S}(t_1,t_2) \approx (1/M)\sum_{i=1}^M \mathrm{E}\left[\mathrm{e}^{-\mu_1(0,t_1]-\mu_2(0,t_2]} \,\middle|\, \boldsymbol{D},\alpha^{(i)}\right]$. In this case the implementation of the Metropolis–Hastings algorithm is straightforward. Indeed the likelihood turns out to be equal to

$$\exp\left\{-\sum_{j=1}^N \left[\gamma(T_{(j)})-\gamma(T_{(j-1)})\right]\psi(\tilde{n}^c_{j,1}+\bar{n}_{j,1},\, \tilde{n}^c_{j,2}+\bar{n}_{j,2})\right\}\times$$
$$\times\prod_{j\in\mathcal{I}}\gamma'(T_{(j)})\int_{(\mathbb{R}^+)^2}\mathrm{e}^{-\sum_{i=1}^2(\tilde{n}^c_{j,i}+\bar{n}_{j+1,i})x_i}\,(1-\mathrm{e}^{-x_1})^{n_{j,1}}(1-\mathrm{e}^{-x_2})^{n_{j,2}}\,\nu(x_1,x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2$$

and it can be computed exactly since the Laplace exponent ψ has a very simple form as described in (7.4). In order to implement the simulation scheme we fixed a prior beta$(0.5,5)$, which is highly concentrated around zero and reflects a strong prior opinion of a high degree of correlation between the marginal survival functions S_1 and S_2. Moreover, we choose a uniform distribution on $(0,1)$ as the proposal of the algorithm and set $A = 0.01$. Of course, one could also set a prior for A and incorporate it into the sampling scheme. We did not address such an issue here. We have performed 10000 iterations, the first 2000 of which were dropped as burn–in moves. The first interesting thing about the results we have obtained is that, despite the particular structure of the prior of α, the posterior estimate of $\hat{\alpha} = (1/M)\sum_{i=1}^M \alpha^{(i)}$ is equal to $\hat{\alpha}\approx 0.7306$. The plots of sections of the estimates of the survival functions $t_1 \mapsto \hat{S}(t_1,t_2)$ for $t_2 \in \{0,11,40,93\}$ are depicted on the left–hand side of Figure 7.2 whereas the plots of the function $t_2 \mapsto \hat{S}(t_1,t_2)$ for $t_1 \in \{0,13,26,43,93\}$ are given on the right–hand side of Figure 7.2.

We have also examined the correlation structure as modified by the data. In particular, one can note that the data have induced the following effects on the dependence between S_1 and S_2: (a) they have determined a sensible reduction of the magnitude of the correlation at any value of t from values in $[0.77,0.95]$ to values ranging between 0.37 and 0.64; (b) the correlation function is not monotone. See Figure 7.3, where a plot of the correlations between survival functions S_1 and S_2 are plotted for values of t up to the maximum value among the observations, i.e. 93. The prior correlation has been evaluated, for any t, by means of a simple Monte Carlo procedure by drawing a sample of α's from the beta$(0.5,5)$ distribution and, then, averaging the expression in (4.8). As

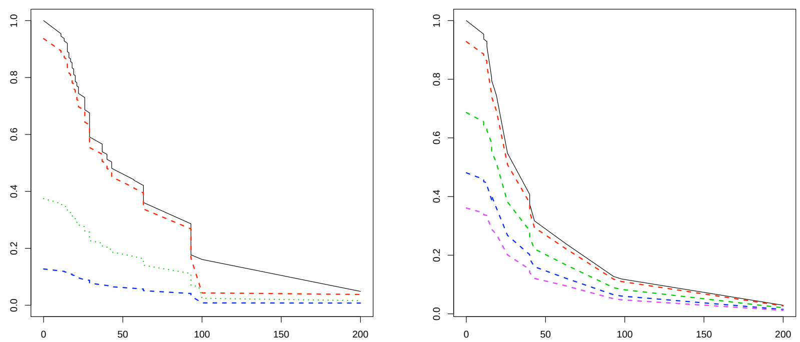

Figure 7.2: The estimated marginal survival functions of $Y^{(1)}$ and $Y^{(2)}$ arising from the application of a MCMC algorithm for the skin grafts data. On the left–hand side the plots of $t_1 \mapsto \hat{S}(t_1,t_2)$ for $t_2 = 0$ (solid line) and $t_2 \in \{11, 40, 93\}$ (dashed lines in decreasing order). On the right–hand side the plots of $t_2 \mapsto \hat{S}(t_1,t_2)$ for $t_1 = 0$ (solid line) and $t_1 \in \{13, 26, 43, 93\}$ (dashed lines in decreasing order).

for the posterior correlations, we have used the output of the MCMC algorithm in order to evaluate mixed and marginal posterior moments of S_1 and S_2 according to formula (7.2). The comparison between prior and posterior correlations is quite interesting. As expected, the prior correlation is a decreasing function of t and it takes on values very close to 1: this is explained by the fact that the prior for α is concentrated around 0 which identifies the situation of complete dependence in the Clayton copula (3.2) with $\theta = 1/\alpha$. On the other hand, *a posteriori* the data have a sensible impact on the correlation. First one notes that there is no monotonicity. Secondly, the points where the correlation is decreasing identify time intervals where the observed behaviour of $Y^{(1)}$ and $Y^{(2)}$ sensibly differs. For example, in $[11, 16)$ one observes only failures for $Y^{(2)}$. The correlations, then, reaches a local minimum at $t = 26$ where two exact observations for $Y^{(2)}$ have been recorded. Other local minima are located at $t = 57$ and $t = 60$ which are censored data for $Y^{(1)}$.

8. Concluding remarks

The results we have achieved allow for Bayesian inference on vectors of survival, or cumulative hazard, functions. Nonetheless, the idea of using Lévy copulas for building vectors of completely random measures might also be the starting point for defining nonparametric priors for vectors of paired survival data $(Y^{(1)}, Y^{(2)})$. While there is a wealth of papers on Bayesian nonparametric estimation of univariate survival functions, we are not aware of many contributions to inference for bivariate survival functions. An example is given by Bulla, Muliere and Walker (2007) where a generalized Pólya urn scheme is used to obtain a bivariate reinforced process which, in turn, can be applied to deduce an estimator of the bivariate survival function. Moreover, in Nieto–Barajas and Walker (2007) the authors assume conditional independence between lifetimes and model

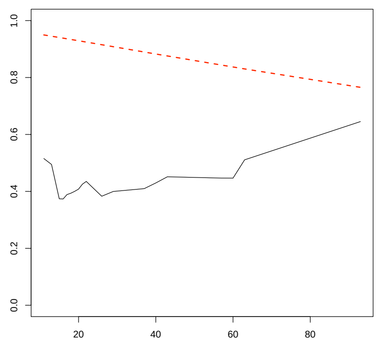

Figure 7.3: Plots of the correlation between $S_1(t)$ and $S_2(t)$ for values of t coinciding with the observed data, both exact and censored. Red dashed lines for prior correlations and continuous line for posterior correlations.

nonparametrically each marginal density: the bivariate density is then obtained as a mixture. In Ghosh, Hjort, Messan and Ramamoorthi (2006) a nonparametric prior based on beta processes is adopted and the updating rule is described: the authors show it does not lead to inconsistencies analogous to those featured by some frequentist nonparametric estimators.

An important issue we did not consider concerns an investigation of the properties of consistency of the prior we have proposed. In other terms, if one supposes the data are independently generated by survival function $S_{1,0}$ and $S_{2,0}$, it is worth to check whether the posterior distribution of (S_1, S_2) concentrates on a suitable neighbourhood of $(S_{1,0}, S_{2,0})$ as the sample size increases. It is expected that in this case one can extend results similar to those achieved in Kim and Lee (2001) for NTR priors or results in Draghici and Ramamoorthi (2003) and De Blasi, Peccati and Prünster (2009) for the mixture models mentioned in the previous Remark 1. This will be pursued in future work.

Acknowledgements. The authors wish to thank Massimo Santini (Dep. of Computer Science, Università di Milano) for his valuable support in developing the MCMC algorithm adopted in Example 2.

Appendix

A1. Proof of Proposition 1. Set $C^{(2)}(U_1(x_1), U_2(x_2)) := \dfrac{\partial^2}{\partial u\, \partial v} C(u,v)\,\big|_{u=U_1(x_1),\, v=U_2(x_2)}$. Since the Lévy intensity $\nu(x_1, x_2)$ in (3.3) is of finite variation, the Laplace exponent is

$$\psi(\lambda_1, \lambda_2) = \int_{(\mathbb{R}^+)^2} \left[1 - \mathrm{e}^{-\lambda_1 x_1 - \lambda_2 x_2}\right] C^{(2)}(U_1(x_1), U_2(x_2))\, \nu_1(x_1)\, \nu_2(x_2)\, \mathrm{d}x_1\, \mathrm{d}x_2$$

$$= (\theta + 1) \int_{(\mathbb{R}^+)^2} \left[1 - \mathrm{e}^{-\lambda_1 x_1 - \lambda_2 x_2}\right] \frac{U_1^\theta(x_1)\, U_2^\theta(x_2)\ \nu_1(x_1) \nu_2(x_2)}{\left\{U_1^\theta(x_1) + U_2^\theta(x_2)\right\}^{1/\theta + 2}}\, \mathrm{d}x_1\, \mathrm{d}x_2\ .$$

When integrating by parts the above integral, one obtains the expression in (3.4). □

A2. Proof of Proposition 2. As $\theta \to 0$, the Lévy density tends to the independence case and in (3.4) one has $\psi(\lambda_1, \lambda_2) = \psi_\perp(\lambda_1, \lambda_2)$ for any $\lambda_1 > 0$ and $\lambda_2 > 0$. This implies (i). Moreover, $C_\theta(U_1(x_1), U_2(x_2)) \le \min\{U_1(x_1), U_2(x_2)\}$ for any $\theta > 0$. The Lévy measure $\nu(\mathrm{d}x_1, \mathrm{d}x_2; \infty)$ corresponding to the perfect dependence case does not admit a density on $(\mathbb{R}^+)^2$ but it still of finite variation. Indeed, if U_i^{-1} denotes the inverse of the i–th tail integral ($i = 1, 2$), one has

$$
\begin{aligned}
\int_{\|\boldsymbol{x}\|\le 1} \|\boldsymbol{x}\|\, \nu(\mathrm{d}x_1, \mathrm{d}x_2; \infty) &= \int_{\{\|\boldsymbol{x}\|\le 1\}\cap\{U_1(x_1)=U_2(x_2)\}} (x_1^2 + x_2^2)^{1/2}\, \nu(\mathrm{d}x_1, \mathrm{d}x_2; \infty) \\
&= \int_{\left\{x^2+\left(U_2^{-1}(U_1(x))\right)^2\le 1\right\}\cap\{U_1(x)\le U_2(x)\}} \left(x^2 + \left(U_2^{-1}(U_1(x))\right)^2\right)^{1/2} \nu_1(x)\,\mathrm{d}x + \\
&\quad + \int_{\left\{x^2+\left(U_2^{-1}(U_1(x))\right)^2\le 1\right\}\cap\{U_1(x)> U_2(x)\}} \left(x^2 + \left(U_2^{-1}(U_1(x))\right)^2\right)^{1/2} \nu_1(x)\,\mathrm{d}x \\
&\le \sqrt{2}\int_{\{x\le U_1^{-1}(U_2(1/\sqrt{2}))\}} x\,\nu_1(x)\,dx + \sqrt{2}\int_{\{x\le 1/\sqrt{2}\}} U_2^{-1}(U_1(x))\,\nu_1(x)\,dx \\
&= \sqrt{2}\int_{\{x\le U_1^{-1}(U_2(1/\sqrt{2}))\}} x\,\nu_1(x)\,dx + \sqrt{2}\int_{\{x\le U_2^{-1}(U_2(1/\sqrt{2}))\}} x\,\nu_2(x)\,dx < \infty
\end{aligned}
$$

and this is finite since both ν_1 and ν_2 are of finite variation. Consequently, the Laplace functional transform of the two–dimensional independent increments process corresponding to the complete dependence case admits a Lévy–Khintchine representation. This implies that $\min\{U_1(x_1), U_2(x_2)\}$ is integrable on $(\mathbb{R}^+)^2$ with respect to $\mathrm{e}^{-x_1-x_2}$. A simple application of the dominated convergence theorem now yields (ii). Finally, (iii) holds true since $\theta \mapsto C_\theta(x, y)$ is an increasing function for any $x, y > 0$. □

A3. Proof of Proposition 3. By virtue of the adopted model $\mathbb{P}(Y^{(1)} > s,\, Y^{(2)} > t) = \mathrm{E}\left[\mathrm{e}^{-\mu_1(0,s]-\mu_2(0,t]}\right]$. If $s \le t$, as noted in Equation (2.3), the independence of the increments of (μ_1, μ_2) implies

$$
\begin{aligned}
\mathbb{P}[Y^{(1)} > s, Y^{(2)} > t] &= \mathrm{E}\left[\mathrm{e}^{-\mu_1(0,s]-\mu_2(0,s]}\right]\ \mathrm{E}\left[\mathrm{e}^{-\mu_2(s,t]}\right] \\
&= \exp\{-\gamma(s)\,\psi(1,1) - (\gamma(t)-\gamma(s))\,\psi_2(1)\}\ .
\end{aligned}
$$

A similar representation holds true for $s > t$ and the conclusion stated in (4.2) follows. The positivity of the coefficients ξ_1, ξ_2 and $\xi_{1,2}$ follows from the definition of the Laplace exponent ψ. □

A4. Proof of Proposition 4. If $(F_1(s), F_2(t)) \overset{\mathrm{d}}{=} (1 - \mathrm{e}^{-\mu_{1,s}}, 1 - \mathrm{e}^{-\mu_{2,t}})$, then it is easy to find that $F = F_1 F_2$ satisfies condition (4.3), with $V_{i,j} = 1 - \mathrm{e}^{-(\mu_{i,t_j} - \mu_{i,t_{j-1}})}$, for $i \in \{1, 2\}$, $j = 1, \ldots, k$ and $t_0 = 0$. Conversely, let $\mu_{i,t} = -\log(1 - F_i(t))$, for $i \in \{1, 2\}$ and suppose that for any choice of $k \ge 1$ and $0 < t_1 < \cdots < t_k$ there exist k independent random vectors $(V_{1,1}, V_{2,1}), \ldots, (V_{1,k}, V_{2,k})$ such that condition (4.3) holds.

It follows by Theorem 3.1 in Doksum (1974) that both the marginal processes $\mu_{1,s}$ and $\mu_{2,t}$ start from $(0,0)$ and are stochastically continuous, almost surely non-decreasing and transient. Furthermore, $(\mu_{1,t_j}-\mu_{1,t_{j-1}},\mu_{2,t_j}-\mu_{2,t_{j-1}})=(-\log(1-V_{1,j}),-\log(1-V_{2,j}))$, for $j=1,\ldots,k$. Hence, the process $(\mu_{1,t},\mu_{2,t})_{t\geq 0}$ has independent increments. We can conclude that (μ_1,μ_2) is a completely random measure. $\square$

A5. Proof of Proposition 5. First of all, it can be easily seen that

$$\mathrm{Cov}(F_1(t),F_2(t))=\mathrm{Cov}(S_1(t),S_2(t))=\mathrm{e}^{-\gamma(t)\,\psi_\perp(1,1)}\left\{\mathrm{e}^{-\gamma(t)[\psi(1,1)-\psi_\perp(1,1)]}-1\right\}$$

and $\mathrm{Var}(F_i(t))=\mathrm{e}^{-2\gamma(t)\,\psi_i(1)}\left\{\mathrm{e}^{-\gamma(t)[\psi_i(2)-2\psi_i(1)]}-1\right\}$ so that Formula (4.4) follows by noting that $\psi(1,1)-\psi_\perp(1,1)=-\kappa(\theta)$ and $\psi_i(2)-2\psi_i(1)=-\kappa_i$. Moreover, if the two marginal Lévy densities coincide, *i.e.* $\nu_1=\nu_2=\nu^*$, then one has

$$\begin{aligned}\kappa(\infty)&=\int_{(\mathbb{R}^+)^2}\mathrm{e}^{-x_1-x_2}\,\min\{U_1(x_1),U_2(x_2)\}\,\mathrm{d}x_1\mathrm{d}x_2\\&=\int_{(\mathbb{R}^+)^2}\mathrm{e}^{-x_1-x_2}\,U^*(x_1\wedge x_2)\}\,\mathrm{d}x_1\mathrm{d}x_2=2\int_0^\infty\mathrm{e}^{-x_2}\int_0^{x_2}\mathrm{e}^{-x_1}U_2(x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2\\&=2\int_0^\infty(1-\mathrm{e}^{-x})\,\mathrm{e}^{-x}\,U^*(x)\,\mathrm{d}x=\int_0^\infty(1-\mathrm{e}^{-x})^2\,\nu^*(\mathrm{d}x)=\kappa_1=\kappa_2\end{aligned}$$

where $U^*(x)=\int_x^\infty\nu^*(s)\,\mathrm{d}s$ for any $x>0$. From this representation of $\kappa(\infty)$ and (4.4) one has $\rho_\theta(t)=[\mathrm{e}^{\gamma(t)\kappa(\theta)}-1]/[\mathrm{e}^{\gamma(t)\kappa(\infty)}-1]$. Recalling the properties of the function γ, one has that $\lim_{t\to 0}\rho_\theta(t)=\bar\kappa(\theta)$. On the other hand, $t\mapsto\rho_\theta(t)$ is a decreasing function since $\kappa(\theta)<\kappa(\infty)$, for any $\theta>0$, with $\lim_{t\to\infty}\rho_\theta(t)=0$. Hence $\rho_\theta(t)<\bar\kappa(\theta)$. $\square$

A6. Proof of Proposition 6. Proposition 3 provides $\mathbb{P}(Y_i>t)=\exp\{-\gamma(t)\psi_i(1)\}$. From this one deduces that $\mathrm{E}[Y_i]=\int_0^\infty P(Y_i>t)\,\mathrm{d}t=\int_0^\infty\mathrm{e}^{-\gamma(t)\,\psi_i(1)}\,\mathrm{d}t$ and

$$\mathrm{Var}[Y_i]=2\int_0^\infty tP(Y_i>t)\,\mathrm{d}t-(\mathrm{E}[Y_i])^2=2\int_0^\infty t\,\mathrm{e}^{-\gamma(t)\psi_i(1)}\,\mathrm{d}t-\left(\int_0^\infty\mathrm{e}^{-\gamma(t)\psi_i(1)}\,\mathrm{d}t\right)^2.$$

Finally

$$\begin{aligned}\mathrm{E}[Y_1Y_2]&=\int_0^\infty\int_0^\infty\mathbb{P}(Y_1>s,\,Y_2>t)\,\mathrm{d}s\,\mathrm{d}t=\int_0^\infty\int_0^\infty\mathrm{e}^{-\gamma(s)\,\xi_1-\gamma(t)\,\xi_2-\gamma(s\vee t)\xi_{1,2}}\,\mathrm{d}s\,\mathrm{d}t\\&=\int_0^\infty\int_0^\infty\mathrm{e}^{-\gamma(t)\,\psi_1(1)-\gamma(s)\,\psi_2(1)+\gamma(s\wedge t)\,\kappa(\theta)}\,\mathrm{d}s\,\mathrm{d}t\,.\end{aligned}$$

The expression in (4.5) now easily follows. $\square$

A7. Proof of Proposition 7. In order to prove Proposition 7 we adopt a technique similar to the one exploited in Lijoi, Prünster and Walker (2008). Firstly, we need to introduce a preliminary lemma.

LEMMA A.1. *Let (μ_1,μ_2) be a bivariate completely random measure and suppose that μ_1*

and μ_2 are not independent. Let the Lévy intensity $\nu_t(x_1, x_2)$ of (μ_1, μ_2) be differentiable with respect to t on $\mathbb{R}^+$. If s_1 and s_2 are two integers such that $\max\{s_1, s_2\} \geq 1$ *and r_1, r_2 are two non–negative numbers with* $\min\{r_1, r_2\} \geq 1$, *then*

$$\mathrm{E}\left[\mathrm{e}^{-r_1\mu_1(A_\varepsilon)-r_2\mu_2(A_\varepsilon)}\left(1-\mathrm{e}^{-\mu_1(A_\varepsilon)}\right)^{s_1}\left(1-\mathrm{e}^{-\mu_2(A_\varepsilon)}\right)^{s_2}\right]$$
$$= \varepsilon \int_{(\mathbb{R}^+)^2} \mathrm{e}^{-r_1 x_1 - r_2 x_2}(1-\mathrm{e}^{-x_1})^{s_1}(1-\mathrm{e}^{-x_2})^{s_2}\nu'_{t_0}(x_1,x_2)\,\mathrm{d}x_1\mathrm{d}x_2 + o(\varepsilon)$$

as $\varepsilon \downarrow 0$, where $A_\varepsilon = \{t > 0 :\ t_0 - \varepsilon < t \leq t_0\}$.

Proof. Note that the left–hand–side above can be rewritten as

$$\sum_{j_1=0}^{s_1}\sum_{j_2=0}^{s_2}\binom{s_1}{j_1}\binom{s_2}{j_2}(-1)^{j_1+j_2}\,\mathrm{e}^{-\psi_{t_0}(r_1+j_1,r_2+j_2)+\psi_{t_0-\varepsilon}(r_1+j_1,r_2+j_2)}$$
$$= \mathrm{e}^{-\psi_{t_0}(r_1,r_2)+\psi_{t_0-\varepsilon}(r_1,r_2)} + \mathrm{e}^{-\psi_{t_0}(r_1,r_2)+\psi_{t_0-\varepsilon}(r_1,r_2)} \times$$
$$\times\left\{\sum_{j_1=1}^{s_1}\sum_{j_2=1}^{s_2}\binom{s_1}{j_1}\binom{s_2}{j_2}(-1)^{j_1+j_2}\ \exp\left\{-\Delta_{t_0-\varepsilon}^{t_0}\left[\psi_t(r_1+j_1,r_2+j_2)-\psi_t(r_1,r_2)\right]\right\}\right.$$
$$+\sum_{j_1=1}^{s_1}\binom{s_1}{j_1}(-1)^{j_1}\ \exp\left\{-\Delta_{t_0-\varepsilon}^{t_0}\left[\psi_t(r_1+j_1,r_2)-\psi_t(r_1,r_2)\right]\right\}$$
$$\left.+\sum_{j_2=1}^{s_2}\binom{s_2}{j_2}(-1)^{j_2}\ \exp\left\{-\Delta_{t_0-\varepsilon}^{t_0}\left[\psi_t(r_1,r_2+j_2)-\psi_t(r_1,r_2)\right]\right\}\right\}$$

where $\psi_t(\lambda_1, \lambda_2)$ is given in Equation (2.2) and $\Delta_{t_0-\varepsilon}^{t_0}\psi_t = \psi_{t_0} - \psi_{to-\varepsilon}$. Note now that

$$\Delta_{t_0-\varepsilon}^{t_0}\left[\psi_t(r_1+j_1,r_2+j_2)-\psi_t(r_1,r_2)\right] =$$
$$= \int \mathrm{e}^{-r_1x_1-r_2x_2}(1-\mathrm{e}^{-j_1x_1-j_2x_2})\left(\nu_{t_0+\varepsilon}(x_1,x_2)-\nu_{t_0}(x_1,x_2)\right)\mathrm{d}x_1\mathrm{d}x_2$$

and that as $\varepsilon \downarrow 0$:

$$\exp\left\{-\Delta_{t_0-\varepsilon}^{t_0}\left[\psi_t(r_1+j_1,r_2+j_2)-\psi_t(r_1,r_2)\right]\right\} =$$
$$= 1 - \varepsilon\left[\int \mathrm{e}^{-r_1x_1-r_2x_2}(1-\mathrm{e}^{-j_1x_1-j_2x_2})\,\nu'_{t_0}(x_1,x_2)\,\mathrm{d}x_1\mathrm{d}x_2\right] + o(\varepsilon)\ .$$

Furthermore we have $\sum_{i=1}^{s}\binom{s}{i}(-1)^i(1-\mathrm{e}^{-ix}) = -(1-\mathrm{e}^{-x})^s$ and

$$\sum_{j_1=1}^{s_1}\sum_{j_2=1}^{s_2}\binom{s_1}{j_1}\binom{s_2}{j_2}(-1)^{j_1+j_2}(1-\mathrm{e}^{-j_1x_1-j_2x_2}) =$$
$$= (1-\mathrm{e}^{-x_1})^{s_1} + (1-\mathrm{e}^{-x_2})^{s_2} - (1-\mathrm{e}^{-x_1})^{s_1}(1-\mathrm{e}^{-x_2})^{s_2}. \tag{8.1}$$

This yields, as $\varepsilon \downarrow 0$, the desired result. $\square$

Note that the case of independence between μ_1 and μ_2 can be included into the statement of Lemma A.1. The result would be slightly modified and one has

$$\begin{aligned} \mathrm{E}\left[\mathrm{e}^{-r_1\mu_1(A_\varepsilon)-r_2\mu_2(A_\varepsilon)}\left(1-\mathrm{e}^{-\mu_1(A_\varepsilon)}\right)^{s_1}\left(1-\mathrm{e}^{-\mu_2(A_\varepsilon)}\right)^{s_2}\right] = \varepsilon^2 \times \\ \times \left(\int_{\mathbb{R}^+} \mathrm{e}^{-r_1x_1}(1-\mathrm{e}^{-x_1})^{s_1}\nu'_{t_0}(x_1)\,\mathrm{d}x_1\right)\left(\int_{\mathbb{R}^+} \mathrm{e}^{-r_2x_2}(1-\mathrm{e}^{-x_2})^{s_2}\nu'_{t_0}(x_2)\,\mathrm{d}x_2\right) + o(\varepsilon^2) \end{aligned}$$

as $\varepsilon \downarrow 0$, where $A_\varepsilon = \{t > 0 :\ t_0 - \varepsilon < t \le t_0\}$.

Define, now, the set

$$\Gamma_{n,\varepsilon} := \bigcap_{i=1}^{2}\bigcap_{j=1}^{k}\left\{(t_1^{(i)},\Delta_1^{(i)},\ldots,t_{n_i}^{(i)},\Delta_{n_i}^{(i)}) :\ \kappa_i(A_{j,\varepsilon}) = n_{j,i},\ \kappa_i^c(\{T_{(j)}\}) = n_{j,i}^c\right\}$$

where $A_{j,\varepsilon} = (T_{(j)} - \varepsilon, T_{(j)}]$. The value of ε is chosen in such a way that the sets $A_{j,\varepsilon}$ are pairwise disjoint. It follows from the partial exchangeability of two samples $Y^{(1)}, Y^{(2)}$ that, in order to establish a description of the posterior distribution of (μ_1,μ_2), given data $\boldsymbol{D}$, we have to evaluate

$$\mathrm{E}\left[\mathrm{e}^{-\lambda_1\,\mu_1(0,t]-\lambda_2\,\mu_2(0,t]}\,\middle|\,\boldsymbol{D}\right] = \lim_{\varepsilon\downarrow 0}\frac{\mathrm{E}\left[\mathrm{e}^{-\lambda_1\mu_1(0,t]-\lambda_2\,\mu_2(0,t]}\,\mathbb{1}_{\Gamma_{n,\varepsilon}}(\boldsymbol{D})\right]}{\mathbb{P}\left[\boldsymbol{D}\in\Gamma_{n,\varepsilon}\right]}\,. \tag{8.2}$$

We can now prove the main result in Proposition 7.

Proof of Proposition 7. As anticipated, the proof consists in the determination of the posterior Laplace transform of $(\mu_1(0,t],\mu_2(0,t])$. As for the numerator in (8.2), one has that it coincides with the expected value of the following quantity

$$\begin{aligned} &\mathrm{e}^{-\lambda_1\,\mu_1(0,t]-\lambda_2\,\mu_2(0,t]}\prod_{j=1}^{k}\mathrm{e}^{-n_{j,1}^c\mu_1(0,T_{(j)}]-n_{j,2}^c\mu_2(0,T_{(j)}]}\prod_{i=1}^{2}\left(\mathrm{e}^{-\mu_i(0,T_{(j)}-\varepsilon]}-\mathrm{e}^{-\mu_i(0,T_{(j)}]}\right)^{n_{j,i}} \\ &= \mathrm{e}^{-\lambda_1\,\mu_1(0,t]-\lambda_2\,\mu_2(0,t]}\prod_{j=1}^{k}\mathrm{e}^{-n_{j,1}^c\mu_1(0,T_{(j)}]-n_{j,2}^c\mu_2(0,T_{(j)}]}\times\prod_{i=1}^{2}\mathrm{e}^{-n_{j,i}\mu_i(0,T_{(j)}-\varepsilon]}\left(1-\mathrm{e}^{-\mu_i(A_{j,\varepsilon})}\right)^{n_{j,i}}\,. \end{aligned}$$

If we suppose that $t \in [T_{(l)}, T_{(l+1)})$, then $\mu_i(0,t] = \sum_{j=1}^{l}\{\mu_i(A_{j,\varepsilon}) + \mu_i(C_j)\} + \mu_i(T_{(l)},t]$ where $C_j = (T_{(j-1)}, T_{(j)} - \varepsilon]$, for any $j \in \{1,\ldots,k\}$, provided that $T_{(0)} \equiv 0$. Moreover,

$$\mu_i(0,T_{(j)}] = \sum_{r=1}^{j}\mu_i(A_{r,\varepsilon}) + \sum_{r=1}^{j}\mu_i(C_r)$$

$\mu_i(0,T_{(1)}-\varepsilon] = \mu_i(C_1)$ and $\mu_i(0,T_{(j)}-\varepsilon] = \sum_{r=1}^{j-1}\mu_i(A_{r,\varepsilon}) + \sum_{r=1}^{j}\mu_i(C_r)$, for $j \ge 2$. These also imply that

$$\sum_{j=1}^{k} n_{j,i}^{c}\mu_i(0,T_{(j)}] = \sum_{j=1}^{k} \tilde n_{j,i}^{c}\mu_i(A_{j,\varepsilon}) + \sum_{j=1}^{k} \tilde n_{j,i}^{c}\mu_i(C_j)$$

and

$$\sum_{j=1}^{k} n_{j,i}\mu_i(0,T_{(j)}-\varepsilon] = \sum_{j=1}^{k-1} \bar n_{j+1,i}\mu_i(A_{j,\varepsilon}) + \sum_{j=1}^{k} \bar n_{j,i}\,\mu_i(C_j)\ .$$

If we define $C'_\varepsilon = \mathbb{R}^+ \setminus (\cup_{j=1}^k A_{j,\varepsilon})$, it is easily seen that

$$\mathrm{E}\left[\mathrm{e}^{-\lambda_1\mu_1(0,t]-\lambda_2\,\mu_2(0,t]}\ \mathbb{1}_{\Gamma_{n,\varepsilon}}(\{(T_j^{(i)},\Delta_j^{(i)})\}_{j=1,\dots,n_i;i=1,2}\right] = \mathrm{E}[I_{1,\varepsilon}]\ \mathrm{E}[I_{2,\varepsilon}]$$

where

$$I_{1,\varepsilon} = \prod_{j=1}^{k}\prod_{i=1}^{2} \mathrm{e}^{-[\lambda_i\,\mathbb{1}_{(0,t]}(T_{(j)})+\tilde n_{j,i}^{c}+\bar n_{j+1,i}]\,\mu_i(A_{j,\varepsilon})}\ \left(1-\mathrm{e}^{-\mu_i(A_{j,\varepsilon})}\right)^{n_{j,i}}$$

(with $\bar n_{k+1,i}=0$) and

$$I_{2,\varepsilon} = \prod_{i=1}^{2}\ \mathrm{e}^{-\int_{C'_\varepsilon}[\lambda_i\,\mathbb{1}_{(0,t]}(s)+\tilde N_i^c(s)+\bar N_i(s)]\,\mu_i(\mathrm{d}s)}\ .$$

The independence of the increments yields

$$\mathrm{E}[I_{1,\varepsilon}] = \prod_{j=1}^{k}\mathrm{E}\left[\prod_{i=1}^{2}\ \mathrm{e}^{-[\lambda_i\,\mathbb{1}_{(0,t]}(T_{(j)})+\tilde n_{j,i}^{c}+\bar n_{j+1,i}]\,\mu_i(A_{j,\varepsilon})}\left(1-\mathrm{e}^{-\mu_i(A_{j,\varepsilon})}\right)^{n_{j,i}}\right]\ .$$

In order to simplify the notation that is going to appear in the sequel, let $\zeta(\boldsymbol{x},\boldsymbol{n}_j) := \prod_{i=1}^{2}(1-\mathrm{e}^{-x_i})^{n_{j,i}}$, where $\boldsymbol{x}=(x_1,x_2)\in(\mathbb{R}^+)^2$ and $\boldsymbol{n}_j=(n_{j,1},n_{j,2})$ is vector of non–negative integers. If $\mathcal{I}=\{j\ :\ T_{(j)}$ is an exact observation$\}$, for any $j\in\mathcal{I}$ one has $\max\{n_{j,1},n_{j,2}\}\ge 1$ and Lemma A.1 applies, i.e., as $\varepsilon\downarrow 0$

$$\begin{aligned}&\mathrm{E}\left[\prod_{i=1}^{2}\ \mathrm{e}^{-[\lambda_i\,\mathbb{1}_{(0,t]}(T_{(j)})+\tilde n_{j,i}^{c}+\bar n_{j+1,i}]\,\mu_i(A_{j,\varepsilon})}\left(1-\mathrm{e}^{-\mu_i(A_{j,\varepsilon})}\right)^{n_{j,i}}\right]\\ &\qquad = \varepsilon\int_{(\mathbb{R}^+)^2}\mathrm{e}^{-\sum_{i=1}^{2}(\lambda_i\mathbb{1}_{(0,t]}(T_{(j)})+\tilde n_{j,i}^{c}+\bar n_{j+1,i})x_i}\,\zeta(\boldsymbol{x},\boldsymbol{n}_j)\,\nu'_{T_{(j)}}(x_1,x_2)\,\mathrm{d}x_1\mathrm{d}x_2+o(\varepsilon)\end{aligned}$$

If $j\notin\mathcal{I}$, then $n_{j,i}=0$ and the continuity of $\nu_t(x,y)$ implies

$$\begin{aligned}&\lim_{\varepsilon\downarrow 0}\ \mathrm{E}\left[\prod_{i=1}^{2}\ \mathrm{e}^{-[\lambda_i\,\mathbb{1}_{(0,t]}(T_{(j)})+\tilde n_{j,i}^{c}+\bar n_{j+1,i}]\,\mu_i(A_{j,\varepsilon})}\left(1-\mathrm{e}^{-\mu_i(A_{j,\varepsilon})}\right)^{n_{j,i}}\right] =\\ &\qquad\qquad = \lim_{\varepsilon\downarrow 0}\ \mathrm{E}\left[\prod_{i=1}^{2}\ \mathrm{e}^{-[\lambda_i\,\mathbb{1}_{(0,t]}(T_{(j)})+\tilde n_{j,i}^{c}+\bar n_{j+1,i}]\,\mu_i(A_{j,\varepsilon})}\right] = 1\ .\end{aligned}$$

Reasoning along the same line, it is immediate that

$$\mathrm{E}\left[I_{2,\varepsilon}\right] = \mathrm{e}^{-\int_{C'_\varepsilon} \psi_s\left(\lambda_1 \mathbb{1}_{(0,t]}+\tilde N_1^c+\bar N_1\,,\,\lambda_2 \mathbb{1}_{(0,t]}+\tilde N_2^c+\bar N_2\right)\,\mathrm{d}s} \to \mathrm{e}^{-\int_{\mathbb{R}^+} \psi_s\left(\lambda_1 \mathbb{1}_{(0,t]}+\tilde N_1^c+\bar N_1\,,\,\lambda_2 \mathbb{1}_{(0,t]}+\tilde N_2^c+\bar N_2\right)\,\mathrm{d}s}.$$

As far as the denominator in (8.2), this follows along the same lines and it turns out to be equal to

$$\mathrm{e}^{-\int_{C'_\varepsilon} \psi_s\left(\tilde N_1^c+\bar N_1\,,\,\tilde N_2^c+\bar N_2\,\right)\,\mathrm{d}s}\times$$
$$\times\ \varepsilon^{k_e}\prod_{j\in\mathcal{I}}\left\{\int_{(\mathbb{R}^+)^2}\mathrm{e}^{-\sum_{i=1}^2(\tilde n_{j,i}^c+\bar n_{j+1,i})x_i}\,\zeta(\boldsymbol{x},\boldsymbol{n}_j)\,\nu'_{T_{(j)}}(x_1,x_2)\mathrm{d}x_1\mathrm{d}x_2+o(1)\right\}$$

where k_e denotes the total number of exact (distinct) observations in the sample. Hence, if one considers the ratio of the two terms we have just determined and let ε tend to 0, one obtains that the posterior Laplace transform in (8.2) is given by

$$\mathrm{e}^{-\int_0^\infty\left[\psi_s\left(\lambda_1\mathbb{1}_{(0,t]}+\tilde N_1^c+\bar N_1,\,\lambda_2\mathbb{1}_{(0,t]}+\tilde N_2^c+\bar N_2\right)-\psi_s\left(\tilde N_1^c+\bar N_1,\,\tilde N_2^c+\bar N_2\right)\right]\,\mathrm{d}s}\times$$
$$\times\ \prod_{j\in\mathcal{I}}\frac{\int_{(\mathbb{R}^+)^2}\mathrm{e}^{-\sum_{i=1}^2(\lambda_i\mathbb{1}_{(0,t]}(T_{(j)})+\tilde n_{j,i}^c+\bar n_{j+1,i})x_i}\,\zeta(\boldsymbol{x},\boldsymbol{n}_j)\,\nu'_{T_{(j)}}(x_1,x_2)\mathrm{d}x_1\mathrm{d}x_2}{\int_{(\mathbb{R}^+)^2}\mathrm{e}^{-\sum_{i=1}^2(\tilde n_{j,i}^c+\bar n_{j+1,i})x_i}\,\zeta(\boldsymbol{x},\boldsymbol{n}_j)\,\nu'_{T_{(j)}}(x_1,x_2)\mathrm{d}x_1\mathrm{d}x_2}$$

and this proves the statement. □

References

Basu, B. (1971). Bivariate failure rate. *J. Amer. Statist. Assoc.* **66**, 103–104.

Bulla, P., Muliere, P. and Walker, S.G. (2007). Bayesian nonparametric estimation of a bivariate survival function. *Statistica Sinica* **17**, 427–444.

Cont, R. and Tankov, P. (2004). *Financial modelling with jump processes.* Chapman & Hall/CRC, Boca Raton, FL.

De Blasi, P., Peccati, G. and Prünster, I. (2009). Asymptotics for posterior hazards. To appear in *Ann. Statist.*

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

Dey, J., Erickson, R.V. and Ramamoorthi, R.V. (2003). Some aspects of neutral to the right priors. *Internat. Statist. Rev.* **71**, 383–401.

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

Draghici, L. and Ramamoorthi, R.V. (2003). Consistency of Dykstra–Laud priors. *Sankhya* **65**, 464-481.

Dunson, D.B., Xue, Y. and Carin, L. (2008). The matrix stick-breaking process: flexible Bayes meta-analysis. *J. Amer. Statist. Assoc.* **103**, 317–327.

Dunson, D.B. and Park, J.-H. (2008). Kernel stick-breaking processes *Biometrika* **95**, 307–323.

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

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

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

Ghosh, J.K., Hjort, N.L., Messan, C. and Ramamoorthi, R.V. (2006). Bayesian bivariate survival estimation. *J. Statist. Plann. Inference* **136**, 2297–2308.

Gill, R.D. and Johansen, S. (1990). A survey of product-integration with a view toward application in survival analysis. *Ann. Statist.* **18**, 1501–1555.

Griffin, J.E. and Steel, M.F.J. (2006). Order-based dependent Dirichlet processes. *J. Amer. Statist. Assoc.* **101**, 179–194.

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

Ishwaran, H. and Zarepour M. (2008). Series representations for multivariate generalized gamma processes via a scale invariance principle. To appear in *Statistica Sinica.*

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

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

Kim, Y. and Lee, J. (2001). On posterior consistency of survival models. *Ann. Statist.* **29**, 666-686.

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

Lijoi, A., Prünster, I. and Walker, S.G. (2008). Posterior analysis for some classes of nonparametric models. *J. Nonpar. Statist.* **20**, 447-457.

Lin, D.Y. and Ying, Z. (1993). A simple nonparametric estimator of the bivariate survival function under univariate censoring. *Biometrika* **80**, 573–581.

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

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

MacEachern, S.N. (2000). Dependent Dirichlet processes. *Technical Report.* Department of Statistics, Ohio State University.

MacEachern, S.N. (2001). Decision theoretic aspects of dependent nonparametric processes. In *Bayesian methods with applications to science, policy and official statistics* (E. George, Ed.), 551–560. International Society for Bayesian Analysis, Crete.

Nieto-Barajas, L.E. and Walker, S.G. (2007). A Bayesian semi-parametric bivariate failure time model. *Computat. Statist. Data Anal.* **11**, 529–543.

Rodríguez, A. Dunson, D. and Gelfand, A. (2008). The nested Dirichlet process. *J. Amer. Statist. Assoc.* **103**, 1131–1144.

Tankov, P. (2003). Dependence structure of spectrally positive multidimensional Lévy processes. *Unpublished Manuscript.*

Walker, S.G. and Muliere, P. (1997). Beta-Stacy processes and a generalization of the Pólya-urn scheme. *Ann. Statist.* **25**, 1762–1780.

Walker, S.G. and Muliere, P. (2003). A bivariate Dirichlet process. *Statist. Probab. Lett.* **64**, 1–7.

Woolson, R.F. and Lachenbruch, P.A. (1980). Rank tests for censored matched pairs, *Biometrika* **67**, 597–606.

Dipartimento Matematica "F. Brioschi", Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy.

E-mail: ilenia.epifani@polimi.it

Dipartimento di Economia Politica e Metodi Quantitativi, Università di Pavia, via San Felice 5, 27100 Pavia, Italy.

E-mail: lijoi@unipv.it