Scuola di Dottorato
Università degli Studi di Milano-Bicocca

Department of

Economics, Management and Statistics

Ph.D. program: **Economics and Statistics** Cycle: **XXXVI°**
Curriculum: **Statistics**

Bayesian Nonparametric Models for Survival Analysis

Surname: **Cogo**
Name: **Riccardo**
Registration number: **868839**

Supervisor: Prof. **Federico Camerlenghi**
Co-Supervisor: Prof. **Tommaso Rigon**
Tutor: Prof. **Matteo Pelagatti**

Academic Year: 2023-2024

Abstract

The aim of this thesis is to explore innovative methods in Bayesian nonparametrics (BNP) for modeling observations in the context of survival analysis. To achieve this, the goal is to leverage the flexibility of BNP methods to address some issues present in the literature on this subject, while also considering the practical implementability of the proposed models to ensure they are applicable in real-case scenarios. This manuscript is a collection of two projects, each of which explores different BNP tools to address two distinct problems in survival analysis.

The manuscript is organized around two distinct research questions. In particular, in Chapter 2, we focus on the study of survival datasets in contexts where it is reasonable to assume a strictly positive probability of recovery, called *cure rate*. Such datasets are usually modeled using the so-called cure rate models, for which various examples exist in both the frequentist and Bayesian literature. We align ourselves with the BNP framework by proposing the use of a class of processes typically employed in feature sampling problems, called *stable-beta scaled processes*, which prove to be particularly suitable for survival analysis. We will demonstrate how they are interesting both in terms of theoretical properties and applicability in real-case scenarios. Chapter 3, on the other hand, focuses on datasets consisting of different groups of survival times, raising the question of whether it is more advantageous to model each group of observations independently each others or if it is more reasonable to use a single model that estimates the survival of each group simultaneously. We adopt the latter approach, introducing a novel BNP hierarchical model that accounts for the survival of each group of observations while considering potential correlations between them. This approach allows the estimation of each group to improve the estimation of all others, following the principle of *borrowing of information*, which is well-known in the BNP literature.

In summary, this thesis aims to provide two examples of how BNP methods allow addressing research questions using similar sets of theoretical tools, highlighting their potential in terms of depth and flexibility.

Listen, for the party scene I'll give you a nice jam of light, I'll lay it all out, okay? I'm not going to do anything fancy. That way, wherever you go, it'll look good.

D. Patanè

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Bayesian Nonparametrics (BNP) and Survival Analysis have developed independently within the statistical field, but in recent decades the application of BNP methods for inference on censored survival times has been studied extensively in the literature. The aim of this chapter is to summarize some basic concepts of BNP and survival analysis, with particular focus on tools that will be used in the next chapters. Specifically, Section 1.1 is dedicated to introducing the main assumptions of BNP and some classical results. In Section 1.2, the main concepts of survival analysis are summarized, with particular attention to classical models for inference. Since the next chapters are dedicated to the introduction of BNP models for survival analysis, a summary of the main tools used in the literature is provided in Section 1.3. In Section 1.1.1 the concepts of exchangeability and partial exchangeability, which are the basis of Bayesian models, will be summarized. Then, the main methods for building nonparametric priors and some classical examples will be reviewed in Section 1.1.2. Finally, in Section 1.1.3 we will briefly introduce hierarchical models in the case of partially exchangeable data.

1.1 Bayesian Nonparametrics

Even in the frequentist framework, the nonparametric (or semiparametric) approach is generally preferred since it allows to avoid arbitrary assumptions about the parametric model, which can even be unverifiable. Therefore, while the preference for Bayesian methods over the frequentist ones can be traced back to the coherency of the Bayesian framework, the choice of developing *Bayesian Nonparametric (or semiparametric) models* is motivated by the same reasoning. The aim of this section is to introduce the Bayesian Nonparametric framework, its modeling assumptions and our main tools. Note how this topic is addressed in a complete and exhaustive manner in Ghosal and van der Vaart (2017) and Regazzini (1996), to which we refer for further information.

1.1.1 Exchangeability and partial exchangeability

The standard Bayesian approach relies on the exchangeability assumption of the observations. Let us therefore consider a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, a measurable space $\mathbb{X}_0$ equipped with its Borel σ-algebra $\mathcal{X}_0$ and an infinite sequence of observations $(X_i)_{i\geq 1}$ defined on the common space $(\Omega, \mathcal{A})$ and taking values in $(\mathbb{X}_0, \mathcal{X}_0)$. The usual assumption on $\mathbb{X}_0$ is that it is a Polish space, i.e., a separable and completely metrizable topological space. Typically, the assumption on the data is that they are exchangeable: in particular, the infinite sequence $(X_i)_{i\geq 1}$ is *exchangeable* if its law is invariant under finite permutations of its elements. Let us summarize the concept in the following definition.

Definition 1.1.1 (Exchangeability)**.** A sequence of observations $(X_i)_{i\geq 1}$ is *exchangeable* if and only if

$$(X_1, \dots, X_n) \overset{\mathrm{d}}{=} (X_{\pi(1)}, \dots, X_{\pi(n)}),$$

where π is a permutation of the set $\{1, \dots, n\}$ for any $n \geq 1$.

Exchangeability is basically an assumption of homogeneity within the observations, in a mathematical sense, which is quite reasonable in a variety of contexts with different datasets, since it can be seen as an irrelevance of the order in which the observations are recorded. The well known de Finetti's representation theorem, introduced in Finetti (1937), states the equivalence between exchangeability and conditional independence and identically distribution of infinite sequences. The main upside of the nonparametric approach is that in this framework the distribution of the observations is completely unknown. Hence, the "parameter space" is the space $\boldsymbol{P}$ of all the probability measures on $(\mathbb{X}_0, \mathcal{X}_0)$, while the law of the data is the *random probability measure* $\tilde{p}$. In order to precisely define this notion, note first that we are interested in the probability distribution of the whole sequence $X := (X_i)_{i\geq 1}$, so if we set $\mathbb{X} := \mathbb{X}_0^\infty$, we can choose the product σ-algebra $\mathcal{X}$ on $\mathbb{X}$, which is the σ-algebra generated by all the sets

$$A = A_1 \times \ldots \times A_n \times \mathbb{X}_0 \times \ldots,$$

where A_i is a measurable element of $\mathcal{X}$ for each $i = 1, \dots, n$ and $n \geq 1$. Under this notation, the sequence X can be seen as the random element

$$X : (\Omega, \mathcal{A}) \to (\mathbb{X}, \mathcal{X}),$$

where X is measurable with respect to the corresponding σ-algebra. In order to define a σ-algebra over the space $\boldsymbol{P}$, i.e., over the space of all the probability measures on $(\mathbb{X}_0, \mathcal{X}_0)$, for any $A \in \mathcal{X}_0$ let us define the map

$$\begin{aligned} \nu_A : \boldsymbol{P} &\to [0,1] \\ p &\mapsto p(A), \end{aligned}$$

that is a function of p for each A. Let us call $\mathcal{P}$ the σ-algebra on $\boldsymbol{P}$ generated by the family of functions $(\nu_A)_{A\in\mathcal{X}_0}$. We now introduce the following definition.

Definition 1.1.2 (Random probability measure)**.** A measurable map

$$\tilde{p} : (\Omega, \mathcal{A}) \to (\boldsymbol{P}, \mathcal{P})$$

is called *random probability measure* on $(\mathbb{X}_0, \mathcal{X}_0)$.

Note how a random probability measure is a function of two variables. In particular, for any element $\omega \in \Omega$, $\tilde{p}(\omega, \cdot) \in \boldsymbol{P}$ is a probability measure on $(\mathbb{X}_0, \mathcal{X}_0)$. On the other hand, for any event $A \in \mathcal{X}_0$, $\tilde{p}(\cdot, A)$ is a random variable: it follows that the random probability measure $\tilde{p}$ can be seen as a stochastic process indexed by the events of the σ-algebra $\mathcal{X}_0$ instead of being indexed by time. This is why in BNP literature, specific random measures are often referred to as *processes*. Note also that considering the probability measure $\mathbb{P}$ on the space $(\Omega, \mathcal{A})$, a random measure $\tilde{p}$ has its own probability distribution $\gamma = \mathbb{P} \circ \tilde{p}^{-1}$ on $(\boldsymbol{P}, \mathcal{P})$. Let us write $\tilde{p} \sim \gamma$.
In the Bayesian parametric framework the conditional independence of the data induces exchangeability. In the nonparametric framework, if we consider a random probability measure $\tilde{p}$ with distribution γ and the sequence of observations X_is, assuming that the observations are i.i.d. with distribution $\tilde{p}$ conditionally on $\tilde{p}$ means that

$$\mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) = \int_{\boldsymbol{P}} \prod_{i=1}^{n} p(A_i) \gamma(\mathrm{d}p)$$

as $n \geq 1$. In short, let us write

$$X_i \mid \tilde{p} \overset{\text{i.i.d.}}{\sim} \tilde{p}, \quad i \geq 1.$$

While in the parametric context we talk about independence of the observations conditionally on a parameter, here we refer to the previous expression saying that the distribution of the observations X is a *mixture of laws of a sequence of i.i.d. random variables*, and we refer to γ as the *mixing measure*. Similarly to what happens in the parametric case, this implies the exchangeability of the data. In fact, let us consider a probability measure γ

on $(\boldsymbol{P}, \mathcal{P})$, i.e., γ is the distribution of a random probability measure $\tilde{p}$ on $(\mathbb{X}_0, \mathcal{X}_0)$. Let us denote by p^∞ the probability distribution of the sequence of observations X; then X is exchangeable and it has law

$$\mathbb{P}(X \in A) = \int_{\boldsymbol{P}} p^\infty(A)\gamma(\mathrm{d}p),$$

where and $A \in \mathcal{X}$. Indeed, $\mathcal{X}$ is a σ-algebra generated by sets of the form

$$A = A_1 \times \ldots \times A_n \times \mathbb{X}_0 \times \ldots,$$

and the previous equality for such sets holds true since

$$\begin{aligned} \mathbb{P}(X \in A) = \mathbb{P}(X_1 \in A_1, \ldots, X_n \in A_n) = \int_{\boldsymbol{P}} \prod_{i=1}^n p(A_i)\gamma(\mathrm{d}p) = \\ = \mathbb{P}(X_{\pi(1)} \in A_1, \ldots, X_{\pi(n)} \in A_n), \end{aligned}$$

for any permutation π of $\{1, \ldots, n\}$. Moreover, the de Finetti representation theorem guarantees the reverse implication, i.e., a sequence of observations is exchangeable if and only if its distribution is a mixture of laws of a sequence of i.i.d. random variables. So the exchangeability assumption guarantees the existence of the mixing measure γ such that

$$\begin{aligned} X_i \mid \tilde{p} &\overset{\text{i.i.d.}}{\sim} \tilde{p}, \quad i \geq 1 \\ \tilde{p} &\sim \gamma, \end{aligned} \tag{1.1}$$

i.e., the observations are i.i.d. with distribution $\tilde{p}$ conditionally on $\tilde{p}$, and $\tilde{p}$ is distributed according to γ. What discussed above is summarized by the following theorem.

Theorem 1.1.1 (de Finetti). *Let $\mathbb{X}_0$ be a Polish space and $\mathcal{X}_0$ its Borel σ-algebra. Then, the following conditions are equivalent.*

(i.) $(X_i)_{i\geq 1}$ *is a sequence of exchangeable observations.*

(ii.) *There exist a random probability measure $\tilde{p}$ on $(\mathbb{X}_0, \mathcal{X}_0)$ such that $(X_i)_{i\geq 1}$ are i.i.d. conditionally on $\tilde{p}$.*

(iii.) *There exist a probability measure γ on $(\boldsymbol{P}, \mathcal{P})$ such that*

$$\mathbb{P}(X_1 \in A_1, \ldots, X_n \in A_n) = \int_{\boldsymbol{P}} \prod_{i=1}^n p(A_i)\gamma(\mathrm{d}p),$$

for any $A_1, \dots, A_n \in \mathcal{X}_0$ *and* $n \geq 1$.

Note that thanks to Theorem 1.1.1, the exchangeability of the observations is a sufficient assumption in order to assure the existence of the measure γ in model (1.1), which therefore can be considered as a prior distribution.
On the other hand, in numerous applications the exchangeability assumption can be too restrictive, in particular when we deal with data coming from multiple studies that shows some relations with each other. For example, this situation can happen when the data come from the same experiment but performed under different conditions. The most appropriate assumption in these case is the *partial exchangeability*, first introduced in Finetti (1938). Data are partially exchangeable when they are organized into a finite number of groups and they are exchangeable, in the sense of Definition 1.1.1, only within each group. In particular, let us consider d groups of $\mathbb{X}_0$-valued observations, where we denote by $X_{i,j}$ the ith observation of group j, for $j = 1, \dots, d$, where N_j is the number of observations in group j; each observation $X_{i,j}$ is a $\mathbb{X}_0$-valued random element defined on the common probability space $(\Omega, \mathcal{A}, \mathbb{P})$.

Definition 1.1.3 (Partial exchangeability)**.** The sequences of observations $((X_{i,j})_{i\geq 1})_{j=1}^d$ are *partially exchangeable* if and only if

$$(X_{1,j}, \dots, X_{n,j}) \stackrel{\mathrm{d}}{=} (X_{\pi_j(1),j}, \dots, X_{\pi_j(n_j),j}),$$

for any $n_j \geq 1$ and π_j is a permutation of the set $\{1, \dots, n_j\}$, for any $j = 1, \dots, d$.

Note that the de Finetti representation theorem holds true even in this context, as summarized in the following theorem.

Theorem 1.1.2 (de Finetti)**.** *The sequences* $(X_{i,j})_{i\geq 1; j=1,\dots,d}$ *are partially exchangeable if and only if there exists a probability measure* γ_d *over* $\boldsymbol{P}^d$*, which is the* d*-dimensional product space with respect to* $\boldsymbol{P}$*, such that*

$$\mathbb{P}\left[\bigcap_{j=1}^{d}\bigcap_{i=1}^{N_j}\{X_{i,j} \in A_{i,j}\}\right] = \int_{\mathbb{P}_X^d} \prod_{j=1}^{d}\prod_{i=1}^{N_j} p_i(A_{i,j})\gamma_d(\mathrm{d}p_1, .., \mathrm{d}p_d),$$

for any $(N_1, \dots, N_d) \in \mathbb{N}^d$ *and for any collection of Borel sets* $A_{i,j} \in \mathcal{X}_0$*, as* $j = 1, \dots, d$ *and* $i = 1, \dots, N_j$.

Note again how the d-dimensional mixing measure of the previous expression works as a prior distribution. Let us refer to the measure γ and γ_d of Theorem 1.1.1 and Theorem 1.1.2 as *de Finetti measures.*

1.1.2 Nonparametric priors

The choice of an appropriate de Finetti measure for the modelling of exchangeable or partially exchangeable observations is a problem that can be addressed in different ways and it depends on the type of data and on the approach. A common way to build a random probability measure, relies on the use of finite-dimensional distributions. Let

$$\tilde{p}: (\Omega, \mathcal{A}) \to (\boldsymbol{P}, \mathcal{P})$$

be a random probability measure on $(\mathbb{X}_0, \mathcal{X}_0)$ with distribution γ and, for any ordered sample $(A_1, \ldots, A_k)$ of distinct elements from $\mathcal{X}_0$, let $\gamma_{A_1,\ldots,A_k}$ be the probability distribution of $(\tilde{p}(A_1), \ldots, \tilde{p}(A_k))$. We refer to $\gamma_{A_1,\ldots,A_k}$ as *a finite-dimensional distribution of* $\tilde{p}$. In fact, $\gamma_{A_1,\ldots,A_k}$ is a probability measure on $([0,1]^k, \mathcal{B}([0,1]^k))$, and if we define the function $\Phi_{A_1,\ldots,A_k}$ as

$$\begin{aligned} \Phi_{A_1,\ldots,A_k} : \boldsymbol{P} &\to [0,1]^k \\ p &\mapsto (p(A_1), \ldots, p(A_k)), \end{aligned}$$

the finite dimensional distribution $\gamma_{A_1,\ldots,A_k}$ can be written in therms of γ as

$$\gamma_{A_1,\ldots,A_k} = \gamma \circ \Phi^{-1}_{A_1,\ldots,A_k}.$$

Finally, let us denote by Γ the set of all finite dimensional distributions of $\tilde{p}$. The following result, discussed in Daley and Vere-Jones (2008), allows to characterize a random probability measure by means of its finite dimensional distributions.

Theorem 1.1.3. *Let $\mathbb{X}_0$ be a Polish space equipped with the Borel σ-algebra $\mathcal{X}_0$. Let Γ be a family of probability distributions as previously described. Then, the measure $\tilde{p}$ is a random probability measure if and only if the following four conditions are satisfied.*

(C1) For any $\gamma_{A_1,\ldots,A_k} \in \Gamma$ and for any permutation π of $\{1, \ldots, k\}$,

$$\gamma_{A_1,\ldots,A_k} = \gamma_{A_{\pi(1)},\ldots,A_{\pi(k)}} \circ f_\pi^{-1},$$

where $f_\pi(x_1, \ldots, x_k) = \left(x_{\pi(1)}, \ldots, x_{\pi(k)}\right)$ for any $x_1, \ldots, x_k$.

(C2) $\gamma_{\mathbb{X}_0} = \delta_{\{1\}}$.

(C3) For any $\gamma_{A_1,\ldots,A_k} \in \Gamma$, let $(B_1, \ldots, B_m)$ be any measurable partition of $\mathbb{X}_0$ that is equal

or finer with respect to the partition generated by $(A_1, \ldots, A_k)$*. Then:*

$$\gamma_{A_1,\ldots,A_k} = \gamma_{B_1,\ldots,B_m} \circ s^{-1},$$

where the function s *is defined as*

$$\begin{aligned} s : [0,1]^m &\to [0,1]^k \\ (x_1, \ldots, x_m) &\mapsto \left(\sum_{(1)} x_i, \ldots, \sum_{(k)} x_i \right), \end{aligned}$$

where the sums are performed over the set $(j) := \{i : A_j \supset B_i\}$*, i.e., over the indexes of the sets of the finer partition included in* A_j*.*

(C4) For any sequence $(A_n)_{n\geq 1}$*,* $A_n \in \mathcal{X}_0$ *such that* $A_n \searrow \emptyset$ *as* $n \to \infty$*, then* γ_{A_n} *weakly converges to* $\delta_{\{0\}}$*.*

Another common way to define random probability measures is the *stick-breaking construction*, which allows to define almost surely discrete random probability measures on $(\mathbb{X}_0, \mathcal{X}_0)$ of the type

$$\tilde{p} = \sum_{j\geq 1} \tilde{p}_j \delta_{\tilde{\theta}_j},$$

where the locations $\tilde{\theta}_j$ are i.i.d. random atoms distributed according to a probability P_0 on $(\mathbb{X}_0, \mathcal{X}_0)$, while the $\tilde{p}_j s$ forms a set of normalized positive weight, i.e., $\sum_{j\geq 1} \tilde{p}_j = 1$ almost surely. These weights are defined according to the *stick-breaking strategy* as follows. Let us consider a sequence of random variables $(Y_n)_{n\geq 1}$ in $[0,1]$; starting from a stick of length 1, it is broken in two peaces of length Y_1 and $1 - Y_1$, putting $\tilde{p}_1 = Y_1$. The residual stick of length $1 - Y_1$ is then splitted again in two sticks of length Y_2 and $1 - Y_2$, putting $\tilde{p}_2 = (1 - Y_1) \cdot Y_2$. Note that the remaining stick has length $(1 - Y_1) \cdot (1 - Y_2)$. Iterating the procedure, we obtain that the weights are equal to

$$\tilde{p}_1 = Y_1, \quad \tilde{p}_j = Y_j \prod_{i=1}^{j-1} (1 - Y_i) \quad \text{for any } j > 1.$$

Note that the weights obtained via the aforementioned stick-breaking construction are normalized thanks to the following result, discussed in Ghosal and van der Vaart (2017).

Theorem 1.1.4. *Let us consider an infinite sequence of i.i.d. random variables* $(Y_n)_{n\geq 1}$ *in* $[0,1]$*. Then* $\sum_{j\geq 1} \tilde{p}_j = 1$ *almost surely if and only if* $\mathbb{P}(Y_1 > 0) > 0$*.*

Note how this condition assure that the random measure $\tilde{p}$ obtained via the stick-breaking construction is a probability measure.
While the construction of nonparametric priors can be performed in different ways, the main issue is the possibility of investigating the posterior distribution. In general, the posterior distribution of a nonparametric prior is unknown, and since it is a infinite-dimensional object, even its approximation is a tricky task. In this sense, a historically important step forward in the theory of BNP has been made by introducing a first example of a tractable prior, both in the sense of posterior distribution and in simplicity of construction. This is the case of the Ferguson-Dirichlet process, introduced in Ferguson (1973). This process can be defined in different ways. Here we report the constructions obtained according to the aforementioned methods, i.e., the finite-dimensional distributions and the stick-breaking methods. Let us therefore consider a finite measure α on a Polish space with its Borel σ-algebra, $(\mathbb{X}_0, \mathcal{X}_0)$, with $\alpha(\mathbb{X}_0) = a < \infty$. Let us consider a set of elements $A_1 \dots, A_k$ in $\mathcal{X}_0$ and let us define the finite-dimensional distribution $\gamma_{A_1,\dots,A_k}$ as

$$\gamma_{A_1,\dots,A_k} = \gamma_{B_1,\dots,B_m} \circ s_k^{-1},$$

where $(B_1, \dots, B_m)$ is the partition generated by $(A_1, \dots, A_k)$ in the sense that

$$A_i = \bigcup_{(i)} B_j, \quad \text{where } (i) := \{j : B_j \subseteq A_i\},$$

s_k is defined as

$$s_k(x_1, \dots, x_k) = \left(\sum_{(1)} x_i, \dots, \sum_{(k)} x_i \right)$$

and we set

$$\gamma_{B_1,\dots,B_m} = \mathcal{D}_m(\alpha(B_1), \dots, \alpha(B_m)),$$

i.e., $\gamma_{B_1,\dots,B_m}$ is the Dirichlet distribution with parameters $(\alpha(B_i))_{i=1}^k$. Finally, if $m = 1$, we set $\gamma_{\mathbb{X}_0} = \delta_{\{1\}}$. The existence of the corresponding random probability measure $\tilde{p}$ is guaranteed by Theorem 1.1.3, and it is usually called *Dirichlet process*. Let us write

$$\tilde{p} \sim \mathcal{D}_\alpha.$$

An equivalent representation of the Dirichlet process exploits the stick-breaking method described above, and was first described by Sethuraman (1994). The construction is summarized by the following theorem, which also guarantees the equivalence between the two definitions of the Dirichlet process.

Theorem 1.1.5 (Sethuraman)**.** *Let* $\tilde{\theta}_1, \tilde{\theta}_1, \ldots \overset{i.i.d.}{\sim} \frac{\alpha}{a}$*, and let* $Y_1, Y_2, \ldots \overset{i.i.d.}{\sim} Beta(1, a)$*, where the two sequences are independent. The random probability measure*

$$\tilde{p} := \sum_{j \geq 1} Y_j \prod_{i=1}^{j-1} (1 - Y_i) \delta_{\tilde{\theta}_j}$$

has distribution $\mathcal{D}_\alpha$*.*

Note that in the previous construction the locations are sampled from the measure α, which characterizes the specific Dirichlet process. Note that clear advantage of a method such as the stick-breaking is that it allows to construct random probability measures in an algorithmic way, thus also allowing an estimation of the approximation error, which is inevitable in the estimation of infinite-dimensional objects.
An appealing feature of the Dirichlet process is its conjugacy. In fact, the posterior distribution of a Dirichlet process is again a Dirichlet process with updated parameter. This result is summarized in the following theorem.

Theorem 1.1.6 (Ferguson)**.** *Let* $(X_n)_{n \geq 1}$ *be an exchangeable sequence of observations on the probability space* $(\Omega, \mathcal{A}, \mathbb{P})$*, with* $\mathcal{D}_\alpha$ *as the de Finetti measure of this sequence, namely*

$$X_i \mid \tilde{p} \overset{i.i.d.}{\sim} \tilde{p}, \quad i \geq 1$$
$$\tilde{p} \sim \mathcal{D}_\alpha,$$

where α *is a finite measure on* $(\mathbb{X}_0, \mathcal{X}_0)$*. Then the posterior distribution of* $\tilde{p}$ *is again a Dirichlet process with parameter* $\alpha_n = \alpha + \sum_{i=1}^n \delta_{X_i}$*, i.e.,*

$$\tilde{p} \mid (X_i)_{i=1}^n \sim \mathcal{D}_{\alpha_n}.$$

The BNP literature provides several random probability measures which generalizes the Dirichlet process. The most famous one is probably the *Pitman-Yor* process, introduced in Pitman and Yor (1997). There are several equivalent definition for this process; here we report the representation via a stick-breaking construction.

Definition 1.1.4 (Pitman-Yor process)**.** Let $\bar{\alpha}$ be a non-atomic probability measure on $(\mathbb{X}_0, \mathcal{X}_0)$ and let $\sigma \in (0, 1)$ and $a > -\sigma$ be two real parameters. The *two-parameter Pitman-Yor process* is a random probability measure $\tilde{p}$ on $(\mathbb{X}_0, \mathcal{X}_0)$ having the stick-breaking representation

$$\tilde{p} = \sum_{j \geq 1} \tilde{p}_j \delta_{\tilde{\theta}_j},$$

where the atoms $\tilde{\theta}_j$s are sampled from $\bar{\alpha}$ and the weights $\tilde{p}_j$s are constructed as follows:

$$\tilde{p}_1 = Y_1, \quad \tilde{p}_j = Y_j \prod_{i=1}^{j-1} (1 - Y_i) \quad \text{for any } j > 1,$$

where $Y_j \overset{\text{i.i.d.}}{\sim} \text{Beta}(1 - \sigma, a + j\sigma)$ as $j \geq 1$. Let us write

$$\tilde{p} \sim \text{PY}(\sigma, a; \bar{\alpha}).$$

Note that when $\sigma = 0$, the Pitman-Yor process reduces to a the Dirichlet process with total mass a. Moreover, it is also possibile to further generalize both the Dirichlet and the Pitman-Yor processes seeing them as part of the larger class of the so called Gibbs-type priors, which represent the most natural generalization of the Dirichlet process that is still tractable. For a review of this class of priors and its properties see De Blasi et al. (2015). The Pitman-Yor process is not conjugate, but it is quasi-conjugate, as summarized by the following theorem.

Theorem 1.1.7 (Pitman & Yor)**.** *Let $(X_n)_{n \geq 1}$ be an exchangeable sequence of observations on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$, with $PY(\sigma, a; \bar{\alpha})$ as the de Finetti measure of this sequence, namely*

$$\begin{aligned} X_i \mid \tilde{p} &\overset{i.i.d.}{\sim} \tilde{p}, \quad i \geq 1 \\ \tilde{p} &\sim PY(\sigma, a; \bar{\alpha}), \end{aligned}$$

where $\bar{\alpha}$ is a probability measure on $(\mathbb{X}_0, \mathcal{X}_0)$. Let K_n be the number of distinct values out of $X_1, \ldots, X_n$, denoted as $X_1^*, \ldots, X_{K_n}^*$, and let $n_1, \ldots, n_{K_n}$ be the corresponding frequencies. Then the posterior distribution of $\tilde{p}$ can be written as

$$\tilde{p} \mid (X_i)_{i=1}^n = \sum_{j=1}^{K_n} \tilde{W}_j \delta_{X_j^*} + \tilde{W}_{K_{n+1}} \tilde{q},$$

where the vector of random variables $(\tilde{W}_1, \ldots, \tilde{W}_{K_n+1})$ is jointly distributed as a $K_n + 1$ dimensional Dirichlet distribution with parameters $(n_1 - \sigma, \ldots, n_{K_n} - \sigma, a + K_n\sigma)$, and

$$\tilde{q} \sim PY(\sigma, a + K_n\sigma; \bar{\alpha}).$$

1.1.3 Hierarchical processes

As mentioned above, the motivation to relax the exchangeability assumption on the observations emerges naturally from applications. A classic example regards document analysis: the overall population of observed values are the words in a collection of documents, but it is reasonable to assume that each document constitutes a sub-population with its own distribution. In these cases, the partial exchangeability as defined in Definition 1.1.3 is the most suitable assumption. In fact, first of all the partial exchangeability implies a intrinsic heterogeneity of the observations, and it is consistent with a dependence assumption between groups. On the other hand, Theorem 1.1.2 assures that it is sufficient to guarantee the existence of a de Finetti measures, which encapsulate the prior opinion on the data and on the dependence between their groups. In general, an extensive literature has been developed to underline the inferential problems arising under a ill-posed exchangeability assumption in Bayesian nonparametrics. See for example the seminal works MacEachern (1999) and MacEachern (2000), or the reviews Dunson (2010) and Teh and Jordan (2010). Therefore, in a partially exchangeable framework the issue becomes the construction of an adequate de Finetti measure, which must adequately model the observations and the dependence between the groups, and at the same time be tractable, in particular to what extent the inference from its posterior distribution. A class of models widely used in the literature concerning Bayesian nonparametrics for modeling partially exchangeable data exploits *hierarchical processes* as nonparametric priors. For a given $d \geq 1$, let us consider d random probability measure $(\tilde{p}_i)_{i=1}^d$ on $(\mathbb{X}_0, \mathcal{X}_0)$ and a further random probability measure $\tilde{p}_0$ on $(\mathbb{X}_0, \mathcal{X}_0)$, independent from the first set of measures. The general structure of a d-dimensional hierarchical processes is

$$\begin{aligned} \tilde{p}_i \mid \tilde{p}_0 &\overset{\text{i.i.d.}}{\sim} \tilde{\mathcal{L}}_0, \quad i = 1, \dots, d \\ \tilde{p}_0 &\sim \mathcal{L}_0, \end{aligned} \tag{1.2}$$

where $\tilde{\mathcal{L}}_0$ is the probability distribution of each $\tilde{p}_i$ conditionally to $\tilde{p}_0$, i.e.,

$$\mathbb{E}\left[\tilde{p}_i \mid \tilde{p}_0\right] = \int_{\boldsymbol{P}} p \tilde{\mathcal{L}}_0(\mathrm{d}p),$$

while $\mathcal{L}_0$ is such that

$$\mathbb{E}\left[\tilde{p}_0\right] = \int_{\boldsymbol{P}} p \mathcal{L}_0(\mathrm{d}p) = P_0,$$

for a certain fixed and non-atomic probability measure P_0. Note that this type of process is designed to be used as nonparametric prior for a set of d-groups of partially exchangeable observations. According to this approach, each group of observations is modeled according to a different random probability measure $\tilde{p}_i$, while the dependence between the groups is modeled by the further random probability measure $\tilde{p}_0$. In fact, in general the $\tilde{p}_i$s are dependent and their marginal distribution is assumed unknown; they are i.i.d. with known distribution $\tilde{\mathcal{L}}_0$ only conditionally on $\tilde{p}_0$. Note also that a majority of literature concerning these topics considers vector of *discrete* random probability measure $(\tilde{p}_1, \dots, \tilde{p}_d)$, thus assuming the possibility of shared values within each sample; moreover, the assumption of shared atoms across the d samples entails positive probabilities of ties across the samples themselves. Assuming partially exchangeable data, the idea behind this Bayesian approach is that it is somehow preferable to model the d groups of data together rather than model each group under independent priors. In fact, the goal is to exploit the information used to model one group of observations to improve the modelling of all the other groups. This concept is called *borrowing of information*, or *borrowing of strength*, and it is a crucial point when it comes to justify the adoption of these kind of models. See for example Dunson (2009).

At a historical level, Teh et al. (2006) were the first to introduce and study an example of a tractable hierarchical process, in particular as regards its posterior distribution. This is the case of the *hierarchical Dirichlet process (HDP)*, introduced in Teh et al. (2006) as a nonparametric extension of the well known *Latent Dirichlet Allocation (LDA)* model. First introduced in Blei et al. (2003) in the field of information retrieval, LDA overcomes the classical exchangeability assumption between the observations (words) of the set of documents which constitute the dataset, also known in this context as "bag of words" assumption, assuming otherwise that each word of a document arises from a number of latent clusters, or "topics", and modeling each cluster as a multinomial probability distribution on words from some basic vocabulary. HDP provides a nonparametric generalization of LDA, in order to allow the sharing of clusters between documents in the corpora, as well as to allow for multiple corpora, thus investigating to what extent the latent topics that are shared among documents are also shared across groups of documents. The approach proposed reposes on the Dirichlet process. In particular, relying on the previous notation, a HDP is a distribution over $\boldsymbol{P}^d$ characterized by a set of random probability measure $(\tilde{p}_i)_{i=1}^d$ as in (1.2), such that

$$\begin{aligned} \tilde{p}_i \mid \tilde{p}_0 &\overset{\text{i.i.d.}}{\sim} \mathcal{D}_{\tilde{p}_0}, \quad i = 1, \dots, d \\ \tilde{p}_0 &\sim \mathcal{D}_\alpha, \end{aligned}$$

where α is a baseline probability measure. The HDP can also be constructed via a stick-breaking scheme, as

$$\tilde{p}_i = \sum_{j\geq 1} \pi_{i,j}\delta_{\tilde{\theta}_j}, \quad i = 1, \dots, d$$
$$\tilde{p}_0 = \sum_{j\geq 1} \beta_j \delta_{\tilde{\theta}_j},$$

where the locations $\tilde{\theta}_j$s are i.i.d. sampled from α and the weights $\beta := (\beta_j)_{j\geq 1}$s are such that

$$\beta_1 = Y_1, \quad \beta_j = Y_j \prod_{i=1}^{j-1}(1 - Y_i) \quad \text{for any } j > 1,$$

where $Y_j \overset{\text{i.i.d.}}{\sim} \text{Beta}(1, a)$ as $j \geq 1$, and

$$(\pi_{i,j})_{j\geq 1} \mid \beta \sim \mathcal{D}_\beta.$$

Here, both $\pi_i := (\pi_{i,j})_{j\geq 1}$ and β are interpreted as probability measures on the positive integers. The original work which introduced HDP also provides different sampling strategy in order to perform inferential tasks; please refer again to Teh et al. (2006). Note that the subsequent literature has introduced other examples of hierarchical processes, as well as investigating general representations and properties for this class of processes; see for example Camerlenghi et al. (2019).

Hierarchical processes are just one example of a nonparametric prior for partially exchangeable data, since in the literature numerous nonparametric priors to model dependent groups of observations can be found. Examples of such models are given, for instance, by the nested processes, such as the nested Dirichlet process introduced in Rodríguez et al. (2008), and the compound random measures introduced in Griffin and Leisen (2017). Another example is given by the Dependent Dirichlet Processes and their extensions: see Quintana et al. (2022) for a complete review.

1.2 Survival Analysis

Survival analysis is a branch of statistics which aims to infer useful information from datasets reporting *survival times*, i.e., times elapsed until the occurrence of an event, be it the event being analyzed or a censorship that prevents verifying whether the event occurred or not. Let us first set the notation.

Let $T_1, \ldots, T_n$ be positive random variables i.i.d. distributed according to the distribution F, and let $C_1, \ldots, C_n$ be another set of random variables distributed according to the distribution G, i.e.,

$$T_1, \ldots, T_n \overset{\text{i.i.d.}}{\sim} F, \quad \text{and}$$
$$C_1, \ldots, C_n \overset{\text{i.i.d.}}{\sim} G.$$

The first set of variables measures the time until the event, while the second set of variables measures the time until the censorship. Note the there can be different types of censorship mechanism; here we will always focus on the *right censorship*. This means that the observed values are

$$\{(X_i, \Delta_i)\}_{i=1}^n, \quad \text{where}$$

$$X_i = \min\{T_i, C_i\} \quad \text{and} \quad \Delta_i = \mathbb{1}_{(0,C_i]}(T_i) \quad \text{for each } i = 1, \ldots, n,$$

i.e., we record the time elapsed until the occurrence of the first occurrence for each subject, be it the event T_i or the censorship C_i, and a variable indicating whether the observed occurrence is the event or the censorship. Let us call the observed variables X_is *survival times*. If the ith observation is such that $\Delta_i = 1$, it is called *exact*. Let us consider for example a pharmaceutical trial where a drug is administered to a group of patients, and the time elapsed until the desired effect occurs is measured. The event may not be observed when a given censoring occurs, which may be given by the end of the observation period, the death of the patient or other mechanisms. Note that the censorship mechanism is common between the patients, but each censorship is independent from all the others.
The statistical methods developed within the context of survival analysis focus on the investigation of some objects of interest. In particular, the first object represents the probability for a patient of being still alive after a given time, i.e., of not having yet experienced the censorship, according to the following definition.

Definition 1.2.1 (Survival function)**.** Let F be the distribution of the times T_is. Then, the *survival function* is defined as

$$S(t) = 1 - F(t),$$

for each time $t \in \mathbb{R}^+$.

Let $T \sim F$. Note that, since $F(t) = \mathbb{P}(T \leq t)$ for each $t > 0$, it follows that $S(t) = \mathbb{P}(T > t)$, i.e., the function S indicates the probability for a generic subject of the considered group of being still alive after time t. Another quantity of interest is the probability of instant death. Assuming that the distribution F has a density, let us introduce the following object.

Definition 1.2.2 (Hazard rate function)**.** Let F be the distribution of the times T_is and $S(\cdot) = 1 - F(\cdot)$ be the corresponding survival function. Then, the *hazard rate function* is defined as

$$h(t) = -\frac{S'(t)}{S(t)},$$

for each time $t \in \mathbb{R}^+$.

Note that, assuming the existence of a density for the distribution F, the hazard rate is the instant probability of death at time t given the survival up to time t. In fact, let us consider a variable $T \sim F$, and let us define the probability of death at t give the survival up to t as

$$h(t)\mathrm{d}t = \mathbb{P}(T \in (t, t + \mathrm{d}t) \mid T > t),$$

and note therefore that

$$h(t)\mathrm{d}t = \frac{\mathbb{P}(T \in (t, t + \mathrm{d}t))}{\mathbb{P}(T > t)} = \frac{F'(t)\mathrm{d}t}{1 - F(t)} = -\frac{S'(t)}{S(t)}\mathrm{d}t,$$

since $S'(t) = -F'(t)$ for any t. The expression reported in Definition 1.2.2 follows. Integrating the instant rate of death $h(t)$ over time, we obtain the cumulative probability of death up to a certain moment. On the other hand, this is only possible when h exists, i.e., when F has a density. Let us therefore introduce the following definition.

Definition 1.2.3 (Cumulative hazard function)**.** Let F be the distribution of the times T_is and $S(\cdot) = 1 - F(\cdot)$ be the corresponding survival function. Then, the *cumulative hazard function* is defined as

$$H(t) = -\int_0^t \frac{\mathrm{d}S(u)}{S(u^-)},$$

for each time $t \in \mathbb{R}^+$, where

$$S(u^-) = 1 - F(u^-) = 1 - \lim_{z \searrow 0} F(u - z).$$

Note that the object defined above always exists, as well as the survival function. They are related through the integral relation reported in Definition 1.2.3, or viceversa, through a product integral relation. Let us recall that given a non-decreasing and right-continuous function

$$g : \mathbb{R}^+ \to \mathbb{R}^+,$$

a positive value $t \in \mathbb{R}^+$ and a partition of $[0, t]$

$$0 = t_0 < t_1 < \ldots < t_{m-1} < t_m = t$$

such that

$$\lim_{m\to\infty} \left(\max_{i=1,\ldots,m} \mid t_i - t_{i-1} \mid \right) = 0,$$

the *product integral* is the function

$$t \mapsto \prod_{s\in(0,t]} (1 + \mathrm{d}g(s)) := \lim_{m\to\infty} \prod_{i=1}^{m} (1 + g(t_i) - g(t_{i-1})).$$

Moreover, let g be a function defined on $\mathbb{R}^+$ as the sum of a discrete and a continuous component, as follows:

$$g(t) := \sum_{j\geq 0} w_j \mathbb{1}_{(0,t_j]}(t) + g_c(t), \quad \text{for any } t > 0,$$

where each t_j is the jth element of the aforementioned partition, $w_j < 1$ for each j and g_c the continuous part. Then the product integral can be equivalently defined as

$$t \mapsto \prod_{s\in(0,t]} (1 + \mathrm{d}g(s)) := e^{-\int_0^t \mathrm{d}g_c(t)} \cdot \prod_{j:t_j\leq t} (1 + w_j).$$

The following result explains the link between the survival function and the cumulative hazard function, thus justifying Definition 1.2.3.

Proposition 1.2.1. *Let $H : \mathbb{R}^+ \to \mathbb{R}^+$ be a non-decreasing and right-continuous function such that*

$$H(t) = \sum_{j\geq 0} w_j \mathbb{1}_{(0,t_j]}(t) + H_c(t), \quad \text{for any } t > 0,$$

where $w_j < 1$ and $H(t) < \infty$ for each $t \in \mathbb{R}^+$. Then the function

$$t \mapsto S(t) = \prod_{s\in(0,t]} (1 + \mathrm{d}H(s))$$

is the survival function associated to H and H is the correspondent cumulative hazard function. Viceversa, if S is a survival function such that supp$(S) = (0, \infty)$, then the function

$$t \mapsto H(t) = -\int_0^t \frac{\mathrm{d}S(u)}{S(u^-)}$$

is the cumulative hazard function associated to S.

Therefore, while the hazard rate could not exist, the survival function and the cumulative hazard functions always exist and they are related via the previous proposition. On the other hand, if the distribution F has a density, then the hazard rate h is defined and the expression of the cumulative hazard reported in Definition 1.2.3 boils down to

$$H(t) = -\int_0^t \frac{\mathrm{d}S(u)}{S(u^-)} = \int_0^t -\frac{S'(u)}{S(u)}\mathrm{d}u = \int_0^t h(u)\mathrm{d}u,$$

which intuitively expresses how the cumulative hazard is the integral of the hazard rate over time. It is also possible to obtain the expression from the ordinary differential equation defining the hazard rate as in Definition 1.2.2, i.e.,

$$h(t) = -\frac{S'(t)}{S(t)},$$

which leads to

$$S(t) = exp\left(-\int_0^t h(s)\mathrm{d}s\right) = e^{-H(t)}.$$

So, if the distribution F has a density, the relationship between the functions S and H expressed in Proposition 1.2.1 boils down to

$$S(t) = e^{-H(t)} \quad \text{and} \quad H(t) = -\log S(t)$$

for each $t > 0$.

The problem of estimating survival and cumulative hazard functions is faced in the literature in different ways. The classical approach relies on frequentist estimators which allow to estimate the survival function and the cumulative hazard function without any parametric assumptions and managing the censored data. In particular, the *Kaplan-Meier* estimator was introduced in Kaplan and Meier (1958) for the estimation of the survival function, while the *Nelson-Aalen* estimator was independently introduced in Nelson (1969) and in Aalen (1978) for the estimation of the cumulative hazard function. In both the estimators the observations are summarized through two counting processes, i.e.,

$$N(t) = \sum_{i=1}^n \mathbb{1}_{(0,t]}(X_i)\Delta_i = \sum_{i=1}^n \mathbb{1}_{(0,t]}(T_i)\mathbb{1}_{(0,C_i]}(T_i) \quad \text{for each } t > 0,$$

which counts the number of observed failures at time t, and

$$Y(t) = \sum_{i=1}^{n} \mathbb{1}_{[t,\infty)}(X_i) = \sum_{i=1}^{n} \mathbb{1}_{[t,\infty)}(T_i)\mathbb{1}_{[t,\infty)}(C_i) \quad \text{for each } t > 0,$$

which counts the number of subjects still alive at time t. The process Y is also called *at-risk process*, since it counts the number of subject in the dataset which are still susceptible to failure at each time t. The *Nelson-Aalen estimator* for the cumulative hazard function H is defined as

$$\hat{H}(t) = \int_0^t \frac{\mathrm{d}N(s)}{Y(s)} \mathbb{1}_{(0,\infty)}(Y(s)),$$

while the *Kaplan-Meier estimator* for the survival function S is defined as the survival function related to the Nelson-Aalen estimator (according to Proposition 1.2.1), i.e.,

$$\hat{S}(t) = \prod_{s\in(0,t]} \left(1 - \mathrm{d}\hat{H}(t)\right) = \prod_{s\in(0,t]} \left(1 - \frac{\mathrm{d}N(s)}{Y(s)} \mathbb{1}_{(0,\infty)}(Y(s))\right).$$

Note that the aforementioned estimators are nonparametric, since they do not rely on any parametric assumption on the structure of the data. The frequentist literature provides numerous alternatives to this models. For example, the *Fleming-Harrington estimator* was introduced in Fleming and Harrington (1984) as a modified version of the Kaplan-Meier estimator including weights for the observations, thus addressing dataset with a high number of censored observations. Moreover, the well known *Cox model*, also known as the *proportional hazards model*, was introduced in Cox (1972) as a semi-parametric regression model for the estimation of the hazard rate in presence of a set of covariates. Several works were then dedicated to the study and extension of the Cox model: see for example Breslow (1974), which introduced a new estimator for the cumulative hazard function, and Efron (1977), which extended the Breslow estimator to the case of tied events. A common alternative to the proportional hazard model is the *accelerated failure time (AFT) model*, a parametric model which assumes a linear relationship between survival times and covariates; please refer to Kalbfleisch and Prentice (2002) for details.
Regarding the Bayesian framework, literature provides different parametric approaches when it comes to the inference on survival times. In these cases, the analysis relies on a parametric assumption on the conditional distribution of the data and on the choice of a prior for the model parameters. Moreover, various frequentist models are suitable for a Bayesian parametric approach; for example, both the Cox model and the AFT models are often used in the Bayesian framework, inferring from then posterior distribution of their

parameters. About this topic, please refer to Carlin and Louis (2000) and Ibrahim et al. (2001). The development of Bayesian nonparametric methods in survival analysis requires more theoretical tools, and a brief introduction on this topic will be the object of the next section.

1.3 BNP models in Survival Analysis

The application of BNP methods to survival analysis is a widely discussed topic in the literature. The aim of this section is to briefly introduce the statistical tools and the general framework, while the original contributions are discussed in the next chapters. For a complete discussion of the subject, see for example Daley and Vere-Jones (2008), Lijoi and Prünster (2010) and Ghosal and van der Vaart (2017).
The main idea is to find a suitable prior on the object of interest, being it for example the survival function or the cumulative hazard function, and to find effective way to perform inference on the posterior distribution. Let us consider the notation introduced in the previous sections. When it comes to BNP methods in survival analysis, a widely used family of nonparametric processes is the family of the *neutral to the right (NTR) processes*, introduced in Doksum (1974). Let us therefore introduce the following definition.

Definition 1.3.1 (NTR)**.** A random probability measure $\tilde{p}$ on $(\mathbb{R}^+, \mathcal{B}(\mathbb{R}^+))$ is called a *neutral to the right (NTR) process* if, for any integer $k \geq 1$ and any partition $0 \leq t_1 < t_2 < \ldots t_k < \infty$ of $\mathbb{R}^+$, the random variables

$$\tilde{F}(t_1), \frac{\tilde{F}(t_2) - \tilde{F}(t_1)}{1 - \tilde{F}(t_1)}, \ldots, \frac{\tilde{F}(t_k) - \tilde{F}(t_{k-1})}{1 - \tilde{F}(t_{k-1})}$$

are independent, where $\tilde{F}(t) = \tilde{p}((0,t])$ for each $t > 0$.

A core feature of these processes is that they have a standard representation as functionals of specific random measures called *completely random measures (CRMs)*, first introduced in Kingman (1967). Let us therefore consider the generic measure space $(\Omega, \mathcal{A}, \mathbb{P})$, the Polish space with its Borel σ-algebra $(\mathbb{X}_0, \mathcal{X}_0)$ and the space $\boldsymbol{M}$ of boundedly finite measures on $(\mathbb{X}_0, \mathcal{X}_0)$, with its corresponding Borel σ-algebra $\mathcal{M}$. Let introduce the following definition.

Definition 1.3.2 (CRM)**.** A measurable map

$$\tilde{\mu} : (\Omega, \mathcal{A}, \mathbb{P}) \to (\boldsymbol{M}, \mathcal{M})$$

is a *completely random measure (CRM)* if for any set $A_1, \ldots, A_n \in \mathcal{X}_0$ such that $A_i \cap A_j = \emptyset$ for each $i \neq j$, the random variables $\tilde{\mu}(A_1), \ldots, \tilde{\mu}(A_n)$ are mutually independent.

In other words, a random measure is completely random if it maps mutually disjoint events into mutually independent random variables. A CRM $\tilde{\mu}$ is usually represented as the sum of three components, as

$$\tilde{\mu} = \tilde{\mu}_c + \tilde{\mu}_f + \tilde{\nu}_d.$$

In the previous expression, $\tilde{\mu}_c$ is functional of a marked Poisson process $(\tilde{J}_i, \tilde{X}_i)_{i\geq 1}$, as

$$\tilde{\mu}_c = \sum_{i\geq 1} \delta_{(\tilde{J}_i, \tilde{X}_i)} = \sum_{i\geq 1} \tilde{J}_i \delta_{\tilde{X}_i},$$

i.e., both the positive jumps $\tilde{J}_i$s and the $\mathbb{X}_0$-valued locations $\tilde{X}_i$s are random, while the measure $\tilde{\mu}_f$ can be represented as the finite sum

$$\tilde{\mu}_f = \sum_{i=1}^{N} V_i \delta_{X_i},$$

where the locations $X_1, \ldots, X_N \in \mathbb{X}$ are fixed values, as $N \in \mathbb{N} \cup \{\infty\}$, and the jumps $V_1, \ldots, V_N$ are positive random variables, mutually independent as well as independent from $\tilde{\mu}_c$. Finally, the component $\tilde{\nu}_d$ is a deterministic drift. Note that, when it comes to the applications of CRMs in the definition of nonparametric priors, the usual choice is to consider only the completely random component, thus ignoring the discrete measure and the deterministic drift. So from now on let us assume that $\tilde{\mu} = \tilde{\mu}_c$. Note that the measure $\tilde{\mu}$ can be also written as

$$\tilde{\mu} = \sum_{i\geq 1} \tilde{J}_i \delta_{\tilde{X}_i} = \int_{\mathbb{R}^+} s\tilde{N}(\mathrm{d}s, \mathrm{d}x),$$

where the element

$$\tilde{N} = \sum_{i\geq 1} \delta_{(\tilde{J}_i, \tilde{X}_i)}$$

can be represented via its Laplace functional as

$$\mathbb{E}\left[e^{-\int g \mathrm{d}\tilde{N}}\right] = \mathbb{E}\left[e^{-\sum_{i\geq 1} g(\tilde{J}_i, \tilde{X}_i)}\right] = \exp\left(-\int_{\mathbb{R}^+\times\mathbb{X}_0} \left(1 - e^{-g(s,x)}\right) \nu(\mathrm{d}s, \mathrm{d}x)\right),$$

for any measurable function $g : \mathbb{R}^+ \times \mathbb{X}_0 \to \mathbb{R}$ such that

$$\int \mid g \mid \mathrm{d}\tilde{N} < \infty \quad \text{almost surely,}$$

where ν is a measure on $\mathbb{R}^+ \times \mathbb{X}_0$ such that for any $B \in \mathcal{X}_0$,

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

So, for any measurable function $f : \mathbb{X}_0 \to \mathbb{R}$ such that

$$\int \mid f \mid \mathrm{d}\tilde{\mu} < \infty \quad \text{almost surely,}$$

the Laplace functional of $\tilde{\mu}$ is

$$\mathbb{E}\left[e^{-\int_{\mathbb{X}_0} f(x)\tilde{\mu}(\mathrm{d}x)}\right] = \exp\left(-\int_{\mathbb{R}^+ \times \mathbb{X}_0} \left(1 - e^{-sf(x)}\right) \nu(\mathrm{d}s, \mathrm{d}x)\right).$$

The measure ν characterizes the CRM $\tilde{\mu}$ since it contains all the information about the distribution of the jumps and the locations, and it is called *Lévy intensity* of $\tilde{\mu}$. Let us write

$$\tilde{\mu} \sim \mathrm{CRM}(\nu).$$

So it follows that a CRM is almost surely discrete, which means that its realizations are discrete measures with probability 1. The intensity of a CRM is usually written as

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

where α is a measure on $(\mathbb{X}_0, \mathcal{X}_0)$ and ρ is a transition kernel on $\mathbb{X}_0 \times \mathcal{B}(\mathbb{R}^+)$, i.e., the function

$$x \mapsto \rho_x(A)$$

is $\mathcal{X}_0$-measurable for any $A \in \mathcal{B}(\mathbb{R}^+)$ and ρ_x is a measure on $(\mathbb{R}^+, \mathcal{B}(\mathbb{R}^+))$ for any $x \in \mathbb{X}_0$. If the jumps and the locations of $\tilde{\mu}$ are independent, then we write $\rho_x = \rho$ and the corresponding Lévy intensity is called *homogeneous*.
The relationship between NTR processes and CRMs is explained by the following result.

Proposition 1.3.1 (Doksum)**.** *A random probability measure $\tilde{p}$ is neutral to the right if* *and only if there exists a CRM $\tilde{\mu}$ on $(\mathbb{R}^+, \mathcal{B}(\mathbb{R}^+))$ satisfying*

$$\lim_{t\to\infty} \tilde{\mu}((0,t]) = \infty \quad \textit{almost surely}$$

such that the law of $\left\{\tilde{F}(t) : t \geq 0\right\}$ coincides with the law of $\left\{1 - e^{-\tilde{\mu}((0,t])} : t \geq 0\right\}$, where $\tilde{F}(t) = \tilde{p}((0,t])$ for any $t \geq 0$. Let us write

$$\tilde{p} \sim \textit{NTR}(\tilde{\mu}).$$

For convenience, we will call $\tilde{F}$ the NTR process. So, from Proposition 1.3.1 it follows that each NTR process can be uniquely identified by the Lévy intensity of its correspondent CRM. For instance, the prior guess for $\tilde{p} \sim \mathrm{NTR}(\tilde{\mu})$ can be written as

$$\mathbb{E}\left[\tilde{F}\right] = 1 - \mathbb{E}\left[e^{-\tilde{\mu}((0,t])}\right] = 1 - e^{-\int_0^t \int_{\mathbb{R}^+} (1-e^{-s})\rho_x(\mathrm{d}s)\alpha(\mathrm{d}x)},$$

where $\tilde{\mu} \sim \mathrm{CRM}(\nu)$ and $\nu(\mathrm{d}s, \mathrm{d}x) = \rho_x(\mathrm{d}s)\alpha(\mathrm{d}x)$. The reason for the popularity of these processes as nonparametric priors in survival analysis is twofold: first of all, as represented above, their prior distribution can be completely characterized by the Lévy intensity of a CRM; on the other hand, NTR processes forms a class of conjugate priors in survival analysis, thus even in presence of a censorship mechanism, and so also their posterior distributions are completely characterized by the Lévy intensity of the correspondent CRM. The conjugacy of the NTR processes is stated in the following theorem, due to Doksum (1974).

Theorem 1.3.1 (Doksum)**.** *Let us consider a set of exchangeable observations $X_1, \ldots, X_n$, and the model*

$$X_1, \ldots, X_n \mid \tilde{F} \overset{i.i.d.}{\sim} \tilde{F},$$

where $\tilde{F}$ is a NTR($\tilde{\mu}$). Then the posterior distribution of $\tilde{F}$ is

$$\tilde{F} \mid X_1, \ldots, X_n \sim \textit{NTR}(\tilde{\mu}^*),$$

where $\tilde{\mu}^*$ is a CRM with fixed point of discontuinity.

Note that the previous result proves that the NTR processes are a class of conjugate priors, but in this context the term "conjugacy" does not indicate that the posterior process has the same probability distribution of the prior process with updated parameters (as for example the Dirichlet process described in Section 1.1), while it indicates a structural conjugacy, i.e., the posterior process belongs to the same class of nonparametric process.

On the other hand, the CRM characterizing the posterior distribution of a NTR is not necessarily of the same type as the prior. Moreover, the posterior distribution of a NTR process can be written as

$$\tilde{F}(t) \mid X_1, \dots, X_n = 1 - e^{-\tilde{\mu}((0,t])} = 1 - e^{-\sum_{i\geq 1} \tilde{J}_i \delta_{\tilde{X}_i} - \sum_{i=1}^{N} V_i \delta_{X_i}};$$

note that the result stated in Theorem 1.3.1 does not provide an explicit description of this distribution. An explicit description of the posterior CRM $\tilde{\mu}^*$ has been provided in Ferguson (1974). Let us report the result in the following theorem.

Theorem 1.3.2 (Ferguson)**.** *Let us consider a set of exchangeable observations $X_1, \dots, X_n$, and the model*

$$X_1, \dots, X_n \mid \tilde{F} \overset{i.i.d.}{\sim} \tilde{F},$$

where $\tilde{F}$ is a NTR($\tilde{\mu}$), $\tilde{\mu} \sim$ CRM(ν) and $\nu(\mathrm{d}s, \mathrm{d}x) = \rho_x(\mathrm{d}s)\alpha(\mathrm{d}x)$. Let $X_1^*, \dots, X_k^*$ be the k distinct observations among $X_1, \dots, X_n$. Then the posterior distribution of $\tilde{F}$ can be written as

$$\begin{aligned}\tilde{F} \mid X_1, \dots, X_n &\sim NTR(\tilde{\mu}^*),\\ \tilde{\mu}^* &= \tilde{\mu}_c^* + \sum_{i=1}^{k} V_i \delta_{X_i^*},\\ \tilde{\mu}_c^* &\sim CRM(\nu^*),\\ \nu^*(\mathrm{d}s, \mathrm{d}x) &= e^{-Y(x)s} \rho_x(\mathrm{d}s)\alpha(\mathrm{d}x),\end{aligned}$$

where the CRM $\tilde{\mu}_c$ is independent from the jumps $V_1, \dots, V_k$ and the V_ks are mutually independent, and the function $Y(x)$ is the at-risk process related to the dataset $X_1, \dots, X_n$.

It is also possible to determine the probability density of the jumps V_i. Let us therefore consider the distinct observations and let us assume, without loss of generality, that they are increasingly ordered, i.e.

$$X_1^* < \dots < X_1^*.$$

Moreover, let us consider the counting process

$$n_i = \sum_{j=1}^{n} \delta_{X_j}\left(\{X_i^*\}\right) = Y\left(X_i^*\right) - Y\left(X_{i+1}^*\right), \quad \text{for any } i = 1, \dots, k,$$

i.e., the frequency of the ith distinct observation X_i^*, assuming $X_{k+1}^* = \infty$. Then the probability density of the jump V_i is infinitesimally proportional on $\mathbb{R}^+$ to

$$(1-e^{-s})^{n_i} e^{-s\bar{n}_{i+1}} \rho_{X_i^*}(\mathrm{d}s),$$

where $\bar{n}_i$ is the number of ordered and distinct observations from the ith onwards, i.e.,

$$\bar{n}_i = \sum_{j=i}^{k} n_j = Y\left(X_i^*\right) \quad \text{for any } i = 1, \ldots, k.$$

Note that if the Leévy intensity of the CRM $\tilde{\mu}$ is homogeneous, then $\rho_{X_i^*} = \rho$ for any i, and the distribution of each V_i does not depend on the location where the jumps occurs. Note also that the posterior characterization does not take into account the possibility that the data are subject to a censoring mechanism according to which not all observations are exact. The following result, introduced in Ferguson and Phadia (1979), provides a posterior characterization for NTR priors even when the observations are survival times, thus with censoring.

Theorem 1.3.3 (Ferguson & Phadia)**.** *Let us consider a set of exchangeable survival data* $\boldsymbol{D} = \{(X_i, \Delta_i)\}_{i=1}^n$*, and the model*

$$(X_i, \Delta_i) \mid \tilde{F} \overset{i.i.d.}{\sim} \tilde{F} \quad \textit{for each } i,$$

where $\tilde{F}$ *is a NTR*$(\tilde{\mu})$*,* $\tilde{\mu} \sim$ *CRM*(ν) *and* $\nu(\mathrm{d}s, \mathrm{d}x) = \rho_x(\mathrm{d}s)\alpha(\mathrm{d}x)$*. Let* $X_1^*, \ldots, X_k^*$ *be the* k *distinct observations among* $X_1, \ldots, X_n$*. Then the posterior distribution of* $\tilde{F}$ *can be written as*

$$\tilde{F} \mid \boldsymbol{D} \sim NTR(\tilde{\mu}^*),$$
$$\tilde{\mu}^* = \tilde{\mu}_c^* + \sum_{i:\Delta_i=1} V_i \delta_{X_i^*},$$

where:

- $\tilde{\mu}_c^* \sim$ *CRM*(ν^*) *and*
$$\nu^*(\mathrm{d}s, \mathrm{d}x) = e^{-Y(x)s} \rho_x(\mathrm{d}s)\alpha(\mathrm{d}x);$$
- *the jumps* V_i*s are mutually independent and the probability density of the jump* V_i *is infinitesimally proportional on* $\mathbb{R}^+$ *to*
$$(1-e^{-s})^{n_i} e^{-s(\bar{n}_{i+1}+\tilde{n}_i^c)} \rho_{X_i^*}(\mathrm{d}s),$$

where

$$n_i = \sum_{j=1}^{n} \mathbb{1}_{\left(X_j = X_i^*; \Delta_j = 1\right)} \quad \text{and} \quad n_i^c = \sum_{j=1}^{n} \mathbb{1}_{\left(X_j = X_i^*; \Delta_j = 0\right)}$$

are the number of exact and censored observations equal to X_i^**, for each* $i = 1, \ldots, k$*, while*

$$\bar{n}_i = \sum_{j=i}^{k} n_i \quad \text{and} \quad \tilde{n}_i^c = \sum_{j=i}^{k} n_j^c \quad \text{for any } i = 1, \ldots, k.$$

Moreover, $\tilde{\mu}_c^*$ *is independent from the jumps* V_i*s.*

So the result stated in Theorem 1.3.3 guarantees that NTR processes form class of conjugate priors even when the data are survival times, and provides a closed form representation for the posterior distribution.

1.3.1 Conjugate NTR processes

As said before, under a NTR prior the posterior is distributed again as a NTR process, but in general it is distributed as a different one. This is the case of the Dirichlet process.

Proposition 1.3.2. *Let us consider a random probability measure* $\tilde{p}$ *on* $(\mathbb{R}^+, \mathcal{B}(\mathbb{R}^+))$ *and a probability measure* α *on* $\mathbb{R}^+$*. Then,* $\tilde{p} \sim \mathcal{D}_\alpha$ *if and only if there exists a* $\tilde{\mu} \sim CRM(\nu)$ *such that*

$$\tilde{p} \sim NTR(\tilde{\mu}) \quad \text{and} \quad \nu(\mathrm{d}s, \mathrm{d}x) = \frac{e^{-s\alpha((x,\infty))}}{1 - e^{-s}} \alpha(\mathrm{d}x)\mathrm{d}s.$$

So, each Dirichlet process is a NTR process. Note that according to Theorem 1.3.3 it follows that the posterior distribution of the Dirichlet random distribution function $\tilde{F}$ is NTR($\tilde{\mu}^*$), where $\tilde{\mu}^*$ is a CRM whose Lévy intensity

$$\nu^*(\mathrm{d}s, \mathrm{d}x) = \frac{e^{-(\alpha((x,\infty)) + Y(x))s}}{1 - e^{-s}} \alpha(\mathrm{d}x)\mathrm{d}s,$$

and the distribution of each jump V_i at each exact distinct observation X_i^* such that $\Delta_i = 1$ is equal to the distribution of the variable $-\log(B_i)$, where $B_i \sim \text{Beta}\left(\alpha((X_i^*, \infty)) + \bar{n}_{i+1} + \tilde{n}_i^c, n_i\right)$, for any $i = 1, \ldots, k$. Note that if the observations are all exact, then $\tilde{F} \mid \boldsymbol{D}$ is a Dirichlet process with parameter $\alpha + \sum_{i=1}^{n} \delta_{X_i}$, recovering the well-known result proved by Ferguson (1973). On the other hand, if the data are censored, the posterior process is not a Dirichlet process anymore, while having the distribution of a different NTR process. As discussed before, this is a topic that naturally arises when it comes to search for NTR priors in survival analysis: while they all shows a structural conjugacy, one may ask whether there exist particular NTR processes that show even a parametric conjugacy in this context. A first example is given by the *beta-stacy process*, introduced in Walker and Muliere (1997) as nonparametric prior for survival functions.

Definition 1.3.3 (Beta-stacy process)**.** Let α be a probability measure on $\mathbb{R}^+$ which is absolutely continuous with respect to the Lebesgue measure, and let $c : \mathbb{R}^+ \to \mathbb{R}^+$ be some piecewise continuous function. A random probability distribution $\tilde{F}$ is a *beta-stacy process with parameters* α *and* c if it is a $NTR(\tilde{\mu})$ such that $\tilde{\mu} \sim \text{CRM}(\nu)$, where

$$\nu(\mathrm{d}s, \mathrm{d}x) = \frac{e^{-sc(x)\alpha((x,\infty))}}{1 - e^{-s}} c(x) \mathrm{d}s \alpha(\mathrm{d}x).$$

Let us write

$$\tilde{p} \sim \text{Beta-Stacy}(\alpha, c),$$

where $\tilde{p}$ is the random probability measure whose correspondent distribution is $\tilde{F}$.

Note that when $c(x) \equiv c \in \mathbb{R}^+$ for any $x \in \mathbb{R}^+$, this process boils down to the Dirichlet process. Moreover, the posterior beta-stacy process is distributed as another beta-stacy process, with updated parameter. Let us therefore state the following theorem, that can be found in Walker and Muliere (1997).

Theorem 1.3.4 (Walker & Muliere)**.** *Let us consider a set of exchangeable survival data* $\boldsymbol{D} = \{(X_i, \Delta_i)\}_{i=1}^n$*, and the model*

$$\begin{aligned} \boldsymbol{D} \mid \tilde{F} &\overset{i.i.d.}{\sim} \tilde{F} \\ \tilde{F} &\sim \textit{Beta-Stacy}(\alpha, c), \end{aligned}$$

for a certain probability measure α *on* $\mathbb{R}^+$ *and a piecewise continuous function* $c : \mathbb{R}^+ \to \mathbb{R}^+$*. Then the posterior process* $\tilde{F} \mid \boldsymbol{D}$ *is distributed as another beta-stacy process, with parameters* α^* *and* c^**, where*

$$\alpha^*((0,t]) = 1 - \prod_{x \in [0,t]} \left[1 - \frac{c(x)\mathrm{d}\alpha((0,x]) + \mathrm{d}N(x)}{c(x)\alpha([x,\infty)) + Y(x)} \right], \quad \textit{and}$$

$$c^*(x) = \frac{c(x)\alpha([x,\infty)) + Y(x) - \sum_{i=1}^n \delta_{X_i}(\{x\})\delta_{\Delta_i}(\{1\})}{\alpha^*([x,\infty))},$$

where N *is the counting process for the uncensored observations introduced in the previous section, while* $\prod_{x \in [0,t]}$ *is the product integral.*

Note that the parameters of the posterior beta-stacy process show a clear dependence on the data. According to Theorem 1.3.4, the prior random survival function is

$$\tilde{S}(t) = e^{-\tilde{\mu}((0,t])} \quad \text{for any } t \geq 0,$$

where $\tilde{\mu}$ is the CRM of a beta-stacy process. Moreover, the posterior distribution of the survival function under the beta-stacy model is

$$\tilde{S}(t) \mid \boldsymbol{D} = e^{-\tilde{\mu}^*((0,t])} \quad \text{for any } t \geq 0,$$

where $\tilde{\mu}^*$ is the CRM of the posterior NTR process $\tilde{F} \mid \boldsymbol{D}$. Note also that the posterior characterizations for NTR process presented in Theorem 1.3.3 and Theorem 1.3.4 have an additional advantage, since they reduce the problem of sampling from the posterior distribution of a NTR to the problem of approximating a CRM. This will be discussed in the next chapters.
A second example of NTR process which shows parametric conjugacy is the *beta process*, introduced in Hjort (1990) for the estimation of the cumulative hazard function. The idea is to assess a prior for the cumulative hazard, which is linked to the random distribution function $\tilde{F}$ via the usual relation

$$\tilde{F}(t) = 1 - \prod_{x \in [0,t]} \left\{ 1 - \mathrm{d}\tilde{H}_x \right\} \quad \text{for any } t \geq 0,$$

where $\prod_{x\in[0,t]}$ is the product integral. It follows that assessing a prior for $\tilde{F}$ implies assessing a prior for the random cumulative hazard process $\tilde{H} = \left\{ \tilde{H}_t : t \geq 0 \right\}$. Note that in the following we write $\tilde{H}_t$ to indicate the cumulative hazard function $\tilde{H}$ evaluated at time t. The idea is to model $\tilde{H}$ via a suitable CRM $\tilde{\mu}$ by setting

$$t \mapsto \tilde{H}_t := \tilde{\mu}((0,t]).$$

Let us therefore introduce the following definition.

Definition 1.3.4 (Beta process)**.** Let $c : \mathbb{R}^+ \to \mathbb{R}^+$ be some piecewise continuous function and let H_0 be a baseline cumulative hazard. The random process

$$\tilde{H} = \{\tilde{\mu}((0,t]) : t \geq 0\}$$

is a *beta process with parameters* c *and* H_0 if $\tilde{\mu} \sim \mathrm{CRM}(\nu)$, where

$$\nu(\mathrm{d}s, \mathrm{d}x) = c(x)s^{-1}(1-s)^{c(x)-1}\mathrm{d}s\mathrm{d}H_{0,x}$$

as $s \in (0,1)$ and $x \geq 0$. Let us write

$$\tilde{H} \sim \mathrm{Beta}(c, H_0).$$

Note that

$$\mathbb{E}\left[\tilde{H}_t\right] = H_{0,t} \quad \text{for any } t \geq 0,$$

so the baseline hazard H_0 plays the role of prior guess on the cumulative hazard function under a beta prior. Moreover, the relation between modelling the cumulative hazard with a CRM and specifying a NTR prior for the distribution function is clarified by the following theorem, proved in Hjort (1990).

Theorem 1.3.5 (Hjort)**.** *Let us consider a CRM* $\tilde{\mu}$*. Then, a random distribution function* $\tilde{F}$ *is NTR*($\tilde{\mu}$) *if and only if the corresponding cumulative hazard is a independent increments process with Lévy intensity* $\nu(\mathrm{d}s, \mathrm{d}x) = \rho_x(s)\mathrm{d}s\alpha(\mathrm{d}x)$ *satisfying the condition*

$$\int_1^\infty \rho_x(s)\mathrm{d}s = 0.$$

Note that the previous theorem states that the Lévy intensity should be concentrated on $[0,1]$, implying that the jumps of the CRM $\tilde{\mu}$ should be in $[0,1]$. According to Definition 1.3.4, the beta process satisfies this condition. As far as the posterior characterization is concerned, note that Hjort (1990) also provides a general theory for a generic CRM cumulative hazard, in terms of an updated CRM with fixed point of discontinuity corresponding to the exact and distinct observations. The link between the priors for cumulative hazards and NTR processes is deepened and expanded for example in Dey et al. (2003). The appealing feature of the beta process is its parametric conjugacy: the posterior beta process is again a beta process, with updated parameters. Let us state the result, due to Hjort (1990).

Theorem 1.3.6 (Hjort)**.** *Let us consider a set of exchangeable survival data* $\boldsymbol{D} = \{(X_i, \Delta_i)\}_{i=1}^n$

and the model

$$
\begin{aligned}
(X_1, \dots, X_N) \mid \tilde{F} \sim & \tilde{F} \\
\tilde{F}(t) = & 1 - \prod_{s \in (0,t]} \left\{1 - \mathrm{d}\tilde{H}_s\right\} \quad \textit{for any } t \geq 0 \\
\tilde{H} \sim & \textit{Beta}(c, H_0),
\end{aligned}
$$

for a certain basline hazard H_0 and a piecewise continuous function $c : \mathbb{R}^+ \to \mathbb{R}^+$. Then the posterior process $\tilde{H} \mid \boldsymbol{D}$ is distributed as

$$
\tilde{H} \mid \boldsymbol{D} \sim \textit{Beta}\left(c + Y, \int \frac{c\mathrm{d}H_0 + \mathrm{d}N}{c + Y}\right),
$$

where N and Y are the usual counting processes introduced before.

Note that according to Theorem 1.3.6, the Bayes estimator of $\tilde{H}$ and $\tilde{F}$ with respect to a squared loss function are

$$
\mathbb{E}\left[\tilde{H}_t \mid \boldsymbol{D}\right] = \int_0^t \frac{c\mathrm{d}H_0 + \mathrm{d}N}{c + Y} \quad \text{and}
$$

$$
\mathbb{E}\left[\tilde{F}(t) \mid \boldsymbol{D}\right] = 1 - \prod_{[0,t]} \left[1 - \frac{c\mathrm{d}H_0 + \mathrm{d}N}{c + Y}\right]
$$

respectively. As discussed for the beta-stacy process, letting $c \to 0$ the two estimators converges to the Nelson-Aalen and to the Kaplan-Meier estimators respectively. Note that beta-stacy and beta processes are also discussed in Lijoi and Prünster (2010), where authors consider other priors proposed in the BNP literature applied to survival analysis, which arise as suitable transformations of CRMs. Other noteworthy BNP contributions to survival problems include Kim (1999) and James (2006). In particular, in James (2006) the author propose a new family of priors called *spatial neutral to the right processes*, which are useful when one is interested in modelling survival times coupled with variables which take place in a general space (usually, a spatial component).

1.3.2 Other contributions

A number of papers have focused on the issue of specifying a prior for the hazard rate, instead of the cumulative hazard or the survival function. In many of these works, this is achieved via the so-called *life-testing models*. In particular, in this models a prior for the hazard rate

$$h(t) = \frac{F'(t)}{1 - F(t)}$$

is specified in terms of a mixture with respect to a CRM. Let $k(\cdot \mid \cdot)$ be a kernel on $\mathbb{R}^+ \times \mathbb{Y}$, for a further Polish space $(\mathbb{Y}, \mathcal{Y})$, i.e., k is measurable with respect to both its variables and one has

$$\int_B k(x \mid y) \mathrm{d}x \quad \text{for any bounded } B \in \mathcal{B}(\mathbb{R}^+).$$

Then, in this context a prior for the hazard rate is the probability density of the random hazard rate

$$\tilde{h}(t) = \int_{\mathbb{Y}} k(x \mid y) \tilde{\mu}(\mathrm{d}y),$$

where $\tilde{\mu}$ is a CRM on $(\mathbb{Y}, \mathcal{Y})$. For example, in Dykstra and Laud (1981) the authors introduced a random hazard called *extended gamma process*. Other models were proposed in Lo and Weng (1989) and James (2005). In particular, in James (2005) the author obtains a representation of the posterior distribution of the CRM $\tilde{\mu}$, considering the second variable y of the kernel k as a latent variable. For further references on this topic, see for example Nieto-Barajas and Walker (2002), Nieto-Barajas and Walker (2004) and Ishwaran and James (2004).

1.4 Main contributions of the Thesis

In this chapter, after briefly introducing the basic concepts of Bayesian nonparametrics and survival analysis, we provided an essential review of BNP methods in survival analysis, with a specific focus on the case of exchangeable survival times. We also provided some notable examples of nonparametric priors which, like the beta and the beta-stacy processes, not only exhibit parametric conjugacy, but prove also to be a standard in the literature and have been the subject of numerous extensions. The rest of this work is dedicated to the analysis and development of innovative BNP methods to address certain research questions related to survival analysis.

In particular, in Chapter 2 we consider the class of survival models known as cure-rate models, which are used to handle survival time datasets which show a strictly positive probability (called *cure rate*) of not observing the failure event. This implies that a fraction of the population is either cured or not susceptible to the event and, therefore, will not experience failure. Assuming exchangeable survival times, we then introduce a class of nonparametric priors useful for modeling both the survival and the cumulative hazard functions, which allow for the estimation of the strictly positive cure rate. In addition to providing a comprehensive analysis of the prior and posterior processes, we also demonstrate how priors of this class can be estimated particularly efficiently through appropriate approximation algorithms.

In Chapter 3, we contribute to the line of research dedicated to exploring BNP methods for the modeling of partially exchangeable data, an assumption that proves reasonable in survival analysis across various applications. For example, consider a clinical trial conducted with the same drug but on groups of patients located in different facilities, or more generally, under different conditions. Motivated by these observations, we present a hierarchical extension of the well-known NTR processes, which are traditionally used as nonparametric priors for exchangeable times. This extension allows for modeling survival functions of groups of observations conditionally on a baseline measure, which captures the dependence between the groups. The analysis of the posterior distribution of the introduced process enables us to appreciate the borrowing of information from each group to all others. Additionally, by leveraging an appropriate representation of the posterior distribution of our hierarchical process, it is possible to develop a conditional algorithm that allows for sampling trajectories of the survival functions for each group.

Bibliography

Aalen, O. (1978). Nonparametric inference for a family of counting processes. *The Annals of Statistics*, 6(4):701–726.

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022.

Breslow, N. (1974). Covariance analysis of censored survival data. *Biometrics*, 30(1):89–99.

Camerlenghi, F., Lijoi, A., Orbanz, P., and Prünster, I. (2019). Distribution theory for hierarchical processes. *The Annals of Statistics*, 47(1):67–92.

Carlin, B. P. and Louis, T. A. (2000). *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman and Hall, 2 edition.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–220.

Daley, D. J. and Vere-Jones, D. (2008). *An introduction to the theory of point processes.* Probability and Its Applications, a Series of the Applied Probability Trust. New York: Springer, 2nd edition.

De Blasi, P., Favaro, S., Lijoi, A., Mena, R. H., Prunster, I., and Ruggiero, M. (2015). Are Gibbs-type priors the most natural generalization of the Dirichlet process? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(2):212–229.

Dey, J., Erickson, R. V., and Ramamoorthi, R. V. (2003). Some aspects of neutral to the right priors. *International Statistical Review*, 71:383–401.

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

Dunson, D. B. (2009). Bayesian nonparametric hierarchical modeling. *Biometrical Journal*, 51(2):273–284.

Dunson, D. B. (2010). Nonparametric Bayes applications to biostatistics. In *Bayesian Nonparametrics*, pages 223–273. Cambridge University Press, Cambridge.

Dykstra, R. L. and Laud, P. (1981). A Bayesian nonparametric approach to reliability. *The Annals of Statistics*, 9(2):356–367.

Efron, B. (1977). The efficiency of Cox's likelihood function for censored data. *Journal of the American Statistical Association*, 72(359):557–565.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1:209–230.

Ferguson, T. S. (1974). Prior distributions on spaces of probability measures. *The Annals of Statistics*, 2(4):615–629.

Ferguson, T. S. and Phadia, E. G. (1979). Bayesian nonparametric estimation based on censored data. *The Annals of Statistics*, 7(1):163–186.

Finetti, B. d. (1937). La prévision : ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré'*, 7(1):1–68.

Finetti, B. d. (1938). Sur la condition d'equivalence partielle. *Actualités scentiques et industrielles*, (739):5–18.

Fleming, T. R. and Harrington, D. P. (1984). Nonparametric estimation of the survival distribution in censored data. *Communications in Statistics - Theory and Methods*, 13(20):2469–2486.

Ghosal, S. and van der Vaart, A. (2017). *Fundamentals of nonparametric Bayesian inference*, volume 44 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge.

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.

Hjort, N. L. (1990). Nonparametric Bayes estimators based on beta processes in models for life history data. *The Annals of Statistics*, 18(3):1259–1294.

Ibrahim, J. G., Chen, M., and Sinha, D. (2001). *Bayesian Survival Analysis*. Springer New York.

Ishwaran, H. and James, L. F. (2004). Computational methods for multiplicative intensity models using weighted Gamma processes. *Journal of the American Statistical Association*, 99(465):175–190.

James, L. F. (2005). Bayesian Poisson process partition calculus with an application to Bayesian Lévy moving averages. *The Annals of Statistics*, 33(4):1771–1799.

James, L. F. (2006). Poisson calculus for spatial neutral to the right processes. *The Annals of Statistics*, 34(1):416–440.

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*. Wiley, 2 edition.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282):457–481.

Kim, Y. (1999). Nonparametric Bayesian estimators for counting processes. *The Annals of Statistics*, 27(2):562–588.

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

Lijoi, A. and Prünster, I. (2010). Models beyond the Dirichlet process. In *Bayesian nonparametrics*, volume 28 of *Camb. Ser. Stat. Probab. Math.*, pages 80–136. Cambridge Univ. Press, Cambridge.

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

MacEachern, S. N. (1999). Dependent nonparametric processes. In *ASA Proceedings of the SBSS*, pages 50–55. Alexandria: American Statistical Association.

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

Nelson, W. (1969). Theory and applications of hazard plotting for censored failure data. *Technometrics*, 11(4):945–966.

Nieto-Barajas, L. E. and Walker, S. (2002). Markov beta and gamma processes for modelling hazard. *Scandinavian Journal of Statistics*, 29(3):413–424.

Nieto-Barajas, L. E. and Walker, S. (2004). Bayesian nonparametric survival analysis via Lévy driven Markov processes. *Statistica Sinica*, 14(4):1127–1146.

Pitman, J. and Yor, M. (1997). The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *The Annals of Probability*, 25(2):855–900.

Quintana, F. A., Müller, P., Jara, A., and MacEachern, S. N. (2022). The dependent Dirichlet process and related models. *Statistical Science*, 37(1):24–41.

Regazzini, E. (1996). Impostazione non parametrica di problemi d'inferenza statistica Bayesiana. Technical Report 96.21, IAMI CNR.

Rodríguez, A., Dunson, D. B., and Gelfand, A. E. (2008). The nested Dirichlet process. *Journal of the American Statistical Association*, 103(483):1131–1154.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica*, 4(2):639–650.

Teh, Y. W. and Jordan, M. I. (2010). Hierarchical Bayesian nonparametric models with applications. In *Bayesian Nonparametrics*, pages 158–207. Cambridge University Press, Cambridge.

Teh, Y. W., Jordan, M. I., Beal, M. J., and Blei, D. M. (2006). Hierarchical Dirichlet processes. *Journal of the American Statistical Association*, 101(476):1566–1581.

Walker, S. and Muliere, P. (1997). Beta-Stacy processes and a generalization of the Pólya-urn scheme. *The Annals of Statistics*, 25(4):1762–1780.

Chapter 2

Cure rate models based on scaled processes

2.1 Introduction

In various applications of survival analysis, it is natural to consider a strictly positive probability of not observing the failure event. For example, consider the case of a clinical study on a cohort of patients, a percentage of whom are long-term survivors. The standard cure rate model introduced in Berkson and Gage (1952) is a mixture model defined by

$$S_{\text{pop}}(t) = \pi + (1-\pi)S(t) \quad \text{for any } t \geq 0,$$

where $\pi \in (0,1)$ is the cure rate and $S(t)$ is a proper survival function, i.e.,

$$\lim_{t\to\infty} S(t) = 0.$$

Therefore, the survival function of the population $S_{\text{pop}}(t)$ converges to the positive cure rate π as $t \to \infty$. Thus, if we assume that the distribution of the survival times has density and we consider the usual relationship with the cumulative hazard function of the population $H_{\text{pop}}(t)$, we have that

$$H_{\text{pop}}(t) = -\log\left(S_{\text{pop}}(t)\right),$$

which implies that the cumulative hazard function converges to $-\log\pi$ as $t \to \infty$. Literature provides many works which extensively study the standard cure rate model: see for example the classical Kuk and Chen (1992), Maller and Zhou (1992), Maller and Zhou (1995), Sy and Taylor (2000) and the alternative model proposed in Yakovlev and Tsodikov (1996). Comprehensive reviews and discussion on cure rate models can be found for example in Ibrahim et al. (2001b), Tsodikov et al. (2003) and more recently in Amico and Van Keilegom (2018). Among the Bayesian literature, cure rate models were introduced for example in Chen et al. (1999), in Ibrahim et al. (2001a) and in Nieto-Barajas and Yin

(2008). On the other hand, while the BNP literature provides a substantial amount of contributions regarding survival analysis, it addresses this problem to a much lesser extent when there is a non-zero probability of cure, i.e., a cure rate. This is also due to the fact that the Bayesian nonparametric literature for survival analysis focuses usually on priors which induce a proper estimate of the survival function, i.e., a survival function which converges to zero as time goes to infinity. The reason for this is partly practical, as in various applications it is reasonable to assume zero survival at infinite time, and partly theoretical: classical neutral to the right (NTR) nonparametric priors, for example, induce a survival function that naturally converges to zero. For more information, see for example Daley and Vere-Jones (2008) and Lijoi and Prünster (2010). This work therefore falls within the field of BNP research applied to survival analysis, with the aim of proposing a model applicable in the presence of a strictly positive probability of cure.

More specifically, in this work we aim to introduce a class of nonparametric priors in order to deal with survival datasets with a cure rate. In particular, we show how this can be done by considering the *scaled processes (SP)*, a class of nonparametric processes that are useful in various settings; as shown in Camerlenghi et al. (2022), for example, they are valuable in species sampling models. Scaled processes can be constructed starting from a completely random measure (CRM) (see Kingman (1967)), whose jumps are ordered and re-scaled by the biggest one, which is further re-scaled by a random factor. The law of the re-scaling factor and the Lévy intensity of the starting CRM define the specific SP. In particular, again in Camerlenghi et al. (2022), the authors introduce a specific scaled process called *stable-beta scaled process (SB-SP)*, whose parametric assumptions allows, when they are assumed as prior distribution, to obtain closed form representation of the quantities of interest, such as the posterior distribution. In this work we aim to apply SB-SPs as priors in survival analysis. We will show that SB-SPs are a flexible class of processes in the survival framework. First, it is possible to develop a general theory for SPs in survival analysis and derive the results for SB-SPs as a special case. Second, the main results concerning SB-SPs lead to closed-form formulas. Finally, SB-SPs can also be obtained from different nonparametric processes, such as the well-known 3-parameter Indian Buffet Process (IBP); see Teh and Gorur (2009).

The outline of the chapter is as follows. In Section 2.2, following Camerlenghi et al. (2022) we introduce the scaled processes and its main particular case, i.e., the stable-beta scaled processes, only focusing on the definitions. Our survival model is introduced in Section 2.3: in particular, we focus on the general theory for SPs as nonparametric priors in survival analysis, particularly when adopted as priors for the cumulative hazard function. The main results of this work are discussed in Section 2.4: here we develop the theory of SB-SPs as nonparametric priors for cumulative hazard functions, obtaining relatively simple forms for the marginal and the posterior distributions. Moreover, we show how SB-SPs can be also obtained as gamma-mixtures of 3-parameter Indian Buffet Process, thus underlying the link between these processes and the standard nonparametric prior for cumulative hazard functions, the beta process introduced in Hjort (1990). The closed form of the posterior distribution of a SB-SP allows to develop a conditional algorithm in order to sample trajectories for the cumulative hazard function from the posterior distribution itself, as well as a marginal algorithm: both the strategies are discussed in Section 2.5. Finally, Section 2.6 is focused on the application of the conditional algorithm on the estimation of the survival functions of a set of simulated survival times and, finally, of the well known Bone-Marrows Transplantation dataset already used for example in Nieto-Barajas and Yin (2008). Proofs and other technical details are deferred to the Appendix.

2.2 Background on scaled processes

The aim of this section is to introduce the main tools of this paper, i.e., the class of random measures known as *scaled processes (SPs)*. First introduced in James et al. (2015), and used in Camerlenghi et al. (2022) for the estimation of unseen genetic variation, our goal is to exploit them to build a survival cure rate model. This will be discussed in the next sections. Therefore, let us now introduce the notation and the definitions, that will be used in the next sections of this paper.

Let us consider a measurable space $(\mathbb{X}, \mathcal{X})$, where $\mathbb{X}$ is a Polish space with its Borel σ-algebra $\mathcal{X}$, and let $\tilde{\mu} \sim \mathrm{CRM}(\nu)$ be a CRM on $\mathbb{X}$ whose homogeneous Lévy intensity on $\mathbb{R}^+ \times \mathbb{X}$ is

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

where α is a measure on $(\mathbb{X}, \mathcal{X})$ and ρ is a transition kernel such that

$$\int_{\mathbb{R}^+} \min\{s, 1\}\rho(s)\mathrm{d}s < \infty.$$

It is well known from Kingman (1967) that $\tilde{\mu}$ is functional of a marked Poisson process $(\tilde{h}_k, \tilde{x}_k)_{k\geq 1}$ on $\mathbb{R}^+ \times \mathbb{X}$, such that $\sum_{k\geq 1} \tilde{h}_k < \infty$ and the $\tilde{x}_k$'s are i.i.d. and independent of the $\tilde{h}_k$'s for each $k \geq 1$.

$$\tilde{\mu} = \sum_{k\geq 1} \tilde{h}_k \delta_{\tilde{x}_k}. \tag{2.1}$$

Starting from the CRM $\tilde{\mu}$, the random measure whose jumps are the ordered jumps of $\tilde{\mu}$ re-scaled by its biggest jump is called *generalized Dickman measure*, as summarized in the following definition. For further details, see again James et al. (2015).

Definition 2.2.1 (Generalized Dickman Measure)**.** Let us consider $\tilde{\mu} \sim \text{CRM}(\nu)$ as in (2.1). Let $\Delta_1 > \Delta_2 > \dots$ be the decreasingly ordered $\tilde{h}_k$'s. The discrete random measure

$$\tilde{\mu}_{\Delta_1} = \sum_{k\geq 1} \frac{\Delta_{k+1}}{\Delta_1} \delta_{\tilde{x}_{k+1}}, \tag{2.2}$$

i.e., the CRM with locations $\tilde{x}_2, \tilde{x}_3, \dots$ and scaled jumps $\frac{\Delta_2}{\Delta_1}, \frac{\Delta_3}{\Delta_1}, \dots$, is called *generalized Dickman measure*. The variables $\Delta_2, \Delta_3, \dots$ are independent given the largest jump Δ_1.

Assuming the Lévy intensity $\nu(\mathrm{d}s, \mathrm{d}x) = \rho(s)\mathrm{d}s\alpha(\mathrm{d}x)$ of the CRM $\tilde{\mu}$, it is well known from Wolpert and Ickstadt (1998) that the biggest jump Δ_1 is distributed according to

$$\Delta_1 \sim F_{\Delta_1}(\mathrm{d}a) = f_{\Delta_1}(a)\mathrm{d}a, \quad \text{where} \quad f_{\Delta_1}(a) = \exp\left\{-\int_a^\infty \rho(s)\mathrm{d}s\right\}\rho(a). \tag{2.3}$$

Let us furthermore denote by G_a the conditional law of the scaled jumps, i.e.,

$$\left(\frac{\Delta_{k+1}}{\Delta_1}\right)_{k\geq 1} \mid \Delta_1 = a \sim G_a.$$

Our goal is to define the family of scaled processes, starting from the generic measure introduced in Definition 2.2.1. Let us therefore consider a non-negative function h and let us define a re-scaled random biggest jump $\Delta_{1,h}$ with density

$$f_{\Delta_{1,h}}(a) = h(a) f_{\Delta_1}(a).$$

Therefore, let us write the law of the scaled jumps conditional on the random biggest jump $\Delta_{1,h}$ as

$$\left(\frac{\Delta_{k+1}}{\Delta_1}\right)_{k\geq 1} \mid \Delta_1 = \Delta_{1,h} \sim G_{\Delta_{1,h}}.$$

Let us now define the scaled processes, a random measure whose $(0, 1)$-valued jumps are distributed as $G_{\Delta_{1,h}}$.

Definition 2.2.2 (SP)**.** A *Scaled Process (SP)* is defined as

$$\tilde{\mu}_{\Delta_{1,h}} = \sum_{k\geq 1} \tau_k \delta_{\tilde{x}_{k+1}}, \quad \text{where} \quad (\tau_k)_{k\geq 1} \sim G_{\Delta_{1,h}}.$$

We further write

$$\tilde{\mu}_{\Delta_{1,h}} \sim \mathrm{SP}(\nu, h)$$

to denote the distribution.

In other words, scaled processes are random measures whose jumps are $(0,1)$-valued variables distributed as the scaled ordered jumps of the starting CRM $\tilde{\mu}$ conditional on the random and re-scaled bigger jump $\Delta_{1,h}$. Note that, as recalled in (2.3), the distribution of the biggest jump Δ_1 is known. Therefore, a scaled process is completely defined by the Lévy intensity $\nu(\mathrm{d}s, \mathrm{d}x) = \rho(s)\mathrm{d}s\alpha(\mathrm{d}x)$ of $\tilde{\mu}$ and the scaling function $h(\cdot)$. Suitable choices for the transition kernel $\rho(\cdot)$ and the function $h(\cdot)$ can have appealing analytical properties. In particular, stable scaled processes (S-SP) are a subclass of SP introduced in James et al. (2015) and they are characterized by the σ-stable CRM arising from $\tilde{\mu}$, as described in Kingman (1975) and summarized in the following definition.

Definition 2.2.3 (S-SP)**.** Let us consider $\tilde{\mu} \sim \mathrm{SP}(\nu, h)$ as in Definition 2.2.2. If

$$\rho(s) = \sigma s^{-1-\sigma}$$

for some $\sigma \in (0,1)$, we say that $\tilde{\mu}$ is distributed as a *Stable Scaled Process (S-SP)*. We further write

$$\tilde{\mu}_{\Delta_{1,h}} \sim \text{S-SP}(\nu_\sigma, h)$$

to denote the distribution.

As a noteworthy example of S-SPs we will focus on the Stable-Beta Scaled Processes (SB-SP). A SB-SP is a process introduced in Camerlenghi et al. (2022), and it can be obtained by specifying a suitable specification of the non-negative function h. The idea is to define a specific function h in order to obtain a more tractable density for the re-scaled jump. This approach simplifies the further analysis, leading to useful theoretical results, which will be detailed and explained in the following sections, where the practical and theoretical implications of this re-scaling will be explored. This is summarized in the following definition.

Definition 2.2.4 (SB-SP)**.** Let us consider $\tilde{\mu} \sim \text{S-SP}(\nu_\sigma, h)$ as in Definition 2.2.3. If the function h can be written as

$$h_{c,\beta}(a) = \frac{\beta^{c+1}}{\Gamma(c+1)} a^{-c\sigma} \exp\{-(\beta - 1)a^{-\sigma}\}$$

for some positive constants $c, \beta > 0$, the process is called *Stable-Beta Scaled Process (SB-SP)*. We further write

$$\tilde{\mu}_{\Delta_{1,h_{c,\beta}}} \sim \text{SB-SP}(\nu_\sigma, h_{c,\beta})$$

to denote the distribution.

The aim of the next sections is to construct a nonparametric model assuming scaled processes as priors for exchangeable survival time. A further set of results and observations concerning scaled processes is reported in Section 2.A.

2.3 Scaled processes in survival analysis

The aim of this section is to present a general theory for scaled processes as nonparametric priors in survival analysis. In particular, in Section 2.3.1 we introduce the survival model and the result on the prior distribution, while Section 2.3.2 is dedicated to the Bayesian analysis on the posterior distribution.

2.3.1 SP survival model

The aim of this section is to introduce our survival model for the cumulative hazard function under a SP prior, thus showing the expected value a priori. Note that $\mathbb{X} = \mathbb{R}^+$, and let us assume to be provided with a set of survival times $(T_1, \ldots, T_N)$, along with the corresponding right censoring variables $(\Theta_1, \ldots, \Theta_N)$, where

$$\Theta_i = \begin{cases} 1 & \text{if } T_i \leq C_i \\ 0 & \text{otherwise} \end{cases},$$

and let $X_i = \min(T_i, \Theta_i)$ be the ith observation. In the sequel we will denote by $\boldsymbol{D}$ the vector containing all the observations X_i's and the corresponding variables Θ_i's, namely

$$\boldsymbol{D} := ((X_i, \Theta_i) : \ i = 1, \ldots, N),$$

where N is the number of observations. The observations, which are $\mathbb{R}^+$-valued survival times, are assumed to come from an infinite array of exchangeable random variables.
The idea behind this section is to put a SP prior on the cumulative hazard function H. Let us therefore consider the following model:

$$
\begin{aligned}
(X_1,\ldots,X_N) \mid \tilde{F} \sim & \tilde{F} \\
\tilde{F}(t) = & 1 - \prod_{s\in(0,t]} \left\{1 - \mathrm{d}\tilde{H}_s\right\} \quad \text{for any } t \geq 0 \\
\tilde{H}_t = & \tilde{\mu}_{\Delta_{1,h}}(0,t] \quad \text{for any } t \geq 0 \\
\tilde{\mu}_{\Delta_{1,h}} \sim & \mathrm{SP}(\nu, h).
\end{aligned} \tag{2.4}
$$

Note that the second expression in (2.4) is the well-known relations between distribution and hazard functions and it is written in terms of the product integral: if $G : \mathbb{R}^+ \to \mathbb{R}^+$ is a non-decreasing and right-continuous function and $0 = t_0 < t_1 < \ldots < t_m$ is a partition such that

$$
\lim_{m\to\infty} \left[\max_{i\geq 1} \mid t_i - t_{i-1} \mid\right] = 0,
$$

the product integral is the function

$$
t \mapsto \prod_{s\in(0,t]} \{1 + \mathrm{d}G_s\} = \lim_{m\to\infty} \prod_{i=1}^{m} \{1 + G(t_i) - G(t_{i-1})\}.
$$

In the following, we will state and prove the main results, i.e., the expression for the marginal and posterior distributions according to the model in (2.4). Let us first fix the notation. Let us denote by $X_1^*, \ldots, X_K^*$ the distinct observations among $X_1, \ldots, X_N$; let us also assume that they are ordered, i.e., $X_1^* < X_2^* < \ldots < X_K^*$. Moreover, let us write

$$
\Theta_r^* = \max_{i:X_i = X_r^*} C_i \quad \text{for each} \quad r = 1, \ldots, K,
$$

i.e., the variable Θ_r^* indicates whether there is at least a exact observation equal to X_r^*. Let K^* be the number of distinct observations X_r^*'s having $\Theta_r^* = 1$; note that $K^* \leq N$. Moreover, let

$$
n_r = \sum_{i=1}^{N} \mathbb{1}_{(X_i = X_r^*, \Theta_i = 1)} \quad \text{and} \quad n_r^c = \sum_{i=1}^{N} \mathbb{1}_{(X_i = X_r^*, \Theta_i = 0)},
$$

be the counting processes indicating the number of exact and censored observations equal to X_r^* respectively, and let

$$
\bar{n}_r = \sum_{i=1}^{N} \mathbb{1}_{(X_i >= X_r^*, \Theta_i = 1)} \quad \text{and} \quad \tilde{n}_r^c = \sum_{i=1}^{N} \mathbb{1}_{(X_i >= X_r^*, \Theta_i = 0)},
$$

be the cumulative counting processes indicating the number of exact and censored observations greater or equal to X_r^* respectively. Finally, let us define the at-risk process $Y(x)$ as

$$Y(x) = \sum_{i=1}^{N} \mathbb{1}_{[x,\infty)}(X_i).$$

It is now possible to obtain the expression of the prior expected value for the cumulative hazard $\tilde{H}$ in model (2.4), as summarized in the following proposition; the proof is reported in Section 2.B.

Proposition 2.3.1. *Let us consider the model* (2.4)*. Then the prior expected value for* $\tilde{H}$ *under a SP*(ν, h) *prior is*

$$\mathbb{E}\left[\tilde{H}_t\right] = \alpha((0,t]) \cdot \int_{\mathbb{R}^+} \int_0^1 sy\rho(ys)\mathrm{d}s f_{\Delta_{1,h}}(y)\mathrm{d}y$$

for any $t \geq 0$*, where* $f_{\Delta_{1,h}}(y)$ *is the density of the jump* $\Delta_{1,h}$*.*

Note that the expected cumulative hazard function $\tilde{H}$ is proportional to the measure α; such as in the beta process, introduced in Hjort (1990), the base measure of the CRM $\tilde{\mu}$ can be interpreted as the prior guess on the function $\tilde{H}$. In particular, under a Beta prior, the prior guess of $\tilde{H}$ is exactly equal to α.

2.3.2 Bayesian analysis of SP survival model

The aim of this section is to analyze the marginal and posterior distributions under the survival model 2.4. First of all, it is possible to obtain the expression of the marginal distribution of the data arising from model (2.4), as summarized by the following theorem, whose proof is reported in Section 2.B.

Theorem 2.3.1. *The distribution of the data* $\boldsymbol{D}$ *arising from the model* (2.4) *is infinitesimally equal to*

$$\begin{aligned}&\int_{\mathbb{R}^+} y^{K^*} \prod_{r=1:\Theta_r^*=1}^{K} \left[\alpha(\mathrm{d}X_r^*) \int_0^1 s^{n_r}(1-s)^{\bar{n}_{r+1}+\tilde{n}_r^c}\rho(ys)\mathrm{d}s\right] \times \\ &\times exp\left\{-y\int_{\mathbb{R}^+}\int_0^1 \left(1-(1-s)^{Y(x)}\right)\rho(ys)\mathrm{d}s\alpha(\mathrm{d}x)\right\} f_{\Delta_{1,h}}(y)\mathrm{d}y,\end{aligned} \tag{2.5}$$

where $f_{\Delta_{1,h}}(y)$ *is the density of the jump* $\Delta_{1,h}$.

Note how from the results reported in Proposition 2.3.1 and Theorem 2.3.1 it is evident the dependence of the expressions on the Lévy intensity $\nu(\mathrm{d}s, \mathrm{d}x) = \rho(s)\mathrm{d}s\alpha(\mathrm{d}x)$ of $\tilde{\mu}$ and on the distribution of the scaling function h. The expressions reported in these results can be therefore semplified via a suitable choise of ν and h. A notable example of a specific scaled process is the SB-SP introduced in Section 2.2; the next section will be dedicated to analyze how the expression of the generic results for scaled processes become sensibly easier in the case of this particular process.
The expression of the posterior distribution from (2.4) under a $\mathrm{SP}(\nu, h)$ prior can be described via a hierarchical construction. The result is summarized in the following theorem.

Theorem 2.3.2. *Let us consider the model in* (2.4)*. The posterior distribution of* $\tilde{\mu}_{\Delta_{1,h}} \sim$ *SP*(ν, h) *can be described as*

$$\begin{aligned} \tilde{\mu}_{\Delta_{1,h}} \mid \boldsymbol{D}, \Delta_{1,h} &= \tilde{\mu}_c^* + \sum_{r:\Theta_r^*=1}^{K} J_r \delta_{X_r^*} \\ \Delta_{1,h} \mid \boldsymbol{D} &\sim F_{\Delta_{1,h}|\boldsymbol{D}}, \end{aligned} \tag{2.6}$$

where:

i. $\tilde{\mu}_c^* \sim$ *CRM*(ν^*)*, where*

$$\nu^*(\mathrm{d}s, \mathrm{d}x) = (1-s)^{Y(x)} \Delta_{1,h} \rho\left(\Delta_{1,h} s\right) \mathbb{1}_{(0,1)}(s) \mathrm{d}s\alpha(\mathrm{d}x),$$

ii. the jumps J_r*'s, as* $r = 1, \ldots, K$ *and* $\Theta_r^* = 1$*, have density proportional to*

$$s^{n_r}(1-s)^{\bar{n}_{r+1}+\tilde{n}_r^c} \Delta_{1,h} \rho\left(\Delta_{1,h} s\right) \mathbb{1}_{(0,1)}(s) \mathrm{d}s,$$

iii. $F_{\Delta_{1,h}|\boldsymbol{D}}$ *is the posterior distribution of the jump* $\Delta_{1,h}$.

The previous hierarchical expression is useful when the distribution of the jump $\Delta_{1,h}$, in particular when it comes to exploit the posterior representation to construct a conditional sampling algorithm; this will be the case of the SB-SP, discussed in the next section. On the other hand, in the case of the SB-SP it is also possible to obtain a straightforward representation of the posterior distribution without relying on hierarchical constructions, thus obtaining appealing analytical properties that will be as well described in the following section.

The last result of this section is the expression of the posterior estimator for the random cumulative hazard function under a quadratic loss assuming model (2.4). This is summarized in the following proposition, proved in Section 2.B.

Proposition 2.3.2. *Let us consider the model in* (2.4). *Then the posterior estimator under a quadratic loss for the cumulative hazard* $\tilde{H}$ *is*

$$\mathbb{E}\left[\tilde{H}_t \mid \boldsymbol{D}\right] = \int_{\mathbb{R}^+} \left[\int_0^t \int_0^1 s(1-s)^{Y(x)} y\rho(ys)\mathrm{d}s\alpha(\mathrm{d}x) + \right.$$
$$\left. + \sum_{r=1;\Theta_r^*=1,X_r^*\leq t}^{K} \frac{\int_0^1 s^{n_r+1}(1-s)^{\bar{n}_{r+1}+\tilde{n}_r^c}\rho\left(ys\right)\mathrm{d}s}{\int_0^1 t^{n_r}(1-t)^{\bar{n}_{r+1}+\tilde{n}_r^c}\rho\left(yt\right)\mathrm{d}t}\right] f_{\Delta_{1,h}|\boldsymbol{D}}(y)\mathrm{d}y,$$

where $f_{\Delta_{1,h}|\boldsymbol{D}}(y)$ *is the posterior density for the jump* $\Delta_{1,h}$.

Remark 2.3.1. Note that the ratio

$$\frac{\int_0^1 s^{n_r+1}(1-s)^{\bar{n}_{r+1}+\tilde{n}_r^c}\rho\left(ys\right)\mathrm{d}s}{\int_0^1 t^{n_r}(1-t)^{\bar{n}_{r+1}+\tilde{n}_r^c}\rho\left(yt\right)\mathrm{d}t}$$

from the expression in Proposition 2.3.2 is the expected value of a variable with density

$$s^{n_r}(1-s)^{\bar{n}_{r+1}+\tilde{n}_r^c}\rho\left(ys\right)\mathbb{1}_{(0,1)}(s).$$

The expression from Proposition 2.3.2 is clearly obtained marginalizing the posterior jump $\Delta_{1,h} \mid \boldsymbol{D}$ out from the expect value of a CRM with both a continuous and discrete component. More details of this construction are reported in Sections 2.A and 2.B. Note also that the expression can be semplified when the posterior distribution $\Delta_{1,h} \mid \boldsymbol{D}$ is known, as in the case of a SB-SP. The next section will be focused on an extensive analysis of the SB-SP as nonparametric prior in survival analysis, as a particular case of scaled process.

2.4 Stable-beta scaled priors

The results discussed in the previous section provide a general theory for scaled processes as nonparametric priors for cumulative hazard functions of exchangeable survival times. The expressions obtained show the dependence of the results on the choice of the base CRM $\tilde{\mu}$ and the re-scaling function h.

The aim of this section is to study the survival model (2.4) under a particular scaled prior, namely a stable-beta scaled process. There are different appealing features of a SB-SP as nonparametric prior in survival analysis. First of all, the prior and posterior properties can be derived from the general theory for scaled priors presented in Section 2.3, and the main results (e.g., marginal and posterior distributions) are available in a closed and relatively simple form; the study of these properties will be discussed in Section 2.4.1. Moreover, SB-SPs can also be derived as a Gamma-mixture of the well known Indian Buffet Process (IBP), therefore the prior and posterior results can be derived from the theory about IBP (see Teh and Gorur (2009)); this link will be shown and discussed in Section 2.4.2. Finally, thanks to a suitable hierarchical construction of the posterior SB-SPs, the process can be simulated via stanadrd sampling algorithms; this topic will be described in Section 2.5.

2.4.1 Main results

The aim of this section is to specialize the results obtained in the previous section to the case of a SB-SP prior, which is characterized by a specific expression for both $\tilde{\mu}$ and h, which lead to a close and analytically tractable expressions. An appealing feature of the stable-beta scaled prior is that it can be represented in different ways. The first result of this section shows how a SB-SP prior in model (2.4) can be also represented as a negative binomial point process, as described for example in Griffiths et al. (2024). Let us first recall the following definition (see Gregoire (1984); Griffiths et al. (2024)).

Definition 2.4.1 (negative binomial random measure)**.** Let $\rho(s \mid x)$ be a transition kernel, let α be a random probability measure and let $r > 0$. A *negative binomial random measure* $\tilde{\mu}$ is characterized by a Laplace functional equal to

$$\mathbb{E}\left[e^{-\tilde{\mu}(g)}\right] = \left(1 + \int_{\mathbb{R}^+ \times \mathbb{R}^+} \left(1 - e^{-sg(x)}\right) \rho(s \mid x) \mathrm{d}s \alpha(\mathrm{d}x)\right)^{-r}$$

for any bounded non-negative measurable function g on $\mathbb{R}^+$. Let us write

$$\tilde{\mu} \sim \mathrm{BN}(r, \rho, \alpha).$$

Let us now state the aforementioned result. The proof is reported in Section 2.C.

Proposition 2.4.1. *Let $\tilde{\mu}_{\Delta_{1,h_{c,\beta}}} \sim$ SB-SP$(\nu_\sigma, h_{c,\beta})$. Then*

$$\tilde{\mu}_{\Delta_{1,h_{c,\beta}}} \sim BN\left(c+1, \frac{\sigma}{\beta} s^{-1-\sigma}, \alpha\right).$$

Proposition 2.3.1 and Theorem 2.3.1 provide general results for the expression of the prior expected value, as well as the marginal and distribution, according to model (2.4) under a SP prior. The results about the SB-SP priors can be obtained simply specializing

the aforementioned results, specifying the Lévy intensity of the generating CRM $\tilde{\mu}$ and the expression of the re-scaling factor h.
Let us now summarize the expected value of the cumulative hazard H under SB-SP priors. The result is proved in Section 2.C.

Corollary 2.4.1. *Let us consider the model* (2.4) *with s SB-SP prior. Then the expected value for $\tilde{H}$ is equal to*

$$\mathbb{E}\left[\tilde{H}_t\right] = \frac{\sigma(c+1)}{(1-\sigma)\beta}\alpha((0,t]).$$

for any $t \geq 0$.

Remark 2.4.1. From the Corollary 2.4.1, since

$$\alpha((0,t]) \to 1 \quad \text{as} \quad t \to \infty,$$

it is possible to observe that the prior guess for the limit value of $\tilde{H}_t$ is

$$\frac{\sigma(c+1)}{(1-\sigma)\beta}.$$

Under our model, this is the prior guess for the cure rate at the cumulative hazard scale.

Let us now state the expression for the marginal distribution under a SB-SP prior as a corollary of Theorem 2.3.1. The proof is reported in Section 2.C.

Corollary 2.4.2. *Let us consider the model in* (2.13) *with a SB-SP prior. Then, the distribution of the data $\boldsymbol{D}$ arising from this model is infinitesimally equal to*

$$\frac{\Gamma\left(K^*+c+1\right)\sigma^{K^*}\beta^{c+1}}{\Gamma(c+1)\left(\beta+\sigma\eta_\sigma\right)^{K^*+c+1}} \times \left[\prod_{r=1:\Theta_r^*=1}^{K} \alpha(\mathrm{d}X_r^*) B\left(n_r-\sigma, \bar{n}_{r+1}+\tilde{n}_r^c+1\right)\right], \tag{2.7}$$

where

$$\eta_\sigma = \sum_{r=1}^{K} \alpha((X_{r-1}^*, X_r^*]) \sum_{k=1}^{N_r^*} (-1)^{k-1} \binom{N_r^*}{k} \frac{1}{k-\sigma} \quad \text{and} \quad N_r^* = \sum_{g=r}^{K} n_g^*,$$

and n_g^* is the number of observations equal to X_g^*. Moreover, the posterior distribution of $\Delta_{1,h_{c,\beta}}$ is such that

$$\Delta_{1,h_{c,\beta}}^{-\sigma} \mid \boldsymbol{D} \sim Gamma\left(K^*+c+1, \beta+\sigma\eta_\sigma\right),$$

i.e.,

$$f_{\Delta_{1,h_{c,\beta}}|\boldsymbol{D}}(y) = \frac{\sigma\,(\beta+\sigma\eta_\sigma)^{K^*+c+1}}{\Gamma(K^*+c+1)} y^{-(K^*+c+1)\sigma-1} e^{-(\beta+\sigma\eta_\sigma)y^{-\sigma}}.$$

Remark 2.4.2. Note that in the expression (2.7) the product is performed over the distinct observations X_r^* such that $\Theta_r^* = 1$, i.e., over the distinct observations with which at least one exact observation coincides. For this reason, it follows that in the product in (2.7) $n_r \geq 1$, so $n_r - \sigma \geq 0$ for each r, so the Beta function is well-defined.

The importance of the result reported in the previous corollary is twofold. First of all, it provides a closed and tractable expression for the marginal distribution, that allows to sample the value of the parameters σ, c and β from their full-conditional distributions via a suitable MCMC simulation algorithm. Moreover, it provides the posterior distribution of the jump $\Delta_{1,h_{c,\beta}}$; this result, along with the hierarchical expression for the posterior distribution of a generic SP prior provided in Theorem 2.3.2, reduces the problem of sampling from the posterior of SPs to the problem of sampling from a CRM conditionally to a variable with known distribution, i.e., the posterior jump $\Delta_{1,h_{c,\beta}} \mid \boldsymbol{D}$. Although the aforementioned hierarchical expression is useful for computational purposes, the following result provides a closed form for the posterior distribution of a SB-SP prior, which allows us to capture some analytical properties of interest. The proof is reported in Section 2.C.

Corollary 2.4.3. *Let us consider the model in* (2.4) *with a SB-SP prior. Then the posterior distribution can be written as:*

$$\tilde{\mu}_{\Delta_{1,h_{c,\beta}}} \mid \boldsymbol{D} = \tilde{\mu}_c^* + \sum_{r:\Theta r^*=1}^{K} J_r \delta_{X_r^*}, \tag{2.8}$$

where:

i. $\tilde{\mu}_c^*$ *is a negative binomial random measure, i.e.,* $\tilde{\mu}_c^* \sim BN(K^*+c+1, \rho^*, \alpha)$, *where*

$$\rho^*(s \mid x)\mathrm{d}s = \frac{\sigma}{\beta+\sigma\eta_\sigma}(1-s)^{Y(x)} s^{-1-\sigma}\mathrm{d}s,$$

ii. the jumps J_r*'s are distributed as a*

$$Beta(n_r - \sigma, \bar{n}_{r+1} + \tilde{n}_r^c + 1).$$

Remark 2.4.3. Note that the sum in (2.8) is performed over the distinct observations X_r^* such that $\Theta_r^* = 1$, i.e., over the distinct observations with which at least one exact observation coincides. For these indices r's, $n_r \geq 1$, so $n_r - \sigma \geq 0$. It follows that the Beta distributions

$$\text{Beta}(n_r - \sigma, \bar{n}_{r+1} + \tilde{n}_r^c + 1)$$

are well-defined.

The importance of the posterior representation summarized in Corollary 2.4.3 lies in the fact that it shows the quasi-conjugacy property of the SB-SPs. Combining the results of Proposition 2.4.1 and Corollary 2.4.3, in fact, it emerges that SB-SPs forms a family of quasi-conjugate processes in the class of negative binomial point processes. On the other hand, they are not conjugate in the scaled processes class. The posterior process represented in Corollary 2.4.3, in fact, is not a scaled process anymore.

Let us now state the last result of the section, that reports a closed form expression of the estimator for the cumulative hazard function under a quadratic loss, under the SB-SP model as in (2.4). The proof is reported in Section 2.C.

Corollary 2.4.4. *Let us consider the model in* (2.4) *with a SB-SP prior, i.e., with*

$$\tilde{\mu}_{\Delta_{1,h_{c,\beta}}} \sim \mathit{SB\text{-}SP}(\nu_\sigma, h_{c,\beta}).$$

Then the estimator for the cumulative hazard under a quadratic loss is

$$\mathbb{E}\left[\tilde{H}_t \mid \boldsymbol{D}\right] = \frac{\sigma(K^* + c + 1)}{\beta + \sigma\eta_\sigma}\eta_t + \sum_{r=1;\Theta_r^*=1,X_r^*\leq t}^{K} \frac{n_r - \sigma}{\bar{n}_{r+1} + \tilde{n}_r^c + n_r + 1 - \sigma},$$

where

$$\eta_t := \sum_{r=1}^{K_t} \alpha\left((X_{r-1}^*, X_r^* \wedge t]\right) B(1-\sigma, N_r^* + 1), \quad \textit{and} \quad K_t = \min\{g : X_g^* \geq t\}$$

and η_σ is defined in Corollary 2.4.2.

Remark 2.4.4. Note that

$$\lim_{t\to\infty} \mathbb{E}\left[\tilde{H}_t \mid \boldsymbol{D}\right] = \frac{\sigma(K^* + c + 1)}{\beta + \sigma\eta_\sigma}\eta_\infty + \sum_{r=1;\Theta_r^*=1}^{K} \frac{n_r - \sigma}{\bar{n}_{r+1} + \tilde{n}_r^c + n_r + 1 - \sigma},$$

where

$$\eta_\infty := \sum_{r=1}^{K} \alpha\left((X_{r-1}^*, X_r^*]\right) B(1-\sigma, N_r^* + 1),$$

that is, the posterior estimate for the cure rate at the cumulative hazard scale.

The previous result is useful in order to construct a marginal algorithm for the estimation of the posterior distribution of the cumulative hazard function. More details about this topic will be provided in Section 2.5. Also note how the results presented in this section, although concerning the cumulative hazard function, can naturally be extended to the survival function scale. Section 2.E dedicated to this purpose.

2.4.2 Gamma-mixtures of Indian Buffet Processes

Up to this point we have shown how SB-SPs can be obtained equivalently in different ways. In particular, in Section 2.2 we showed how they are a particular case of scaled processes, therefore inheriting their prior and posterior properties described in Section 2.3. On the other hand, in Section 2.4.1 we showed how they are part of the family of negative binomial point processes, within which they show a quasi-conjugate behaviour. Another useful hierarchical construction of the scaled processes, already cited in the previous sections and inherited by the SB-SPs, is described in Section 2.A and shows how they are a CRM conditionally on the jump $\Delta_{1,h}$.
The aim of this section is to show a further construction for this family of processes, hence providing a useful link to the well known 3-parameter Indian Buffet Process, also known as *Stable Beta Process* and introduced in Teh and Gorur (2009). Let us first recall the definition.

Definition 2.4.2 (IBP)**.** A random measure $\tilde{\mu}$ induces a *Stable Beta Process*, or *3-parameter Indian Buffet Process (IBP)*, if

$$\tilde{\mu} \sim \text{CRM}(\nu) \quad \text{and}$$

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

where $\gamma > 0$ is a mass parameter, $\theta > -\sigma$ is a concentration parameter, $\sigma \in [0, 1)$ is a stability exponent and α is a probability measure. Let us write

$$\tilde{\mu} \sim \text{IBP}(\gamma, \theta, \sigma).$$

Since a IBP is a CRM, the theory for the IBP as nonparametric prior for the cumulative hazard function is entirely induced by the general theory for the CRMs. In particular, the survival model is the following:

$$
\begin{aligned}
(X_1, \ldots, X_N) \mid \tilde{F} \sim & \tilde{F} \\
\tilde{F}(t) = & 1 - \prod_{s \in (0,t]} \left\{ 1 - \mathrm{d}\tilde{H}_s \right\} \quad \text{for any } t \geq 0 \\
\tilde{H}_t = & \tilde{\mu}(0,t] \quad \text{for any } t \geq 0 \\
\tilde{\mu} \sim & \mathrm{IBP}(\gamma, \theta, \sigma).
\end{aligned}
\tag{2.9}
$$

The link between the model in (2.4) and (2.9) is provided by the following result, which shows how a SB-SP can be obtained as a Gamma-mixture of IBPs with $\theta = 1 - \sigma$. The proof is reported in Section 2.C.

Proposition 2.4.2. *Let us consider the following hierarchical definition:*

$$
\begin{aligned}
\tilde{\mu} \mid \gamma \sim & IBP(\gamma, 1-\sigma, \sigma), \\
\gamma \sim & Gamma\left(c+1, \frac{\beta(1-\sigma)}{\sigma}\right),
\end{aligned}
\tag{2.10}
$$

where $\sigma > 0$ *and* $\beta, c > 0$*. Then*

$$
\tilde{\mu} \sim \textit{SB-SP}(\nu_\sigma, h_{c,\beta}),
$$

where ν_σ *is the intensity of the IBP defined in Definition 2.4.2 and* $h_{c,\beta}$ *is the re-scaling factor of a SB-SP defined in Definition 2.2.4.*

Note that the main results concerning IBP as nonparametric prior in model (2.9) can be obtained as a direct application of the general results for CRMs (see, for example, Ferguson and Phadia (1979); Lijoi and Prünster (2010)). Section 2.D reports the expression for the expected value of the cumulative hazard function under a IBP prior in model (2.9), and its posterior estimator under a quadratic loss, as well as the marginal and posterior distributions. It follows that, according to Proposition 2.4.2, all the main results stated in Section 2.4.1 can also be obtained starting from the general results for the IBP, assuming $\theta = 1 - \sigma$ and integrating out the gamma variable γ. Moreover, note how the link between IBP and SB-SP is useful to underline the connection between the SB-SP and the beta process (see Hjort (1990)), since a beta process with parameters θ and α can be recovered as a IBP with parameters $\gamma = 1$, and $\sigma = 0$. Finally, note that the hierarchical definition of a SB-SP reported in Proposition 2.4.2 leads to an alternative strategy for constructing a sampling algorithm, since it reduces the problem of approximating a SB-SP to the problem of approximating a CRM, i.e., approximating a IBP with a random parameter γ. Note that this observations holds true even for the posterior distribution of a SB-SP, since the IBP in model (2.9) is conjugate.

2.5 Sampling Algorithm

The aim of this section is to describe a sampling algorithm that allows to sample from the posterior distribution of the survival model (2.4) under a SB-SP prior. In particular, we are interested in the estimation of the survival function arising from this model; let us denote by $\tilde{S}_t$ such random survival function evaluated at time $t \geq 0$. Note that all the results reported in Section 2.4.1 refer to the cumulative hazard function, but that they can all be adapted to the survival function; obviously, the expression of the marginal distribution does not change. Please refer to Section 2.E for more details about these results on the survival function arising from model (2.4). In this section we introduce two different strategies: a marginal algorithm, which obtain a survival function marginalizing out the random measures; and a conditional algorithm, which allow to simulate the trajectories of the posterior survival functions.
Note that the prior process is entirely defined by the measure α and the parameters σ, β and c. Let us therefore assume that α is a uniform measure

$$\text{Uniform}(0, \tau), \quad \tau \in \mathbb{R}^+,$$

where τ is a suitable positive value. Finally, in order to rely on a full Bayesian approach, let us assume gamma priors on the parameters $\beta, c > 0$, and a beta prior on the parameter $\sigma \in (0, 1)$.

2.5.1 Marginal algorithm

A marginal algorithm can be obtained exploiting the posterior estimator of the survival function $\tilde{S}_t \mid \boldsymbol{D}$ under a quadratic loss arising from SB-SP model. Let us report the expression of the estimator for each time $t \geq 0$ (please refer to Section 2.E for the details):

$$\mathbb{E}\left[\tilde{S}_t \mid \boldsymbol{D}\right] = \left(\frac{\beta + \sigma\eta_\sigma}{\beta + \sigma\eta_\sigma + \sigma\eta_t}\right)^{K^*+c+1} \times \prod_{r=1;\Theta_r^*=1,X_r^*\leq t}^{K} \frac{\bar{n}_{r+1} + \tilde{n}_r^c + 1}{\bar{n}_{r+1} + \tilde{n}_r^c + n_r + 1 - \sigma}, \quad (2.11)$$

where

$$\eta_t := \sum_{r=1}^{K_t} \alpha\left((X^*_{r-1}, X^*_r \wedge t]\right) B(1-\sigma, N^*_r + 1), \quad \text{and} \quad K_t = \min\{g : X^*_g \geq t\}.$$

A numerical approximation of $\mathbb{E}\left[\tilde{S}_t \mid \boldsymbol{D}\right]$ can be obtained via a suitable MCMC algorithm; to this end, the full conditional distributions of the parameters σ, β, c must be identified. The expressions of the full-conditional distributions derive directly from the marginal distribution reported in Corollary 2.4.2, and they are reported in Section 2.F. At each step of the marginal algorithm the model parameters are sampled according to their full-conditional distributions; then, the corresponding estimated survival function can be evaluated on a time-grid according to the expression (2.11). Finally, the posterior estimate can be obtained as a Monte-Carlo approximation exploiting the samples obtained with the marginal algorithm.

2.5.2 Conditional algorithm

The aim of this section is to describe a conditional algorithm that generates trajectories from the posterior distribution of the survival function assuming the model (2.4) under a SB-SP prior, with the assumption about the measure α and the model parameters described before. Note that this algorithm is useful since it allows to estimate the actual posterior distribution of the survival function under our model, as well as credible intervals for the estimated quantities.
The goal can be achieved relying on a hierarchical construction derived from Theorem 2.3.2 in the case of a stable-beta scaled prior. In particular, the posterior random survival function for each $t \geq 0$ can be represented as

$$\tilde{S}_t \mid \boldsymbol{D} = e^{-\tilde{\mu}^F_{\Delta_{1,h_{c,\beta}}}((0,t])|\boldsymbol{D}},$$

where the posterior distribution of the random measure $\tilde{\mu}^F_{\Delta_{1,h_{c,\beta}}}$, i.e., the distribution of $\tilde{\mu}^F_{\Delta_{1,h_{c,\beta}}} \mid \boldsymbol{D}$, can be hierarchically described as

$$\tilde{\mu}^F_{\Delta_{1,h_{c,\beta}}} \mid \boldsymbol{D}, (\Delta_{1,h_{c,\beta}} = y) = \tilde{\mu}^{*,F}_c + \sum_{r:\Theta^*_r=1}^{K} J^F_r \delta_{X^*_r}$$

$$\Delta^{-\sigma}_{1,h_{c,\beta}} \mid \boldsymbol{D} \sim \text{Gamma}\left(K^* + c + 1, \beta + \sigma\eta_\sigma\right).$$

In the previous expression:

i. $\tilde{\mu}_c^{*,F} \sim \text{CRM}(\nu_F^*)$, with Lévy intensity

$$\nu_F^*(\mathrm{d}s, \mathrm{d}x) = \sigma y^{-\sigma} e^{-s(Y(x)+1)}(1-e^{-s})^{-1-\sigma}\mathbb{1}_{(0,\infty)}(s)\mathrm{d}s\alpha(\mathrm{d}x),$$

and

ii. the jumps J_r^F's, as $r = 1, \dots, K$ and $\Theta_r^* = 1$, are such that

$$1 - e^{-J_r^F} \sim \text{Beta}(n_r - \sigma, \bar{n}_{r+1} + \tilde{n}_r^c + 1).$$

Please refer to Section 2.E for more details about the previous construction. Note that, according to this construction, the discrete jumps J_r^F's can be sampled as transformed Beta variables. Moreover, the sampling procedure from the absolutely continuous part can be divided in two parts: first of all, the jump $\Delta_{1,h_{c,\beta}}$ can be sampled from its posterior distribution (a transformed Gamma distributions); then, the problem of approximate the random measure $\tilde{\mu}^F_{\Delta_{1,h_{c,\beta}}}$ reduces to sample from a CRM, i.e., the measure $\tilde{\mu}^F_{\Delta_{1,h_{c,\beta}}}$ conditional on the jump $\Delta_{1,h_{c,\beta}}$. The CRM with the inhomogeneous Lévy intensity as in (i) can be approximated relying on the algorithm described in Wolpert and Ickstadt (1998): relying on the representation of CRMs described in Ferguson and Klass (1972), it approximates the CRM $\tilde{\mu}_c^{*,F}$ as the finite sum

$$\tilde{\mu}_c^{*,F,M} := \sum_{k=1}^{M} \hat{h}'_k \delta_{\hat{x}'_k},$$

for $M > 0$ big enough, where the set of locations $(\hat{x}'_k)_{k=1}^M$ is independently sampled from the base measure $\alpha \sim \text{Uniform}(0, \tau)$. Moreover, as $k = 1, \dots, M$, the kth approximated jump $\hat{h}'_k$ is obtained as the zero of the function in the variable s

$$\xi(s) = \int_s^{\infty} \nu'(v, \hat{x}'_k)\mathrm{d}v - \sigma_k,$$

where

$$\nu_F^*(\mathrm{d}s, \mathrm{d}x) = \nu'(s, x)\mathrm{d}s\alpha(\mathrm{d}x) \quad \text{and} \quad \nu'(s, x) = \sigma y^{-\sigma} e^{-s(Y(x)+1)}(1-e^{-s})^{-1-\sigma}\mathbb{1}_{(0,\infty)}(s),$$

and $\sigma_k = \sum_{b=1}^{k} E_b$, as $E_k \sim \text{Exponential}(1)$ for each k. Since ξ is a decreasing function, we simply assume $\hat{h}'_k = 0$ if $\xi(\epsilon) < 0$, where ϵ is a defined tolerance. Finally, note that at each iteration of the MCMC algorithm the parameters σ, c and β can be sampled from their full-conditional distributions, which are reported in Section 2.F. Note that each trajectory of the survival function sampled following the previous scheme can then be evaluated on a grid of times, and that functional of the posterior distribution can be Monte Carlo-approximated, such as the mean survival curve and the quantiles. The conditional algorithm will be applied in the next section to a set of simulations studies and to a real dataset.

2.6 Illustrations

In the Section 2.5.2 we described an algorithmic strategy in order to sample trajectories of the survival function from the posterior distribution of a stable-beta scaled process, as presented in Corollary 2.4.3. The aim of this section is to apply the aforementioned algorithm to a set of simulated survival times and to a real dataset. Note that the aim of a sampling algorithm for our cure rate model is twofold: first of all, the resulting estimated survival curve should approximate the real survival function; then, the value of the tail of the estimated survival curve should approximate the cure rate arising from the available data. This will be evident in the simulation study, since the data will be sampled from a known distribution having a strictly positive cure rate value, so that the comparison between the estimated and real survival curves are straightforward.
In particular, in Section 2.6.1 we apply the conditional algorithm to 3 different sets of exchangeable survival times, simulated from a geometric distribution. Then, in Section 2.6.2 we exploit the algorithm to find suitable estimations of the survival curves arising from the well-known bone marrow transplantation (BMT) dataset, according to the SB-SP model.

2.6.1 Simulation study

In order to study the estimation of survival curves under the SB-SP model, the aim of this section is apply the conditional algorithm described in Section 2.5.2 to different sets of simulated datasets. In particular, our aim is to compare the estimated survival curve from the SB-SP model with the real one, as well as to the Kaplan-Meier estimation, with an increasing number of available and exchangeable observations.
Let us therefore consider a geometric distribution with parameter $p = 0.3$. For our simulation study we consider a cure rate equal to $p_c = 0.2$. Therefore, the simulation of the set of survival times from this distribution can be performed in two steps: first of all, we sample survival times from the aforementioned Geometric ($p = 0.3$) distribution with probability

0.8; otherwise, we assign a large value to the survival time, i.,e., we sample a survival time associated with a cured patient with probability 0.2. In the second step, we independently apply the censorship mechanism.
Following the previous schema, we sample 3 sets of observations containing an increasing number of observations: 10, 50 and 200. To each resulting survival dataset we apply the conditional algorithm described in Section 2.5.2 under the same assumptions; in particular we assume that the measure α is distributed as

$$\alpha \sim \text{Uniform}(0, \tau),$$

where τ is a positive real number large enough such that all the observations are included in $(0, \tau)$. Moreover, we assume a Beta prior on the parameter σ and two Gamma priors on the parameters c and β. In particular, let us consider the priors

$$\begin{aligned} \sigma &\sim \text{Beta}(0.1, 0.1), \\ \beta &\sim \text{Gamma}(\text{shape} = 2, \text{scale} = 2) \quad \text{and} \\ c &\sim \text{Gamma}(\text{shape} = 2, \text{scale} = 2). \end{aligned}$$

Moreover, for each dataset we also compute the Kaplan-Meier estimation of the survival functions. Note that since the survival function for geometrically distributed survival times is equal to

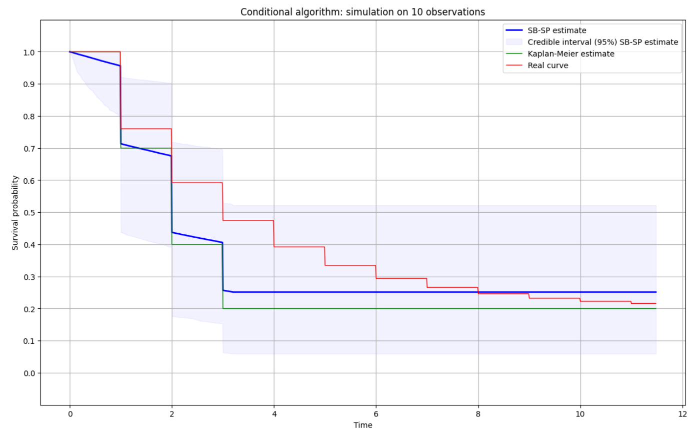
$$S_g(t) = (1-p)^{\lfloor t \rfloor} \quad \text{for each } t \geq 0,$$

it follows that the real survival curve in our simulated framework is equal to

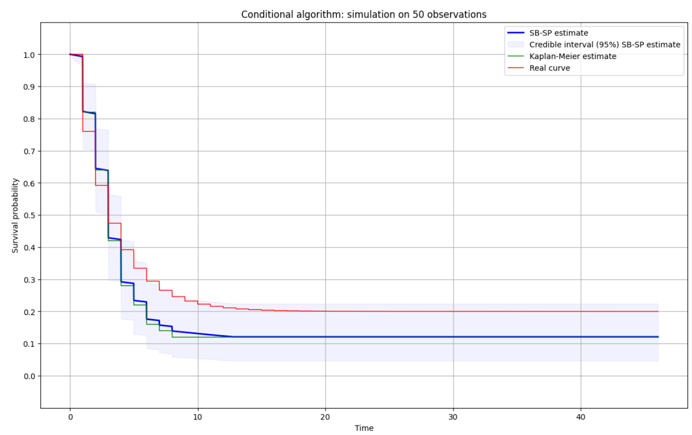
$$S(t) = p_c + (1 - p_c) S_g(t) \quad \text{for each } t \geq 0,$$

where p is the parameter of the geometric distribution and p_c is the cure rate defined above.

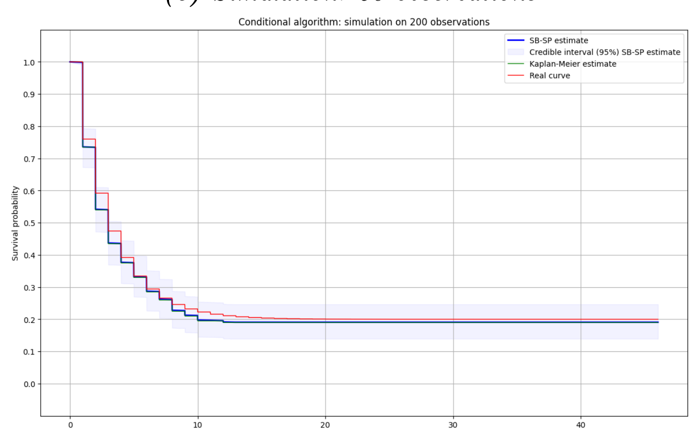
Figure 2.1 reports the results of the simulation study. First of all, note how with an increasing number of observations both the Kaplan-Meier and the SB-SP estimations become closer and closer to the real curve, providing a more precise approximation. Note also how the SB-SP estimation becomes closer to the Kaplan-Meier estimation as the size of the dataset increases. In fact, note how in Figure 2.1a the small number of data induces a greater effect of the prior in estimating the posterior distribution of the survival curve. On the other hand, while the number of observations increases, Figure 2.1b and 2.1c show that the posterior estimates under the SB-SP model relies more and more on the data and the effect of the prior assumptions is less perceptible. Therefore, since the Kaplan-Meier



(a) *Simulation: 10 observations*



(b) *Simulation: 50 observations*



(c) *Simulation: 200 observations*

Figure 2.1: *Plots of estimated survival curve from the SB-SP model (blue line), the Kaplan-Meier estimator (green line) and the real survival curve (red line). The blue area is the the 95% confidence interval of the SB-SP estimate.*

estimator completely relies on the data, this means that as the number of observations increases the Kaplan-Meier and SB-SP estimates tend to get closer to each other, and both to the true survival curve. Finally, as expected, the width of the 95% confidence interval becomes increasingly smaller as the size of the dataset increases, since a greater number of data leads to a reduction in the posterior variance. As far as the cure rate is concerned, its estimation becomes more precise increasing the number of observations, as expected for both SB-SP model and Kaplan-Meier estimator.

2.6.2 Application: BMT dataset

Let us now apply the discussed algorithm to a real dataset. In particular, following Nieto-Barajas and Yin (2008) we consider the well-known bone marrow transplantation (BMT) dataset, which contains information about patients with Hodgkin's desease or with non-Hodgkin's lymphoma and who consequently were treated with procedures that led to destruction of bone marrow. BMT is a procedure that replace the destroyed bone marrow in order to restore its ability to produce functioning blood cells in adequate numbers. There are two types of BMT: allogenic BMT, an infusion of bone marrow from a donor, and autogenic BMT, a reinfusion of the patient's own marrow that had been removed prior to the treatment. The BMT dataset is composed by 43 rows, each row reports data on a different patient; in particular, we are interested in the variables reporting the time to death, or relaps, of the patient (in days) and the possible censorship of the observation, and in the variable reporting wheter the adopted BMT procedure for that patient was allogenic or autogenic. We apply the conditional algorithm independently on the datasets composed by the patients treated with allogenic and autogenic BMT; note that the first dataset is composed by 16 rows while the second dataset is composed by 27 rows. In both our cases, we assume again that the measure α underlying the SB-SP model is distributed as

$$\alpha \sim \text{Uniform}(0, \tau),$$

where τ is a positive real number large enough such that all the observations of the group are included in $(0, \tau)$. In both the applications, we assume again the same priors on the model parameters as in Section 2.6.1. The results of the algorithms are reported in Figure 2.2.

Note how, from Figure 2.2, the survival curves estimated independently in the two groups present a different slope, and how for both groups a cure rate is appreciable, i.e., a strictly positive probability of recovery. In particular, the allogenic BMT group, according to our model, shows a slower decline in the probability of survival than the autogenic BMT group. Furthermore, looking at the right tails of the plots, a patient in the allogenic BMT

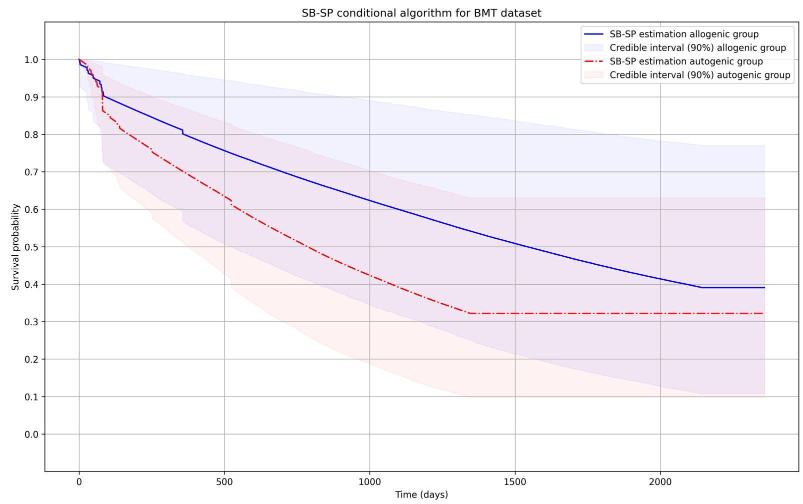


Figure 2.2: *Plots of estimated survival curve from the SB-SP model applied to the group of allogenic BMT patients (blue line) and autogenic BMT patients (dotted red line). The blue and red areas are the the 90% confidence interval of the SB-SP estimate for allogenic and autogenc BMT groups respectively.*

group shows a higher probability of recovery than that of a patient in the autogenic BMT group. The number of data for both groups is relatively small (less than 50, please refer to Section 2.6.1), so for both groups it is possible to see the effect of the priors in the decay of the survival functions, and also how the jumps occur corresponding to the exact and distinct observations, as expected (see Section 2.5.2). Due to the low dimensions of the datasets, note how the confidence intervals are quite large, indicating a large variance of the posterior distribution of the survival functions for both groups.

Bibliography

Amico, M. and Van Keilegom, I. (2018). Cure models in survival analysis. *Annual Review of Statistics and Its Application*, 5:311–342.

Berkson, J. and Gage, R. P. (1952). Survival curve for cancer patients following treatment. *Journal of the American Statistical Association*, 47(259):501–515.

Camerlenghi, F., Favaro, S., Masoero, L., and Broderick, T. (2022). Scaled process priors for Bayesian nonparametric estimation of the unseen genetic variation,. *Journal of the American Statistical Association,*, 119(545):320–331.

Chen, M.-H., Ibrahim, J. G., and Sinha, D. (1999). A new Bayesian model for survival data with a surviving fraction. *Journal of the American Statistical Association*, 94(447):909–919.

Daley, D. J. and Vere-Jones, D. (2008). *An introduction to the theory of point processes*. Probability and Its Applications, a Series of the Applied Probability Trust. New York: Springer, 2nd edition.

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.

Ferguson, T. S. and Phadia, E. G. (1979). Bayesian nonparametric estimation based on censored data. *The Annals of Statistics*, 7(1):163–186.

Gregoire, G. (1984). Negative binomial distributions for point processes. *Stochastic Processes and their Applications*, 16(2):179–188.

Griffiths, R. C., Maller, R. A., and Shemehsavar, S. (2024). A Gibbs sampling scheme for a generalised Poisson-Kingman class.

Hjort, N. L. (1990). Nonparametric Bayes estimators based on beta processes in models for life history data. *The Annals of Statistics*, 18(3):1259–1294.

Ibrahim, J. G., Chen, M., and Sinha, D. (2001a). Bayesian semiparametric models for survival data with a cure fraction. *Biometrics*, 57:383–388.

Ibrahim, J. G., Chen, M., and Sinha, D. (2001b). *Bayesian Survival Analysis*. Springer New York.

James, L. F., Orbanz, P., and Teh, Y. W. (2015). Scaled subordinators and generalizations of the Indian buffet process.

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

Kingman, J. F. C. (1975). Random discrete distributions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 37(1):1–22.

Kuk, A. Y. C. and Chen, C.-H. (1992). A mixture model combining logistic regression with proportional hazards regression. *Biometrika*, 79(3):531–541.

Lijoi, A. and Prünster, I. (2010). Models beyond the Dirichlet process. In *Bayesian nonparametrics*, volume 28 of *Camb. Ser. Stat. Probab. Math.*, pages 80–136. Cambridge Univ. Press, Cambridge.

Maller, R. A. and Zhou, S. (1992). Estimating the proportion of immunes in a censored sample. *Biometrika*, 79(4):731–739.

Maller, R. A. and Zhou, S. (1995). Testing for the presence of immune or cured individuals in censored survival data. *Biometrics*, 51(4):1197–1205.

Nieto-Barajas, L. E. and Yin, G. (2008). Bayesian semiparametric cure rate model with an unknown threshold. *Scandinavian Journal of Statistics*, 35(3):540–556.

Sy, J. P. and Taylor, J. M. (2000). Estimation in a Cox proportional hazards cure model. *Biometrics*, 56(1):227–236.

Teh, Y. and Gorur, D. (2009). Indian buffet processes with power-law behavior. In Bengio, Y., Schuurmans, D., Lafferty, J., Williams, C., and Culotta, A., editors, *Advances in Neural Information Processing Systems*, volume 22. Curran Associates, Inc.

Tsodikov, A. D., G., I. J., and Y., Y. A. (2003). Estimating cure rates from survival data: An alternative to two-component mixture models. *Journal of the American Statistical Association*, 98(464):1603–1078.

Walker, S. and Muliere, P. (1997). Beta-Stacy processes and a generalization of the Pólya-urn scheme. *The Annals of Statistics*, 25(4):1762–1780.

Wolpert, R. L. and Ickstadt, K. (1998). Simulation of Lévy random fields. In *Practical Nonparametric and Semiparametric Bayesian Statistics*, volume 133 of *Lecture Notes in Statistics*, pages 227–242. Springer, New York, NY.

Yakovlev, A. Y. and Tsodikov, A. D. (1996). *Stochastic Models of Tumor Latency and Their Biostatistical Applications*, volume 1 of *Series in Mathematical Biology and Medicine*. World Scientific, New Jersey.

Appendix

This appendix is organized in different sections. In particular, Section 2.A shows some further considerations on scaled processes which are not reported in the main sections of this chapter, with a particular focus on a hierarchical representation of SPs. Sections 2.B and 2.C are dedicated to the proofs of the results reported in Sections 2.3 and 2.4 respectively. Section 2.D delves deeper into what was introduced in Section 2.4.2, explicitly presenting the prior and posterior results of the nonparametric Bayesian survival model under a 3-parameter IBP prior for the cumulative hazard function. Section 2.E presents all the theoretical results obtained for the cumulative hazard function at the survival function scale, specifically deriving the posterior distribution under an SB-SP prior; this result is used in Sections 2.5 and 2.6 to derive the marginal and conditional algorithms. Finally, Section 2.F provides further details on the aforementioned algorithms, hence completing the discussion reported in Section 2.5.

2.A Further considerations on scaled processes

The proofs of the results of this paper make extensive use of a result regarding the behavior of an SP conditional on the re-scaled highest jump. This result, proved in Section S2.1 of Camerlenghi et al. (2022), is reported in the following lemma.

Lemma 2.A.1. *Let* $\mu_{\Delta_{1,h}} \sim SP(\nu, h)$ *be a scaled process governed by the Lévy intensity* $\nu(\mathrm{d}s, \mathrm{d}x) = \rho(s)\mathrm{d}s\alpha(\mathrm{d}x)$ *on* $\mathbb{R}^+ \times \mathbb{X}$. *Then*

$$\begin{aligned} \tilde{\mu}_{\Delta_{1,h}} \mid \Delta_{1,h} &\sim CRM(\nu'), \quad \text{where} \\ \nu'(\mathrm{d}s, \mathrm{d}x) &= \rho'(s)\mathrm{d}s\alpha(\mathrm{d}x), \\ \rho'(s) &= \Delta_{1,h}\rho(\Delta_{1,h}s)\mathbb{1}_{(0,1)}(s), \end{aligned}$$

The law of the jump $\Delta_{1,h}$ characterizes the process, therefore the conditional representation reported in Lemma 2.A.1 allows us to exploit the generic results of the CRMs used as nonparametric priors for exchangeable survival times.
In the case of the SB-SPs, the law of the jump $\Delta_{1,h}$ simplifies to a tractable expression, as summarized by the next lemma.

Lemma 2.A.2. *Let* $\tilde{\mu}_{\Delta_{1,h_{c,\beta}}} \sim SB\text{-}SP(\nu_\sigma, h_{c,\beta})$ *be a stable-beta scaled process. Then the jump* $\Delta_{1,h}$ *has density*

$$f_{\Delta_{1,h_{c,\beta}}}(y) = \frac{\sigma\beta^{c+1}}{\Gamma(c+1)} y^{-1-(c+1)\sigma} e^{-\beta y^{-\sigma}},$$

i.e.,

$$\Delta_{1,h_{c,\beta}}^{-\sigma} \sim \mathit{Gamma}(c+1,\beta).$$

Proof. First of all, applying (2.3) to the transition kernel of a S-SP $\rho(s) = \sigma s^{-1-\sigma}$ we have that

$$f_{\Delta_1}(y) = \sigma y^{-1-\sigma} e^{-y^{-\sigma}}.$$

Since $f_{\Delta_{1,h_{c,\beta}}}(y) = h_{c,\beta}(y) f_{\Delta_1}(y)$ and

$$h_{c,\beta}(y) = \frac{\beta^{c+1}}{\Gamma(c+1)} y^{-c\sigma} \exp\{-(\beta-1)y^{-\sigma}\},$$

the thesis follows. □

Note that the model defined in (2.4) directly induces a nonparametric model for the survival function, since the CRM defined in Lemma 2.A.1 can be transposed at the scale of the survival function with the change of variable

$$s \mapsto 1 - e^{-s}.$$

The following definition summarized this construction.

Definition 2.A.1. Let us consider $\tilde{\mu} \sim \mathrm{SP}(\nu, h)$ as in Definition 2.2.2. The measure $\tilde{\mu}^F_{\Delta_{1,h}}$ such that

$$\begin{aligned} \tilde{\mu}^F_{\Delta_{1,h}} \mid \Delta_{1,h} &\sim \mathrm{CRM}(\nu'_F), \quad \text{where} \\ \nu'_F(\mathrm{d}s, \mathrm{d}x) &= \rho'_F(s)\mathrm{d}t\alpha(\mathrm{d}x), \\ \rho'_F(s) &= \Delta_{1,h}\rho(\Delta_{1,h}(1-e^{-s}))e^{-s}\mathbb{1}_{(0,\infty)}(s). \end{aligned} \tag{2.12}$$

is called *Beta-Stacy Scaled Process (BS-SP)*. Let us write

$$\tilde{\mu}^F_{\Delta_{1,h}} \sim \text{BS-SP}(\nu, h).$$

If $\tilde{\mu}_{\Delta_{1,h_{c,\beta}}} \sim \text{SB-SP}(\nu_\sigma, h_{c,\beta})$, the measure $\tilde{\mu}^F_{\Delta_{1,h_{c,\beta}}}$ is called *Stable-Beta-Stacy Scaled Process (SBS-SP)*. Let us write

$$\tilde{\mu}^F_{\Delta_{1,h_{c,\beta}}} \sim \text{SBS-SP}(\nu_\sigma, h_{c,\beta}).$$

Note that this definition induces a nonparametric model on the survival function the same way the Beta-Stacy process is induced by the beta process; see Walker and Muliere (1997). In particular, the corresponding nonparametric model on the survival function $\tilde{S} = 1 - \tilde{F}$ is

$$\begin{aligned} (X_1, \dots, X_N) \mid \tilde{F} \sim & \tilde{F} \\ \tilde{F}(t) = & 1 - e^{-\tilde{\mu}^F_{\Delta_{1,h}}(0,t]} \quad \text{for any } t \geq 0 \\ \tilde{\mu}^F_{\Delta_{1,h}} \sim & \text{BS-SP}(\nu, h). \end{aligned} \tag{2.13}$$

Note that, given a survival dataset $\boldsymbol{D}$,

$$\mathbb{E}_{\tilde{\mu}_{\Delta_{1,h}} \mid \Delta_{1,h}} \left[\mathbb{1}_{\boldsymbol{D}}\right] = \mathbb{E}_{\tilde{\mu}^F_{\Delta_{1,h}} \mid \Delta_{1,h}} \left[\mathbb{1}_{\boldsymbol{D}}\right],$$

i.e., the models (2.4) and (2.13) induce the same likelihood for the data $\boldsymbol{D}$.

2.B Proofs of Section 2.3

Proof of Proposition 2.3.1. Thanks to the conditional representation reported in Lemma 2.A.1, the expected value can be written as

$$\begin{aligned} \mathbb{E}\left[\tilde{H}_t\right] &= \mathbb{E}\left[\tilde{\mu}_{\Delta_{1,h}}(0,t]\right] = \mathbb{E}_{\Delta_{1,h}}\left[\mathbb{E}_{\tilde{\mu}_{\Delta_{1,h}} \mid \Delta_{1,h}}\left[\tilde{\mu}_{\Delta_{1,h}}(0,t] \mid \Delta_{1,h}\right]\right] \\ &= \mathbb{E}_{\Delta_{1,h}}\left[\int_0^t \int_{\mathbb{R}^+} s\nu'(\mathrm{d}s, \mathrm{d}x)\right] = \mathbb{E}_{\Delta_{1,h}}\left[\int_0^t \int_{\mathbb{R}^+} s\Delta_{1,h}\rho(\Delta_{1,h}s)\mathbb{1}_{(0,1)}(s)\mathrm{d}s\alpha(\mathrm{d}x)\right]. \end{aligned}$$

So the thesis follows. □

Proof of Theorem 2.3.1. Note that the likelihood is

$$\mathbb{E}\left[\mathbb{1}_{\boldsymbol{D}}\right] = \mathbb{E}_{\Delta_{1,h}}\left[\mathbb{E}_{\tilde{\mu}_{\Delta_{1,h}} \mid \Delta_{1,h}}\left[\mathbb{1}_{\boldsymbol{D}}\right]\right]$$

and recall that

$$\mathbb{E}_{\tilde{\mu}_{\Delta_{1,h}} \mid \Delta_{1,h}} \left[\mathbb{1}_{\boldsymbol{D}}\right] = \mathbb{E}_{\tilde{\mu}^F_{\Delta_{1,h}} \mid \Delta_{1,h}} \left[\mathbb{1}_{\boldsymbol{D}}\right],$$

is the likelihood of the data $\boldsymbol{D}$ arising from the models in (2.4) and in (2.13), as underlined in the Section 2.A. For ease of calculation, we are going to compute

$$\mathbb{E}_{\tilde{\mu}^F_{\Delta_{1,h}} \mid \Delta_{1,h}} \left[\mathbb{1}_{\boldsymbol{D}}\right].$$

So according to Definition 2.A.1 we have that

$$\tilde{\mu}^F_{\Delta_{1,h}} \mid \Delta_{1,h} \sim \mathrm{CRM}(\nu'_F),$$

where

$$\nu'_F(\mathrm{d}s, \mathrm{d}x) = \rho'_F(s)\mathrm{d}s\alpha(\mathrm{d}x) \quad \text{and}$$
$$\rho'_F(s) = \Delta_{1,h}\rho(\Delta_{1,h}(1-e^{-s}))e^{-s}\mathbb{1}_{(0,\infty)}(s).$$

Exploiting the general result for the expression of likelihood under CRM priors reported in Ferguson and Phadia (1979), it follows that

$$\mathbb{E}_{\tilde{\mu}_{\Delta_{1,h}}|\Delta_{1,h}}[\mathbb{1}_{\boldsymbol{D}}] = \prod_{r=1\ \Theta^*_r=1}^{K} \left[\alpha(\mathrm{d}X^*_r)\int_{\mathbb{R}^+}(1-e^{-s})^{n_r}e^{-s(\bar{n}_{r+1}+\tilde{n}^c_r)}\rho'(s)\mathrm{d}s\right]$$
$$\times\, exp\left\{-\int_{\mathbb{R}^+}\int_{\mathbb{X}}\left(1-e^{-Y(x)s}\right)\nu'(\mathrm{d}s,\mathrm{d}x)\right\}.$$

Since

$$\mathbb{E}_{\Delta_{1,h}}\left[\mathbb{E}_{\tilde{\mu}_{\Delta_{1,h}}|\Delta_{1,h}}[\mathbb{1}_{\boldsymbol{D}}]\right] = \int_{\mathbb{R}^+}\mathbb{E}_{\tilde{\mu}_{\Delta_{1,h}}|\Delta_{1,h}=y}[\mathbb{1}_{\boldsymbol{D}}]\, f_{\Delta_{1,h}}(y)\mathrm{d}y,$$

it follows that the likelihood is infinitesimally equal to

$$\int_{\mathbb{R}^+} y^{K^*}\prod_{r=1\ \Theta^*_r=1}^{K}\left[\alpha(\mathrm{d}X^*_r)\int_{\mathbb{R}^+}(1-e^{-s})^{n_r}e^{-s(\bar{n}_{r+1}+\tilde{n}^c_r+1)}\rho(y(1-e^{-s}))\mathrm{d}s\right]$$
$$\times\, exp\left\{-y\int_{\mathbb{R}^+}\int_{\mathbb{R}^+}\left(1-e^{-Y(x)s}\right)\rho(y(1-e^{-s}))e^{-s}\mathrm{d}s\alpha(\mathrm{d}x)\right\} f_{\Delta_{1,h}}(y)\mathrm{d}y.$$

The thesis follows with the change of variable $1-e^{-s}\mapsto s$. □

Proof of Proposition 2.3.2. Let us recall that

$$\tilde{\mu}^F_{\Delta_{1,h}} \mid \Delta_{1,h} \sim \mathrm{CRM}(\nu'_F),$$

where $\nu'_F(\mathrm{d}s,\mathrm{d}x)$ is described in (2.12). Exploiting the general results for the CRMs reported in Ferguson and Phadia (1979), it follows that

$$\tilde{\mu}^F_{\Delta_{1,h}} \mid \boldsymbol{D} = \tilde{\mu}^{*,F}_c + \sum_{r:\Theta^*_r=1}^{K} J^F_r \delta_{X^*_r},$$

where:

i. $\tilde{\mu}^{*,F}_c \sim \mathrm{CRM}(\nu^{*,F}_c)$, where

$$\nu^{*,F}_c(\mathrm{d}s, \mathrm{d}x) = e^{-Y(x)s} y e^{-s} \rho\left(y(1-e^{-s})\right) \mathbb{1}_{(0,\infty)}(s) \mathrm{d}s \alpha(\mathrm{d}x),$$

ii. the jumps J^F_r's have density proportional to

$$(1-e^{-s})^{n_r} e^{-s(\bar{n}_{r+1}+\tilde{n}^c_r)} y e^{-s} \rho\left(y(1-e^{-s})\right) \mathbb{1}_{(0,\infty)}(s) \mathrm{d}s.$$

The thesis follows with the change of variable $1-e^{-s} \mapsto s$. □

Proof of Proposition 2.3.2. From Theorem 2.3.2,

$$\begin{aligned}
\mathbb{E}\left[\tilde{H}_t \mid \boldsymbol{D}\right] =& \mathbb{E}\left[\tilde{\mu}_{\Delta_{1,h}}(0,t] \mid \boldsymbol{D}\right] = \mathbb{E}_{\Delta_{1,h}|\boldsymbol{D}}\left[\mathbb{E}_{\tilde{\mu}_{\Delta_{1,h}}|\boldsymbol{D},\Delta_{1,h}}\left[\tilde{\mu}_{\Delta_{1,h}}(0,t] \mid \boldsymbol{D}, \Delta_{1,h}\right]\right] \\
=& \mathbb{E}_{\Delta_{1,h}|\boldsymbol{D}}\left[\mathbb{E}_{\tilde{\mu}_{\Delta_{1,h}}|\boldsymbol{D},\Delta_{1,h}}\left[\tilde{\mu}^*_c(0,t] + \sum_{r:\Theta^*_r=1}^{K} J_r \delta_{X^*_r}(0,t]\right]\right] \\
=& \int_{\mathbb{R}^+}\left[\mathbb{E}_{\tilde{\mu}_{\Delta_{1,h}}|\boldsymbol{D},\Delta_{1,h}=y}\left[\tilde{\mu}^*_c(0,t]\right] + \sum_{r:\Theta^*_r=1, X^*_r\le t}^{K} \mathbb{E}_{\tilde{\mu}_{\Delta_{1,h}}|\boldsymbol{D},\Delta_{1,h}=y}[J_r]\right] f_{\Delta_{1,h}|\boldsymbol{D}}(y).
\end{aligned}$$

Note that from Theorem 2.3.2 follows that

$$\mathbb{E}_{\tilde{\mu}_{\Delta_{1,h}}|\boldsymbol{D},\Delta_{1,h}=y}\left[\tilde{\mu}^*_c(0,t]\right] = \int_0^t \int_{\mathbb{R}^+} s\nu^*(\mathrm{d}s, \mathrm{d}x),$$

and that the density function for the jump J_r is

$$\frac{s^{n_r}(1-s)^{\bar{n}_{r+1}+\tilde{n}^c_r}\rho\left(ys\right)\mathrm{d}s}{\int_0^1 t^{n_r}(1-t)^{\bar{n}_{r+1}+\tilde{n}^c_r}\rho\left(yt\right)\mathrm{d}t}$$

so the thesis follows. □

2.C Proofs of Section 2.4

Proof of Proposition 2.4.1. Let us show it via the Laplace functional. Let us consider a measurable function

$$g : \mathbb{R}^+ \to \mathbb{R} \quad \text{such that} \quad \int \mid g \mid \mathrm{d}\tilde{\mu}^F_{\Delta_{1,h_{c,\beta}}} < \infty.$$

Then,

$$\begin{aligned}\mathbb{E}\left[e^{-\tilde{\mu}_{\Delta_{1,h_{c,\beta}}}(g)} \mid \Delta_{1,h_{c,\beta}} = y\right] &= \exp\left\{-\int_{\mathbb{R}^+}\int_0^1 \left(1 - e^{-sg(x)}\right) y\sigma y^{-1-\sigma} s^{-1-\sigma} \mathrm{d}s\alpha(\mathrm{d}x)\right\} \\ &= \exp\left\{-\sigma y^{-\sigma}\int_{\mathbb{R}^+}\int_0^1 \left(1 - e^{-sg(x)}\right) s^{-1-\sigma} \mathrm{d}s\alpha(\mathrm{d}x)\right\}.\end{aligned}$$

Let us define

$$I := \int_{\mathbb{R}^+}\int_0^1 \left(1 - e^{-sg(x)}\right) s^{-1-\sigma} \mathrm{d}s\alpha(\mathrm{d}x).$$

Therefore, integrating the jump $\Delta_{1,h_{c,\beta}}$ out we have that

$$\begin{aligned}\mathbb{E}\left[e^{-\tilde{\mu}^F_{\Delta_{1,h_{c,\beta}}}(g)}\right] &= \int_{\mathbb{R}^+} e^{-\sigma y^{-\sigma} I}\frac{\sigma\beta^{c+1}}{\Gamma(c+1)} y^{-1-(c+1)\sigma} e^{-\beta y^{-\sigma}} \mathrm{d}y \\ &= \frac{\sigma\beta^{c+1}}{\Gamma(c+1)}\int_{\mathbb{R}^+} y^{-1-(c+1)\sigma} e^{-(\sigma I + \beta) y^{-\sigma}} \mathrm{d}y \\ &= \frac{\sigma\beta^{c+1}}{\Gamma(c+1)} \times \frac{\Gamma(c+1)}{\sigma(\sigma I + \beta)^{c+1}} = \left(\frac{\beta}{\sigma I + \beta}\right)^{c+1} \\ &= \left(1 + \frac{\sigma}{\beta} I\right)^{-(c+1)} = \left(1 + \int_{\mathbb{R}^+}\int_0^1 \left(1 - e^{-sg(x)}\right)\frac{\sigma}{\beta} s^{-1-\sigma}\mathrm{d}s\alpha(\mathrm{d}x)\right)^{-(c+1)}.\end{aligned}$$

So the thesis follows. □

Proof of Corollary 2.4.1. Note that the results of Proposition 2.3.1 can be specialized considering that

$$\rho(ys) = \sigma y^{-1-\sigma} s^{-1-\sigma}$$

and that, from Lemma 2.A.2,

$$f_{\Delta_{1,h_{c,\beta}}}(y) = \frac{\sigma\beta^{c+1}}{\Gamma(c+1)} y^{-1-(c+1)\sigma} e^{-\beta y^{-\sigma}}.$$

So we have that the expected value is

$$
\begin{aligned}
\mathbb{E}\left[\tilde{H}_t\right] &= \alpha((0,t]) \cdot \int_{\mathbb{R}^+} \int_0^1 sy\rho(ys)\mathrm{d}s f_{\Delta_{1,h_{c,\beta}}}(y)\mathrm{d}y \\
&= \alpha((0,t]) \cdot \int_{\mathbb{R}^+} \int_0^1 sy\sigma y^{-1-\sigma} s^{-1-\sigma}\mathrm{d}s \frac{\sigma\beta^{c+1}}{\Gamma(c+1)} y^{-1-(c+1)\sigma} e^{-\beta y^{-\sigma}}\mathrm{d}y \\
&= \frac{\sigma^2\beta^{c+1}\alpha((0,t])}{\Gamma(c+1)} \int_{\mathbb{R}^+} y^{-1-(c+2)\sigma} e^{-\beta y^{-\sigma}}\mathrm{d}y \int_0^1 s^{-\sigma}\mathrm{d}s \\
&= \frac{\sigma^2\beta^{c+1}\alpha((0,t])}{(1-\sigma)\Gamma(c+1)} \int_{\mathbb{R}^+} y^{-1-(c+2)\sigma} e^{-\beta y^{-\sigma}}\mathrm{d}y.
\end{aligned}
$$

Note that the function inside the integral is proportional to the density of variable X such that $X^{-\sigma}$ is distributed as a gamma with shape equal to $c+2$ and rate equal to β. Then

$$
\begin{aligned}
\mathbb{E}\left[\tilde{H}_t\right] &= \frac{\sigma^2\beta^{c+1}\alpha((0,t])}{(1-\sigma)\Gamma(c+1)} \int_{\mathbb{R}^+} y^{-1-(c+2)\sigma} e^{-\beta y^{-\sigma}}\mathrm{d}y \\
&= \frac{\sigma^2\beta^{c+1}}{(1-\sigma)\Gamma(c+1)} \times \frac{\Gamma(c+2)}{\sigma\beta^{c+2}} \times \alpha((0,t]).
\end{aligned}
$$

So the thesis follows. □

Proof of Corollary 2.4.2. In order to specialize the expression in (2.5), let us recall that if $\tilde{\mu}_{\Delta_{1,h_{c,\beta}}} \sim$ SB-SP$(\nu_\sigma, h_{c,\beta})$, then

$$
\begin{aligned}
\rho(ys) &= \sigma y^{-1-\sigma} s^{-1-\sigma}, \\
f_{\Delta_{1,h_{c,\beta}}}(y) &= \frac{\sigma\beta^{c+1}}{\Gamma(c+1)} y^{-1-(c+1)\sigma} e^{-\beta y^{-\sigma}},
\end{aligned}
$$

i.e., thanks to Lemma 2.A.2,

$$
\Delta_{1,h_{c,\beta}}^{-\sigma} \sim \mathrm{Gamma}(c+1, \beta).
$$

So, exploiting (2.5) we obtain that the likelihood is infinitesimally equal to

$$
\begin{aligned}
&\int_{\mathbb{R}^+} \sigma^{K^*} y^{-K^*\sigma} \prod_{r=1:\Theta r^*=1}^{K} \left[\alpha(\mathrm{d}X_r^*) \int_0^1 s^{n_r-1-\sigma}(1-s)^{\bar{n}_{r+1}+\tilde{n}_r^c} \mathrm{d}s \right] \\
&\times \exp\left\{ -\sigma y^{-\sigma} \int_{\mathbb{R}^+} \left[\int_0^1 \left(1-(1-s)^{Y(x)}\right) s^{-1-\sigma} \mathrm{d}s \right] \alpha(\mathrm{d}x) \right\} \\
&\times \frac{\sigma \beta^{c+1}}{\Gamma(c+1)} y^{-1-(c+1)\sigma} e^{-\beta y^{-\sigma}} \mathrm{d}y \\
= &\int_{\mathbb{R}^+} \sigma^{K^*+1} y^{-1-(K^*+c+1)\sigma} e^{-\beta y^{-\sigma}} \frac{\beta^{c+1}}{\Gamma(c+1)} \times \prod_{r=1:\Theta r^*=1}^{K} \mathrm{B}\left(n_r-\sigma, \bar{n}_{r+1}+\tilde{n}_r^c+1\right) \\
&\times \exp\left\{ -\sigma y^{-\sigma} \int_{\mathbb{R}^+} \left[\int_0^1 \left(1-(1-s)^{Y(x)}\right) s^{-1-\sigma} \mathrm{d}s \right] \alpha(\mathrm{d}x) \right\} \mathrm{d}y.
\end{aligned}
$$

Note that the at-risk process $Y(x)$ is a piecewise constant function that can be written as

$$
Y(x) = \sum_{i=1}^{N} \mathbb{1}_{[x,\infty)}(X_i) = \sum_{r=1}^{K} \sum_{g=r}^{K} n_g^* \mathbb{1}_{(X_{r-1}^*, X_r^*]}(x),
$$

where n_g^* is the number of observations equal to X_r^* and we assume $X_0^* = 0$. Let us define

$$
N_r^* := \sum_{g=r}^{K} n_g^*,
$$

and note that $Y(x) = 0$ for any $x > X_K^*$. So we can write the exponential in the expression of the likelihood as

$$
\begin{aligned}
&\exp\left\{ -\sigma y^{-\sigma} \int_{\mathbb{R}^+} \left[\int_0^1 \left(1-(1-s)^{Y(x)}\right) s^{-1-\sigma} \mathrm{d}s \right] \alpha(\mathrm{d}x) \right\} \\
= &\exp\left\{ -\sigma y^{-\sigma} \sum_{r=1}^{K} \int_{X_{r-1}^*}^{X_r^*} \left[\int_0^1 \left(1-(1-s)^{N_r^*}\right) s^{-1-\sigma} \mathrm{d}s \right] \alpha(\mathrm{d}x) \right\} \\
= &\exp\left\{ -\sigma y^{-\sigma} \sum_{r=1}^{K} \alpha((X_{r-1}^*, X_r^*]) \left[\int_0^1 \left(1-(1-s)^{N_r^*}\right) s^{-1-\sigma} \mathrm{d}s \right] \right\}.
\end{aligned}
$$

Note that if $N_r^* = 0$,

$$
\int_0^1 \left(1-(1-s)^{N_r^*}\right) s^{-1-\sigma} \mathrm{d}s = 0.
$$

If $N_r^* \geq 1$ the integral becomes

$$\begin{aligned}\int_0^1 \left(1-(1-s)^{N_r^*}\right) s^{-1-\sigma}\mathrm{d}s &= \int_0^1 \left[1-\sum_{k=0}^{N_r^*}(-1)^k\binom{N_r^*}{k}s^k\right] s^{-1-\sigma}\mathrm{d}s \\ &= \sum_{k=1}^{N_r^*}(-1)^{k-1}\binom{N_r^*}{k}\int_0^1 s^{k-1-\sigma}\mathrm{d}s \\ &= \sum_{k=1}^{N_r^*}(-1)^{k-1}\binom{N_r^*}{k}\frac{1}{k-\sigma},\end{aligned}$$

since $0<\sigma<1\leq k$. Therefore, we get

$$\begin{aligned}&\exp\left\{-\sigma y^{-\sigma}\int_{\mathbb{R}^+}\left[\int_0^1 \left(1-(1-s)^{Y(x)}\right) s^{-1-\sigma}\mathrm{d}s\right]\alpha(\mathrm{d}x)\right\} \\ =&\exp\left\{-\sigma y^{-\sigma}\sum_{r=1}^{K}\alpha((X_{r-1}^*,X_r^*])\sum_{k=1}^{N_r^*}(-1)^{k-1}\binom{N_r^*}{k}\frac{1}{k-\sigma}\right\}.\end{aligned}$$

Let us define

$$\eta_\sigma := \sum_{r=1}^{K}\alpha((X_{r-1}^*,X_r^*])\sum_{k=1}^{N_r^*}(-1)^{k-1}\binom{N_r^*}{k}\frac{1}{k-\sigma},$$

hence, we finally obtain that the likelihood is infinitesimally equal to

$$\begin{aligned}&\prod_{r=1\ \Theta r^*=1}^{K}\alpha(\mathrm{d}X_r^*)\mathrm{B}\left(\bar{n}_{r+1}+\tilde{n}_r^c+1,n_r-\sigma\right) \\ &\times\frac{\sigma^{K^*+1}\beta^{c+1}}{\Gamma(c+1)}\int_{\mathbb{R}^+} y^{-1-(K^*+c+1)\sigma}e^{-(\beta+\sigma\eta_\sigma)y^{-\sigma}}\mathrm{d}y.\end{aligned}$$

From the previous expression we note that the posterior density of the jump $\Delta_{1,h_{c,\beta}}$ is proportional to

$$y^{-(K^*+c+1)\sigma-1}e^{-(\beta+\sigma\eta_\sigma)y^{-\sigma}},$$

i.e.,

$$\Delta_{1,h_{c,\beta}}^{-\sigma}\mid \boldsymbol{D}\sim\mathrm{Gamma}\left(K^*+c+1,\beta+\sigma\eta_\sigma\right)$$

and

$$f_{\Delta_{1,h_{c,\beta}}|\boldsymbol{D}}(y)=\frac{\sigma\left(\beta+\sigma\eta_\sigma\right)^{K^*+c+1}}{\Gamma(K^*+c+1)}y^{-(K^*+c+1)\sigma-1}e^{-(\beta+\sigma\eta_\sigma)y^{-\sigma}}.$$

The thesis now follows. □

Proof of Corollary 2.4.3. Let us first consider the model in (2.13) with a SBS-SP prior, i.e., with

$$\tilde{\mu}^F_{\Delta_{1,h_{c,\beta}}} \sim \text{SBS-SP}(\nu_\sigma, h_{c,\beta}).$$

Let us prove that its posterior distribution can be written as:

$$\tilde{\mu}^F_{\Delta_{1,h_{c,\beta}}} \mid \boldsymbol{D} = \tilde{\mu}^{*,F}_c + \sum_{r:\Theta r^*=1}^{K} J^F_r \delta_{X^*_r}, \tag{2.14}$$

where:

i. $\tilde{\mu}^{*,F}_c$ is a negative binomial random measure, i.e., $\tilde{\mu}^{*,F}_c \sim \text{BN}(K^* + c + 1, \rho^{*,F}, \alpha)$, where
$$\rho^{*,F}(s \mid x)\mathrm{d}s = \frac{\sigma}{\beta + \sigma\eta_\sigma} e^{-s(Y(x)+1)}(1 - e^{-s})^{-1-\sigma}\mathrm{d}s,$$

ii. the jumps J^F_r's have density proportional to
$$(1 - e^{-s})^{n_r - 1 - \sigma} e^{-s(\bar{n}_{r+1} + \tilde{n}^c_r + 1)}\mathrm{d}s.$$

Then the thesis follows from (2.14) with the reparametrization $s \mapsto 1 - e^{-s}$. Note that after the change of variable the jumps J_r's have density proportional to

$$s^{n_r - 1 - \sigma}(1 - s)^{\bar{n}_{r+1} + \tilde{n}^c_r}\mathrm{d}s.$$

Let us now prove the posterior expression in (2.14). The posterior distribution of $\tilde{\mu}^F_{\Delta_{1,h_{c,\beta}}}$ can be obtained via its posterior Laplace functional. In particular, note that for any measurable function

$$g : \mathbb{R}^+ \to \mathbb{R} \quad \text{such that} \quad \int \mid g \mid \mathrm{d}\left(\tilde{\mu}^F_{\Delta_{1,h_{c,\beta}}} \mid \boldsymbol{D}\right) < \infty,$$

$$\begin{aligned}
L_{\tilde{\mu}^F_{\Delta_{1,h_{c,\beta}}} \mid \boldsymbol{D}}(g) &= \mathbb{E}\left[e^{-\int_{\mathbb{R}^+} g(x)\left(\tilde{\mu}^F_{\Delta_{1,h_{c,\beta}}} \mid \boldsymbol{D}\right)(\mathrm{d}x)}\right] \\
&= \mathbb{E}_{\Delta_{1,h_{c,\beta}} \mid \boldsymbol{D}}\left[\mathbb{E}_{\tilde{\mu}^F_{\Delta_{1,h_{c,\beta}}} \mid \Delta_{1,h_{c,\beta}}, \boldsymbol{D}}\left[e^{-\int_{\mathbb{R}^+} g(x)\left(\tilde{\mu}^F_{\Delta_{1,h_{c,\beta}}} \mid \boldsymbol{D}\right)(\mathrm{d}x)}\right]\right] \\
&= \int_{\mathbb{R}^+} L_{\tilde{\mu}^F_{\Delta_{1,h_{c,\beta}}} \mid \left(\Delta_{1,h_{c,\beta}} = y\right), \boldsymbol{D}}(g) f_{\Delta_{1,h_{c,\beta}} \mid \boldsymbol{D}}(y)\mathrm{d}y,
\end{aligned}$$

where $f_{\Delta_{1,h_{c,\beta}}|\boldsymbol{D}}$ is the posterior density of the jump $\Delta_{1,h_{c,\beta}}$ (see Corollary 2.4.2). Since $\tilde{\mu}^F_{\Delta_{1,h_{c,\beta}}} \mid \Delta_{1,h_{c,\beta}} \sim \mathrm{CRM}(\nu'_F)$, from Ferguson and Phadia (1979) it is well known that

$$
\begin{aligned}
&L_{\tilde{\mu}^F_{\Delta_{1,h_{c,\beta}}}|\left(\Delta_{1,h_{c,\beta}}=y\right),\boldsymbol{D}}(g) = \\
&\prod_{r=1:\Theta^*_r=1}^{K} \frac{\int_{\mathbb{R}^+} e^{-sg(X^*_r)}(1-e^{-s})^{n_r} e^{-s(\bar{n}_{r+1}+\tilde{n}^c_r+1)} y\rho(y(1-e^{-s}))\mathrm{d}s}{\int_{\mathbb{R}^+}(1-e^{-s})^{n_r} e^{-s(\bar{n}_{r+1}+\tilde{n}^c_r+1)} y\rho(y(1-e^{-s}))\mathrm{d}s} \\
&\times \exp\left\{-\int_{\mathbb{R}^+}\int_{\mathbb{R}^+}(1-e^{-sg(x)})e^{-s(Y(x)+1)} y\rho(y(1-e^{-s}))\mathrm{d}s\alpha(\mathrm{d}x)\right\} \qquad (2.15)\\
=&\prod_{r=1\ \Theta^*_r=1}^{K} \frac{\int_{\mathbb{R}^+} e^{-sg(X^*_r)}(1-e^{-s})^{n_r} e^{-s(\bar{n}_{r+1}+\tilde{n}^c_r+1)} \rho(y(1-e^{-s}))\mathrm{d}s}{\int_{\mathbb{R}^+}(1-e^{-s})^{n_r} e^{-s(\bar{n}_{r+1}+\tilde{n}^c_r+1)} \rho(y(1-e^{-s}))\mathrm{d}s} \\
&\times \exp\left\{-y\int_{\mathbb{R}^+}\int_{\mathbb{R}^+}(1-e^{-sg(x)})e^{-s(Y(x)+1)} \rho(y(1-e^{-s}))\mathrm{d}s\alpha(\mathrm{d}x)\right\}.
\end{aligned}
$$

Note that (2.15) provides the posterior Laplace functional under a SP prior. When the prior is a SB-SP, we can continue the equalities as follows.

$$
\begin{aligned}
&L_{\tilde{\mu}^F_{\Delta_{1,h_{c,\beta}}}|\left(\Delta_{1,h_{c,\beta}}=y\right),\boldsymbol{D}}(g) = \\
&\prod_{r=1\ \Theta^*_r=1}^{K} \frac{\int_{\mathbb{R}^+} e^{-sg(X^*_r)}(1-e^{-s})^{n_r} e^{-s(\bar{n}_{r+1}+\tilde{n}^c_r+1)} \sigma y^{-1-\sigma}(1-e^{-s})^{-1-\sigma}\mathrm{d}s}{\int_{\mathbb{R}^+}(1-e^{-s})^{n_r} e^{-s(\bar{n}_{r+1}+\tilde{n}^c_r+1)} \sigma y^{-1-\sigma}(1-e^{-s})^{-1-\sigma}\mathrm{d}s} \\
&\times \exp\left\{-y\int_{\mathbb{R}^+}\int_{\mathbb{R}^+}(1-e^{-sg(x)})e^{-s(Y(x)+1)} \sigma y^{-1-\sigma}(1-e^{-s})^{-1-\sigma}\mathrm{d}s\alpha(\mathrm{d}x)\right\} \\
=&\prod_{r=1\ \Theta^*_r=1}^{K} \frac{\int_{\mathbb{R}^+} e^{-sg(X^*_r)}(1-e^{-s})^{n_r-1-\sigma} e^{-s(\bar{n}_{r+1}+\tilde{n}^c_r+1)}\mathrm{d}s}{\int_{\mathbb{R}^+}(1-e^{-s})^{n_r-1-\sigma} e^{-s(\bar{n}_{r+1}+\tilde{n}^c_r+1)}\mathrm{d}s} \\
&\times \exp\left\{-\sigma y^{-\sigma}\int_{\mathbb{R}^+}\int_{\mathbb{R}^+}(1-e^{-sg(x)})e^{-s(Y(x)+1)}(1-e^{-s})^{-1-\sigma}\mathrm{d}s\alpha(\mathrm{d}x)\right\}.
\end{aligned}
$$

Since

$$\Delta^{-\sigma}_{1,h_{c,\beta}} \mid \boldsymbol{D} \sim \mathrm{Gamma}\left(K^* + c + 1, \beta + \sigma\eta_\sigma\right),$$

we obtain that

$$L_{\tilde{\mu}^F_{\Delta_{1,h_{c,\beta}}}|\boldsymbol{D}}(g) = \int_{\mathbb{R}^+} \prod_{r=1 \ \Theta_r^*=1}^{K} \frac{\int_{\mathbb{R}^+} e^{-sg(X_r^*)}(1-e^{-s})^{n_r-1-\sigma} e^{-s(\bar{n}_{r+1}+\tilde{n}_r^c+1)}\mathrm{d}s}{\int_{\mathbb{R}^+}(1-e^{-s})^{n_r-1-\sigma} e^{-s(\bar{n}_{r+1}+\tilde{n}_r^c+1)}\mathrm{d}s} \times$$
$$\times \exp\left\{-\sigma y^{-\sigma}\int_{\mathbb{R}^+}\int_{\mathbb{R}^+}(1-e^{-sg(x)})e^{-s(Y(x)+1)}(1-e^{-s})^{-1-\sigma}\mathrm{d}s\alpha(\mathrm{d}x)\right\}\times$$
$$\times \frac{\sigma\left(\beta+\sigma\eta_\sigma\right)^{K^*+c+1}}{\Gamma(K^*+c+1)} y^{-(K^*+c+1)\sigma-1} e^{-(\beta+\sigma\eta_\sigma)y^{-\sigma}}\mathrm{d}y =$$
$$= \prod_{r=1 \ \Theta_r^*=1}^{K} \frac{\int_{\mathbb{R}^+} e^{-sg(X_r^*)}(1-e^{-s})^{n_r-1-\sigma} e^{-s(\bar{n}_{r+1}+\tilde{n}_r^c+1)}\mathrm{d}s}{\int_{\mathbb{R}^+}(1-e^{-s})^{n_r-1-\sigma} e^{-s(\bar{n}_{r+1}+\tilde{n}_r^c+1)}\mathrm{d}s} \times$$
$$\times \frac{\sigma\left(\beta+\sigma\eta_\sigma\right)^{K^*+c+1}}{\Gamma(K^*+c+1)} \int_{\mathbb{R}^+} e^{-(\beta+\sigma\eta+\sigma\eta_\sigma)y^{-\sigma}} y^{-\sigma(K^*+c+1)}\mathrm{d}y,$$

where

$$\eta = \int_{\mathbb{R}^+}\int_{\mathbb{R}^+}(1-e^{-sg(x)})e^{-s(Y(x)+1)}(1-e^{-s})^{-1-\sigma}\mathrm{d}s\alpha(\mathrm{d}x).$$

Note that

$$\frac{\sigma\left(\beta+\sigma\eta_\sigma\right)^{K^*+c+1}}{\Gamma(K^*+c+1)} \int_{\mathbb{R}^+} e^{-(\beta+\sigma\eta+\sigma\eta_\sigma)y^{-\sigma}} y^{-\sigma(K^*+c+1)}\mathrm{d}y =$$
$$= \frac{\sigma\left(\beta+\sigma\eta_\sigma\right)^{K^*+c+1}}{\Gamma(K^*+c+1)} \times \frac{1}{\sigma} \times \frac{\Gamma(K^*+c+1)}{(\beta+\sigma\eta+\sigma\eta_\sigma)^{K^*+c+1}} = \left(1+\frac{\sigma}{\beta+\sigma\eta_\sigma}\eta\right)^{-(K^*+c+1)}.$$

So finally

$$L_{\tilde{\mu}^F_{\Delta_{1,h_{c,\beta}}}|\boldsymbol{D}}(g) = \prod_{r=1 \ \Theta_r^*=1}^{K} \frac{\int_{\mathbb{R}^+} e^{-sg(X_r^*)}(1-e^{-s})^{n_r-1-\sigma} e^{-s(\bar{n}_{r+1}+\tilde{n}_r^c+1)}\mathrm{d}s}{\int_{\mathbb{R}^+}(1-e^{-s})^{n_r-1-\sigma} e^{-s(\bar{n}_{r+1}+\tilde{n}_r^c+1)}\mathrm{d}s} \times$$
$$\times \left(1+\int_{\mathbb{R}^+}\int_{\mathbb{R}^+}(1-e^{-sg(x)})\frac{\sigma}{\beta+\sigma\eta_\sigma}e^{-s(Y(x)+1)}(1-e^{-s})^{-1-\sigma}\mathrm{d}s\alpha(\mathrm{d}x)\right)^{-(K^*+c+1)},$$

hence the thesis follows. □

Proof of Corollary 2.4.4. To prove the statement it is enough to specialize the expression in Proposition 2.3.2 by noticing that

$$\rho(ys) = \sigma y^{-1-\sigma} s^{-1-\sigma}$$

and

$$f_{\Delta_{1,h_{c,\beta}}|\boldsymbol{D}}(y) = \frac{\sigma\left(\beta+\sigma\eta_\sigma\right)^{K^*+c+1}}{\Gamma(K^*+c+1)} y^{-(K^*+c+1)\sigma-1} e^{-(\beta+\sigma\eta_\sigma)y^{-\sigma}},$$

thanks to Lemma 2.A.2. So it follows that

$$\begin{aligned}\int_0^t\int_0^1 s(1-s)^{Y(x)}\rho(ys)\mathrm{d}s\alpha(\mathrm{d}x) =& \sigma y^{-\sigma}\int_0^t\int_0^1 s^{-\sigma}(1-s)^{Y(x)}\mathrm{d}s\alpha(\mathrm{d}x) = \\ =& \sigma y^{-\sigma}\sum_{r=1}^{K_t}\alpha\left((X^*_{r-1},X^*_r]\right)\int_0^1 s^{-\sigma}(1-s)^{N^*_r}\mathrm{d}s = \\ =& \sigma y^{-\sigma}\sum_{r=1}^{K_t}\alpha\left((X^*_{r-1},X^*_r]\right)B(1-\sigma,N^*_r+1),\end{aligned}$$

where

$$K_t = \min\{g : X^*_g \geq t\}.$$

Note that

$$\begin{aligned}B(1-\sigma,N^*_r+1) =& \int_0^1 s^{-\sigma}(1-s)^{N^*_r}\mathrm{d}s = \\ =& \sum_{k=0}^{N^*_r}\binom{N^*_r}{k}(-1)^k\int_0^1 s^{k-\sigma}\mathrm{d}s = \\ =& \sum_{k=0}^{N^*_r}\binom{N^*_r}{k}(-1)^k\frac{1}{k+1-\sigma}.\end{aligned}$$

Let us define

$$\eta_t := \sum_{r=1}^{K_t}\alpha\left((X^*_{r-1},X^*_r]\right)B(1-\sigma,N^*_r+1).$$

So finally

$$\int_0^t\int_0^1 s(1-s)^{Y(x)}\rho(ys)\mathrm{d}s\alpha(\mathrm{d}x) = \sigma y^{-\sigma}\eta_t. \tag{2.16}$$

Note also that

$$\exp\left\{-\int_0^t\int_{\mathbb{R}^+}(1-e^{-s})e^{-s(Y(x)+1)}\rho(y(1-e^{-s}))\mathrm{d}s\alpha(\mathrm{d}x)\right\} = e^{-\sigma y^{-\sigma}\eta_t}.$$

Moreover,

$$
\begin{aligned}
&\frac{\int_0^1 s^{n_r+1}(1-s)^{\bar{n}_{r+1}+\tilde{n}_r^c}\rho\,(ys)\,\mathrm{d}s}{\int_0^1 t^{n_r}(1-t)^{\bar{n}_{r+1}+\tilde{n}_r^c}\rho\,(yt)\,\mathrm{d}t} = \frac{\sigma y^{-1-\sigma}\int_0^1 s^{n_r-\sigma}(1-s)^{\bar{n}_{r+1}+\tilde{n}_r^c}\mathrm{d}s}{\sigma y^{-1-\sigma}\int_0^1 t^{n_r-\sigma-1}(1-t)^{\bar{n}_{r+1}+\tilde{n}_r^c}\mathrm{d}t} \\
&= \int_0^1 \frac{s^{n_r-\sigma}(1-s)^{\bar{n}_{r+1}+\tilde{n}_r^c}\mathrm{d}s}{\mathrm{B}(n_r-\sigma, \bar{n}_{r+1}+\tilde{n}_r^c+1)}\mathrm{d}s = \frac{n_r-\sigma}{\bar{n}_{r+1}+\tilde{n}_r^c+n_r+1-\sigma},
\end{aligned}
\tag{2.17}
$$

since the last integral is the expected value of a $\mathrm{Beta}(n_r-\sigma, \bar{n}_{r+1}+\tilde{n}_r^c+1)$ random variable. Note also that

$$1-e^{-J_r^F}$$

is distributed again as a $\mathrm{Beta}(n_r-\sigma, \bar{n}_{r+1}+\tilde{n}_r^c+1)$ random variable. So, merging (2.16) and (2.17) into the expression of Proposition 2.3.2, it follows that

$$
\begin{aligned}
\mathbb{E}\left[\tilde{H}_t \mid \boldsymbol{D}\right] &= \int_{\mathbb{R}^+}\left[\int_0^t\int_0^1 s(1-s)^{Y(x)}y\rho(ys)\mathrm{d}s\alpha(\mathrm{d}x)+\right. \\
&\quad\left. + \sum_{r=1;\Theta_r^*=1,X_r^*\le t}^{K} \frac{\int_0^1 s^{n_r+1}(1-s)^{\bar{n}_{r+1}+\tilde{n}_r^c}\rho\,(ys)\,\mathrm{d}s}{\int_0^1 t^{n_r}(1-t)^{\bar{n}_{r+1}+\tilde{n}_r^c}\rho\,(yt)\,\mathrm{d}t}\right] f_{\Delta_{1,h_{c,\beta}}|\boldsymbol{D}}(y)\mathrm{d}y = \\
&= \int_{\mathbb{R}^+}\left[\sigma y^{-\sigma}\eta_t + \sum_{r=1;\Theta_r^*=1,X_r^*\le t}^{K}\frac{n_r-\sigma}{\bar{n}_{r+1}+\tilde{n}_r^c+n_r+1-\sigma}\right] f_{\Delta_{1,h_{c,\beta}}|\boldsymbol{D}}(y)\mathrm{d}y = \\
&= \sigma\eta_t\int_{\mathbb{R}^+} y^{-\sigma} f_{\Delta_{1,h_{c,\beta}}|\boldsymbol{D}}(y)\mathrm{d}y + \sum_{r=1;\Theta_r^*=1,X_r^*\le t}^{K}\frac{n_r-\sigma}{\bar{n}_{r+1}+\tilde{n}_r^c+n_r+1-\sigma}.
\end{aligned}
$$

Note that the integral in the previous expression can be explicitly calculated as follows:

$$
\begin{aligned}
\int_{\mathbb{R}^+} y^{-\sigma} f_{\Delta_{1,h_{c,\beta}}|\boldsymbol{D}}(y)\mathrm{d}y &= \frac{\sigma\,(\beta+\sigma\eta_\sigma)^{K^*+c+1}}{\Gamma(K^*+c+1)}\int_{\mathbb{R}^+} y^{-(K^*+c+2)\sigma-1}e^{-(\beta+\sigma\eta_\sigma)y^{-\sigma}}\mathrm{d}y = \\
&= \frac{\sigma\,(\beta+\sigma\eta_\sigma)^{K^*+c+1}}{\Gamma(K^*+c+1)} \times \frac{\Gamma(K^*+c+2)}{\sigma\,(\beta+\sigma\eta_\sigma)^{K^*+c+2}}.
\end{aligned}
$$

It follows that

$$\int_{\mathbb{R}^+} y^{-\sigma} f_{\Delta_{1,h_{c,\beta}}|\boldsymbol{D}}(y)\mathrm{d}y = \frac{K^*+c+1}{\beta+\sigma\eta_\sigma},$$

hence the thesis follows. □

Proof of Proposition 2.4.2. The random measure hierarchically defined in the statement is

$$\begin{aligned}\tilde{\mu} \mid \gamma \sim& \mathrm{CRM}(\nu),\\ \nu(\mathrm{d}s, \mathrm{d}x) =& \gamma \frac{\Gamma(2-\sigma)}{\Gamma(1-\sigma)\Gamma(1)} s^{-1-\sigma} \mathrm{d}s \alpha(\mathrm{d}x),\\ \gamma \sim& \mathrm{Gamma}\left(c+1, \frac{\beta(1-\sigma)}{\sigma}\right),\end{aligned}$$

i.e.,

$$\begin{aligned}\tilde{\mu} \mid \gamma \sim& \mathrm{CRM}(\nu),\\ \nu(\mathrm{d}s, \mathrm{d}x) =& \gamma(1-\sigma) s^{-1-\sigma} \mathrm{d}s \alpha(\mathrm{d}x),\\ \gamma \sim& \mathrm{Gamma}\left(c+1, \frac{\beta(1-\sigma)}{\sigma}\right).\end{aligned}$$

We are going to show that the Laplace functional of $\tilde{\mu}$ coincides with that of a negative binomial random measure. The Laplace functional of $\tilde{\mu}$, for any measurable function

$$f : \mathbb{R}^+ \to \mathbb{R} \quad \text{such that} \quad \int \mid f \mid \mathrm{d}\tilde{\mu} < \infty,$$

is

$$\begin{aligned}\mathbb{E}\left[e^{-\tilde{\mu}(f)}\right] =& \int_{\mathbb{R}+} \mathbb{E}\left[e^{-\tilde{\mu}(f)} \mid \gamma\right] \frac{1}{\Gamma(c+1)} \left(\frac{\beta(1-\sigma)}{\sigma}\right)^{c+1} \gamma^c \exp\left\{-\frac{\beta(1-\sigma)}{\sigma}\gamma\right\} \mathrm{d}\gamma =\\ =& \frac{1}{\Gamma(c+1)} \left(\frac{\beta(1-\sigma)}{\sigma}\right)^{c+1} \int_{\mathbb{R}^+} \exp\left\{-\gamma(1-\sigma)I\right\} \gamma^c \exp\left\{-\frac{\beta(1-\sigma)}{\sigma}\gamma\right\} \mathrm{d}\gamma,\end{aligned}$$

where

$$I := \int_{\mathbb{R}^+} \int_0^1 (1 - e^{-sf(x)}) s^{-1-\sigma} \mathrm{d}s \alpha(\mathrm{d}x).$$

So the Laplace functional of $\tilde{\mu}$ is

$$\begin{aligned}\mathbb{E}\left[e^{-\tilde{\mu}(f)}\right] =& \frac{1}{\Gamma(c+1)} \left(\frac{\beta(1-\sigma)}{\sigma}\right)^{c+1} \int_{\mathbb{R}^+} \gamma^c \exp\left\{-\gamma\left((1-\sigma)I + \frac{\beta(1-\sigma)}{\sigma}\right)\right\} \mathrm{d}\gamma =\\ =& \frac{1}{\Gamma(c+1)} \left(\frac{\beta(1-\sigma)}{\sigma}\right)^{c+1} \times \Gamma(c+1) \left(\frac{\sigma}{\sigma(1-\sigma)I + \beta(1-\sigma)}\right)^{c+1} =\\ =& \left(1 + \frac{\sigma}{\beta} I\right)^{-(c+1)},\end{aligned}$$

which is the Laplace functional of the negative binomial random measure with distribution

$$\mathrm{BN}\left(c+1, \frac{\sigma}{\beta} s^{-1-\sigma}, \alpha\right).$$

The thesis follows From Proposition 2.4.1. □

2.D IBP as nonparametric prior in Survival Analysis

Let us consider the model (2.9). The aim of this section is to present some prior and posterior results assuming a 3-parameter IBP as nonparametric prior for the cumulative hazard function. The proofs are a direct application of the general results regarding CRMs. The first result is the prior expected value for the cumulative hazard in model (2.9).

Proposition 2.D.1. *The prior expected value for $\tilde{\mu} \sim IBP(\gamma, \theta, \sigma)$ is*

$$\mathbb{E}\left[\tilde{\mu}(0,t]\right] = \gamma \frac{\Gamma(1+\theta)}{\Gamma(1-\sigma)\Gamma(\theta+\sigma)} B(1-\sigma, \theta+\sigma)\alpha((0,t]) \tag{2.18}$$

for any $t \geq 0$.

Then, the following proposition summarizes the expression for the marginal distribution arising from model (2.9).

Proposition 2.D.2. *The distribution of the data $\boldsymbol{D}$ arising from the model in (2.9) is infinitesimally equal to*

$$\begin{aligned}
&\gamma^{K^*}\left(\frac{\Gamma(1+\theta)}{\Gamma(1-\sigma)\Gamma(\theta+\sigma)}\right)^{K^*} \times \\
&\times \prod_{r=1:\Theta_r^*=1}^{K} \left[\alpha(\mathrm{d}X_r^*) B(n_r-\sigma, \bar{n}_{r+1}+\tilde{n}_r^c+\theta+\sigma)\right] \times \\
&\times \exp\left\{-\gamma \frac{\Gamma(1+\theta)}{\Gamma(1-\sigma)\Gamma(\theta+\sigma)} \sum_{r=1}^{K} \alpha((X_{r-1}^*, X_r^*]) \sum_{k=1}^{N_r^*} (-1)^{k-1} \binom{N_r^*}{k} B(k-\sigma, \theta+\sigma)\right\}.
\end{aligned} \tag{2.19}$$

The posterior distribution of a IBP can be described as follows.

Proposition 2.D.3. *Let us consider the model in (2.9) with a IBP prior, i.e., with*

$$\tilde{\mu} \sim IBP(\gamma, \theta, \sigma).$$

Then the posterior distribution can be written as

$$\tilde{\mu} \mid \boldsymbol{D} = \tilde{\mu}_c^* + \sum_{r:\Theta r^*=1}^{K} J_r \delta_{X_r^*}, \tag{2.20}$$

where:

i. $\tilde{\mu}_c^* \sim CRM(\nu^*)$*, where*

$$\nu^*(\mathrm{d}s, \mathrm{d}x) = \gamma \frac{\Gamma(1+\theta)}{\Gamma(1-\sigma)\Gamma(\theta+\sigma)} s^{-1-\sigma}(1-s)^{Y(x)+\theta+\sigma-1} \mathrm{d}s \alpha(\mathrm{d}x),$$

ii. the jumps J_r*'s are distributed as a*

$$Beta(n_r - \sigma, \bar{n}_{r+1} + \tilde{n}_r^c + \theta + \sigma).$$

Note that ν^* is the Leǘy intensity of an IBP with updated parameters

$$\begin{aligned}
\gamma' &= \gamma \frac{\Gamma(1+\theta)\Gamma(Y(x)+\theta+\sigma)}{\Gamma(Y(x)+\theta+1)\Gamma(\theta+\sigma)}, \\
\theta' &= \theta + Y(x), \\
\sigma' &= \sigma,
\end{aligned}$$

i.e., IBP is a conjugate prior for survival times. Finally, let us state the expression for the posterior estimator of a IBP under a quadratic loss.

Proposition 2.D.4. *Let us consider the model in* (2.9) *with a IBP prior, i.e., with*

$$\tilde{\mu} \sim IBP(\gamma, \theta, \sigma).$$

Then the posterior estimator of $\tilde{\mu}$ *is*

$$\begin{aligned}
\mathbb{E}\left[\tilde{\mu}(0,t] \mid \boldsymbol{D}\right] =& \gamma \frac{\Gamma(1+\theta)}{\Gamma(1-\sigma)\Gamma(\theta+\sigma)} \sum_{r=1}^{K_t} \alpha((X_{r-1}^*, X_r^*]) B(1-\sigma, N_r^* + \theta + \sigma) + \\
&+ \sum_{r=1:\Theta_r^*=1, X_r^* \leq t}^{K} \frac{n_r - \sigma}{\bar{n}_{r+1} + \tilde{n}_r^c + n_r + \theta}.
\end{aligned} \tag{2.21}$$

2.E Stable Beta-Stacy Scaled priors for Survival Functions

The aim of this section is to transpose all the results for the Stable-Beta Scaled priors for cumulative hazard functions to the scale of the survival function, hence providing a nonparametric scaled prior for survival functions in a exchangeable framework. Almost all the results can be recovered as direct transformations of the results already discussed for the SB-SPs in Section 2.4.
Let us therefore consider the SBS-SP model (2.13). First of all, let us state and prove the result summarizing the prior expected value for the survival function under this model.

Corollary 2.E.1. *Let us consider the model* (2.13) *with a SBS-SP prior. Then the expected value of* $\tilde{S}$ *under the prior is equal to*

$$\mathbb{E}\left[\tilde{S}_t\right] = \left(\frac{\beta}{\frac{\sigma}{1-\sigma}\alpha((0,t]) + \beta}\right)^{c+1}$$

for any $t \geq 0$.

Proof. Note that

$$\begin{aligned}
\mathbb{E}\left[\tilde{S}_t\right] &= \mathbb{E}\left[e^{-\tilde{\mu}^F_{\Delta_{1,h_{c,\beta}}}(0,t]}\right] = \mathbb{E}\left[\mathbb{E}\left[e^{-\tilde{\mu}^F_{\Delta_{1,h_{c,\beta}}}(0,t]} \mid \Delta_{1,h_{c,\beta}} = y\right]\right] = \\
&= \mathbb{E}\left[\exp\left\{-\int_0^t \int_{\mathbb{R}^+} (1-e^{-s})\nu(\mathrm{d}s, \mathrm{d}x)\right\}\right] = \\
&= \mathbb{E}\left[\exp\left\{-\sigma y^{-\sigma} \int_0^t \int_{\mathbb{R}^+} e^{-s}(1-e^{-s})^{-\sigma} \mathrm{d}s \alpha(\mathrm{d}x)\right\}\right] = \\
&= \mathbb{E}\left[\exp\left\{-\frac{\sigma}{1-\sigma} y^{-\sigma}\alpha((0,t])\right\}\right] = \\
&= \int_{\mathbb{R}^+} \exp\left\{-\frac{\sigma}{1-\sigma} y^{-\sigma}\alpha((0,t])\right\} f_{\Delta_{1,h_{c,\beta}}}(y)\mathrm{d}y.
\end{aligned}$$

Since from Lemma 2.A.2

$$f_{\Delta_{1,h_{c,\beta}}}(y) = \frac{\sigma\beta^{c+1}}{\Gamma(c+1)} y^{-1-(c+1)\sigma} e^{-\beta y^{-\sigma}},$$

it follows that

$$\int_{\mathbb{R}^+} \exp\left\{-\frac{\sigma}{1-\sigma} y^{-\sigma}\alpha((0,t])\right\} f_{\Delta_{1,h_{c,\beta}}}(y)\mathrm{d}y =$$
$$\frac{\sigma\beta^{c+1}}{\Gamma(c+1)} \int_{\mathbb{R}^+} \exp\left\{-\left(\frac{\sigma}{1-\sigma}\alpha((0,t])+\beta\right) y^{-\sigma}\right\} y^{-1-(c+1)\sigma}\mathrm{d}y$$
$$\frac{\sigma\beta^{c+1}}{\Gamma(c+1)} \times \frac{\Gamma(c+1)}{\sigma\left(\frac{\sigma}{1-\sigma}\alpha((0,t])+\beta\right)^{c+1}} = \left(\frac{\beta}{\frac{\sigma}{1-\sigma}\alpha((0,t])+\beta}\right)^{c+1}.$$

Hence the thesis follows. □

Since

$$\lim_{t\to\infty} \alpha((0,t]) = 1$$

it follows that

$$\lim_{t\to\infty} \mathbb{E}\left[\tilde{S}_t\right] = \left(\frac{\beta}{\frac{\sigma}{1-\sigma}+\beta}\right)^{c+1}$$

is the prior estimate for the cure rate under SBS-SP model (2.13).
Let us now state the expression for the posterior distribution of a SBS-SP; the proof is a simple transformation of the posterior expression reported in Corollary 2.4.3 and exploited in its proof (see Section 2.C).

Proposition 2.E.1. *Let us consider the model in* (2.13). *The posterior distribution of* $\tilde{\mu}^F_{\Delta_{1,h_{c,\beta}}} \sim$ *SBS-SP*(ν, h) *can be described as*

$$\tilde{\mu}^F_{\Delta_{1,h_{c,\beta}}} \mid \boldsymbol{D} = \tilde{\mu}^{*,F}_c + \sum_{r:\Theta r^*=1}^{K} J^F_r \delta_{X^*_r}, \tag{2.22}$$

where:

i. $\tilde{\mu}^{*,F}_c$ *is a negative binomial random measure, i.e.,* $\tilde{\mu}^{*,F}_c \sim$ *BN*$(K^*+c+1, \rho^{*,F}, \alpha)$, *where*

$$\rho^{*,F}(s \mid x)\mathrm{d}s = \frac{\sigma}{\beta+\sigma\eta_\sigma} e^{-s(Y(x)+1)}(1-e^{-s})^{-1-\sigma}\mathrm{d}s,$$

ii. the jumps J^F_r*'s are such that*

$$1 - e^{-J^F_r} \sim \textit{Beta}(n_r - \sigma, \bar{n}_{r+1} + \tilde{n}^c_r + 1).$$

On the other hand, in order to sample trajectories from the posterior survival function under the model (2.13) it is useful to re-scale the hierarchical construction reported in Theorem 2.3.2 applying it to the model (2.13). It follows the hierarchical construction

$$\begin{aligned} \tilde{\mu}^F_{\Delta_{1,h_{c,\beta}}} \mid \boldsymbol{D}, (\Delta_{1,h_{c,\beta}} = y) &= \tilde{\mu}^{*,F}_c + \sum_{r:\Delta^*_r=1}^{K} J^F_r \delta_{X^*_r} \\ \Delta^{-\sigma}_{1,h_{c,\beta}} \mid \boldsymbol{D} &\sim \text{Gamma}\left(K^* + c + 1, \beta + \sigma\eta_\sigma\right), \end{aligned} \tag{2.23}$$

where:

i. $\tilde{\mu}^{*,F}_c \sim \text{CRM}(\nu^*_F)$, where

$$\nu^*_F(\mathrm{d}s, \mathrm{d}x) = \sigma y^{-\sigma} e^{-s(Y(x)+1)}(1-e^{-s})^{-1-\sigma}\mathbb{1}_{(0,\infty)}(s)\mathrm{d}s\alpha(\mathrm{d}x),$$

ii. the jumps J^F_r's are such that

$$1 - e^{-J^F_r} \sim \text{Beta}(n_r - \sigma, \bar{n}_{r+1} + \tilde{n}^c_r + 1).$$

Finally, let us state and prove the result summarizing the expression of the posterior estimator of the survival function under a quadratic loss.

Corollary 2.E.2. *Let us consider the model in* (2.13) *with a SBS-SP prior, i.e., with*

$$\tilde{\mu}^F_{\Delta_{1,h_{c,\beta}}} \sim \textit{SBS-SP}(\nu_\sigma, h_{c,\beta}).$$

Then the estimator for the survival function under a quadratic loss is

$$\mathbb{E}\left[\tilde{S}_t \mid \boldsymbol{D}\right] = \left(\frac{\beta + \sigma\eta_\sigma}{\beta + \sigma\eta_\sigma + \sigma\eta_t}\right)^{K^*+c+1} \times \prod_{r=1;\Theta^*_r=1,X^*_r\le t}^{K} \frac{\bar{n}_{r+1} + \tilde{n}^c_r + 1}{\bar{n}_{r+1} + \tilde{n}^c_r + n_r + 1 - \sigma},$$

where

$$\eta_t := \sum_{r=1}^{K_t} \alpha\left((X^*_{r-1}, X^*_r \wedge t]\right) B(1-\sigma, N^*_r + 1), \quad \textit{and} \quad K_t = \min\{g : X^*_g \ge t\}.$$

Proof. If

$$\tilde{\mu}^F_{\Delta_{1,h_{c,\beta}}} \sim \text{BS-SP}(\nu_\sigma, h_{c,\beta}).$$

The posterior estimator under a quadratic loss for the survival function $\tilde{S}$ is

$$\mathbb{E}\left[\tilde{S}_t \mid \boldsymbol{D}\right] = \int_{\mathbb{R}^+}\left[\exp\left\{-\int_0^t\int_{\mathbb{R}^+}(1-e^{-s})e^{-s(Y(x)+1)}y\rho(y(1-e^{-s}))\mathrm{d}s\alpha(\mathrm{d}x)\right\}\times\right.$$
$$\left.\times\prod_{r=1;\Theta_r^*=1,X_r^*\le t}^{K}\left(1-\mathbb{E}\left[1-e^{-J_r^F}\right]\right)\right]f_{\Delta_{1,h_{c,\beta}}|\boldsymbol{D}}(y)\mathrm{d}y,$$

where J_r^F is the rth jump of the posterior expression in Proposition 2.E.1. Note that

$$\begin{aligned}
\mathbb{E}\left[\tilde{S}_t \mid \boldsymbol{D}\right] =&\mathbb{E}\left[e^{-\tilde{\mu}^F_{\Delta_{1,h_{c,\beta}}}(0,t]} \mid \boldsymbol{D}\right] = \mathbb{E}_{\Delta_{1,h_{c,\beta}}|\boldsymbol{D}}\left[\mathbb{E}_{\tilde{\mu}^F_{\Delta_{1,h_{c,\beta}}}|\boldsymbol{D},\Delta_{1,h_{c,\beta}}}\left[e^{-\tilde{\mu}^F_{\Delta_{1,h_{c,\beta}}}(0,t]} \mid \boldsymbol{D},\Delta_{1,h_{c,\beta}}\right]\right]\\
=&\mathbb{E}_{\Delta_{1,h_{c,\beta}}|\boldsymbol{D}}\left[\mathbb{E}_{\tilde{\mu}^F_{\Delta_{1,h_{c,\beta}}}|\boldsymbol{D},\Delta_{1,h_{c,\beta}}}\left[e^{-\tilde{\mu}_c^{*,F}(0,t]}\times\prod_{r:\Delta_r^*=1}^{K}e^{-J_r^F\delta_{X_r^*}(0,t]}\right]\right]\\
=&\mathbb{E}_{\Delta_{1,h_{c,\beta}}|\boldsymbol{D}}\left[\mathbb{E}_{\tilde{\mu}^F_{\Delta_{1,h_{c,\beta}}}|\boldsymbol{D},\Delta_{1,h_{c,\beta}}}\left[e^{-\tilde{\mu}_c^{*,F}(0,t]}\right]\times\mathbb{E}_{\tilde{\mu}^F_{\Delta_{1,h_{c,\beta}}}|\boldsymbol{D},\Delta_{1,h_{c,\beta}}}\left[\prod_{r:\Delta_r^*=1}^{K}e^{-J_r^F\delta_{X_r^*}(0,t]}\right]\right].
\end{aligned}$$

The elements of the previous expression can be can be made explicit. In particular,

$$\begin{aligned}
\mathbb{E}_{\tilde{\mu}^F_{\Delta_{1,h_{c,\beta}}}|\boldsymbol{D},\Delta_{1,h_{c,\beta}}}\left[e^{-\tilde{\mu}_c^{*,F}(0,t]}\right] =&\exp\left\{-\int_0^t\int_{\mathbb{R}^+}(1-e^{-s})\nu_F^*(\mathrm{d}s,\mathrm{d}x)\right\}\\
=&\exp\left\{-\int_0^t\int_{\mathbb{R}^+}(1-e^{-s})e^{-s(Y(x)+1)}y\rho(y(1-e^{-s}))\mathrm{d}s\alpha(\mathrm{d}x)\right\},
\end{aligned}$$

while

$$\begin{aligned}
\mathbb{E}_{\tilde{\mu}^F_{\Delta_{1,h_{c,\beta}}}|\boldsymbol{D},\Delta_{1,h_{c,\beta}}}\left[\prod_{r:\Delta_r^*=1}^{K}e^{-J_r^F\delta_{X_r^*}(0,t]}\right] &= \prod_{r:\Delta_r^*=1}^{K}\mathbb{E}\left[e^{-J_r^F}\mathbb{1}_{(0,t]}(X_r^*)+1-\mathbb{1}_{(0,t]}(X_r^*)\right]\\
&= \prod_{r:\Delta_r^*=1}^{K}\mathbb{E}\left[1-(1-e^{-J_r^F})\mathbb{1}_{(0,t]}(X_r^*)\right]\\
&= \prod_{r:\Delta_r^*=1}^{K}\left(1-\mathbb{E}\left[(1-e^{-J_r^F})\mathbb{1}_{(0,t]}(X_r^*)\right]\right).
\end{aligned}$$

It follows that the posterior estimator of $\tilde{S}$ under a quadratic loss can be written as

$$
\begin{aligned}
\mathbb{E}\left[\tilde{S}_t \mid \boldsymbol{D}\right] =& \mathbb{E}_{\Delta_{1,h_{c,\beta}}|\boldsymbol{D}}\left[\exp\left\{-\int_0^t \int_{\mathbb{R}^+} (1-e^{-s})\nu_F^*(\mathrm{d}s,\mathrm{d}x)\right\} \times \right. \\
&\left. \times \prod_{r:\Delta_r^*=1}^{K} \left(1-\mathbb{E}\left[(1-e^{-J_r^F})\mathbb{1}_{(0,t]}(X_r^*)\right]\right)\right] \\
=& \int_{\mathbb{R}^+}\left[\exp\left\{-\int_0^t \int_{\mathbb{R}^+} (1-e^{-s})e^{-s(Y(x)+1)} y\rho(y(1-e^{-s}))\mathrm{d}s\alpha(\mathrm{d}x)\right\}\right. \\
&\left. \times \prod_{r=1;\Theta_r^*=1,X_r^*\le t}^{K} \left(1-\mathbb{E}\left[1-e^{-J_r^F}\right]\right)\right] f_{\Delta_{1,h_{c,\beta}}|\boldsymbol{D}}(y)\mathrm{d}y
\end{aligned}
$$

The previous expression can be simplified by considering

$$
\tilde{\mu}^F_{\Delta_{1,h_{c,\beta}}} \sim \text{SBS-SP}(\nu_\sigma, h_{c,\beta}).
$$

In this case, the jumps J^rs are such that

$$
1-e^{-J_r^F} \sim \text{Beta}(n_r-\sigma, \bar{n}_{r+1}+\tilde{n}_r^c+1),
$$

so it follows that

$$
\prod_{r=1;\Theta_r^*=1,X_r^*\le t}^{K} \left(1-\mathbb{E}\left[1-e^{-J_r^F}\right]\right) = \prod_{r=1;\Theta_r^*=1,X_r^*\le t}^{K} \frac{\bar{n}_{r+1}+\tilde{n}_r^c+1}{\bar{n}_{r+1}+\tilde{n}_r^c+n_r+1-\sigma}.
$$

Moreover, in this case the kernel density ρ is such that

$$
\rho(y(1-e^{-s})) = \sigma y^{-1-\sigma}(1-e^{-s})^{-1-\sigma},
$$

so it follows that

$$
\begin{aligned}
&\int_0^t \int_{\mathbb{R}^+} (1-e^{-s})e^{-s(Y(x)+1)} y\rho(y(1-e^{-s}))\mathrm{d}s\alpha(\mathrm{d}x) = \\
=&\sigma y^{-\sigma}\int_0^t \int_{\mathbb{R}^+} e^{-s(Y(x)+1)}(1-e^{-s})^{-\sigma}\mathrm{d}s\alpha(\mathrm{d}x) = \\
=&\sigma y^{-\sigma}\int_0^t \mathrm{B}(1-\sigma, Y(x)+1)\alpha(\mathrm{d}x) = \sigma y^{-\sigma}\sum_{r=1}^{K_t}\mathrm{B}(1-\sigma, N_r^*+1)\alpha((X_{r-1}^*, X_r^*\wedge t]) = \sigma y^{-\sigma}\eta_t.
\end{aligned}
$$

So the posterior estimator of $\tilde{S}$ under a quadratic loss with a SBS-SP prior can be written as

$$\int_{\mathbb{R}^+} \left[\exp\left\{ -\int_0^t \int_{\mathbb{R}^+} (1-e^{-s}) e^{-s(Y(x)+1)} y \rho(y(1-e^{-s})) \mathrm{d}s \alpha(\mathrm{d}x) \right\} \times \right.$$
$$\left. \times \prod_{r=1;\Theta_r^*=1,X_r^*\leq t}^{K} \left(1 - \mathbb{E}\left[1 - e^{-J_r^F}\right]\right) \right] f_{\Delta_{1,h_{c,\beta}}|\boldsymbol{D}}(y)\mathrm{d}y =$$
$$= \int_{\mathbb{R}^+} e^{-\sigma y^{-\sigma} \eta_t} f_{\Delta_{1,h_{c,\beta}}|\boldsymbol{D}}(y)\mathrm{d}y \times \prod_{r=1;\Theta_r^*=1,X_r^*\leq t}^{K} \frac{\bar{n}_{r+1} + \tilde{n}_r^c + 1}{\bar{n}_{r+1} + \tilde{n}_r^c + n_r + 1 - \sigma}.$$

Moreover, the posterior density of $\Delta_{1,h_{c,\beta}}$ is

$$f_{\Delta_{1,h_{c,\beta}}|\boldsymbol{D}}(y) = \frac{\sigma\left(\beta + \sigma\eta_\sigma\right)^{K^*+c+1}}{\Gamma(K^*+c+1)} y^{-(K^*+c+1)\sigma-1} e^{-(\beta+\sigma\eta_\sigma)y^{-\sigma}},$$

so the integral in the previous expression can be simplified as follows:

$$\int_{\mathbb{R}^+} e^{-\sigma y^{-\sigma}\eta_t} f_{\Delta_{1,h_{c,\beta}}|\boldsymbol{D}}(y)\mathrm{d}y = \frac{\sigma\left(\beta + \sigma\eta_\sigma\right)^{K^*+c+1}}{\Gamma(K^*+c+1)} \int_{\mathbb{R}^+} y^{-(K^*+c+1)\sigma-1} e^{-(\beta+\sigma\eta_\sigma+\sigma\eta_t)y^{-\sigma}} \mathrm{d}y =$$
$$= \frac{\sigma\left(\beta + \sigma\eta_\sigma\right)^{K^*+c+1}}{\Gamma(K^*+c+1)} \times \frac{\Gamma(K^*+c+1)}{\sigma\left(\beta + \sigma\eta_\sigma + \sigma\eta_t\right)^{K^*+c+1}} =$$
$$= \left(\frac{\beta + \sigma\eta_\sigma}{\beta + \sigma\eta_\sigma + \sigma\eta_t}\right)^{K^*+c+1}.$$

Hence the thesis follows. □

2.F Material on sampling algorithm

The aim of this section is to present the full conditional distributions of the parameters σ, β and c of the model (2.4) under a SB-SP prior, in order to rely on a full Bayesian approach as described in Section 2.5. The full-conditional distributions presented in the following are useful for both a marginal and conditional algorithm; moreover, we assume two Gamma distributed priors for β and c, and a Beta distributed prior for σ.

Let us recall that according to Corollary 2.4.2 the distribution of the data $\boldsymbol{D}$ arising from the model in (2.13) with a SB-SP prior is infinitesimally equal to

$$\frac{\Gamma\left(K^*+c+1\right)\sigma^{K^*}\beta^{c+1}}{\Gamma(c+1)\left(\beta+\sigma\eta_\sigma\right)^{K^*+c+1}}\times\left[\prod_{r=1:\Theta_r^*=1}^{K}\alpha(\mathrm{d}X_r^*)\mathrm{B}\left(n_r-\sigma,\bar{n}_{r+1}+\tilde{n}_r^c+1\right)\right],$$

where

$$\eta_\sigma=\sum_{r=1}^{K}\alpha((X_{r-1}^*,X_r^*])\sum_{k=1}^{N_r^*}(-1)^{k-1}\binom{N_r^*}{k}\frac{1}{k-\sigma}\quad\text{and}\quad N_r^*=\sum_{g=r}^{K}n_g^*,$$

and n_g^* is the number of observations equal to X_g^*. Let us now state the expressions of the full-conditional distributions of σ, β and c, which can be directly derived from Corollary 2.4.2.

Let us consider a Beta(A,B) prior on the parameter σ. Then the full conditional distribution of σ is

$$\pi\left(\sigma\mid\beta,c,\boldsymbol{D}\right)=\frac{\sigma^{K^*+A-1}(1-\sigma)^{B-1}}{\left(\beta+\sigma\eta_\sigma\right)^{K^*+c+1}}\times\left[\prod_{r=1:\Theta_r^*=1}^{K}\mathrm{B}\left(n_r-\sigma,\bar{n}_{r+1}+\tilde{n}_r^c+1\right)\right].$$

On the other hand, if we consider a Gamma(shape $=\kappa,\text{rate}=\theta$) prior on the parameter β, the full conditional distribution of β is

$$\pi\left(\beta\mid\sigma,c,\boldsymbol{D}\right)=\frac{\beta^{c+\kappa}e^{-\theta\beta}}{\left(\beta+\sigma\eta_\sigma\right)^{K^*+c+1}}.$$

Finally, if we consider a Gamma(shape $=\kappa,\text{rate}=\theta$) prior on the parameter c, the full conditional distribution of c is

$$\pi\left(c\mid\sigma,\beta,\boldsymbol{D}\right)=\frac{\beta^{c+1}c^{\kappa-1}e^{-\theta c}}{\left(\beta+\sigma\eta_\sigma\right)^{K^*+c+1}}\prod_{r=0}^{K^*-1}(K^*+c-r).$$

Chapter 3

Hierarchical neutral to the right priors

3.1 Introduction

Neutral to The Right processes are popular tools in Bayesian Nonparametrics because of their central role within the exchangeable framework as nonparametric priors in survival analysis. These processes were firstly introduced in the seminal Doksum (1974) as conjugate nonparametric priors for exchangeable data, even in the presence of censored observations, while a closed form representation of the posterior NTR process was provided in Ferguson and Phadia (1979). On the other hand Completely Random Measures (CRMs), introduced in Kingman (1967), play a key role in Bayesian Nonparametrics since they allow the construction of several nonparametric priors, as described for example in Lijoi and Prünster (2010). In particular, each NTR process can be represented as functional of a CRM: this one-to-one correspondence between NTR processes and CRMs is a key theoretical characterization which will be widely used in this work. The appealing conjugacy property enjoyed by NTR processes has provided a useful theoretical tool for the study of nonparametric priors for exchangeable survival times; see for example Hjort (1990) and Walker and Muliere (1997), where two specific conjugate NTR priors are introduced for cumulative hazards and survival functions respectively. Other approaches to Bayesian nonparametric modelling in survival analysis problems under the exchangeability assumption can be found for example in Dykstra and Laud (1981) and in and Lo and Chung-Sing (1989), who introduced a mixture hazard prior with a gamma process as mixing measure, and in Ishwaran and James (2004), who used a general class of random hazard rate-based models generalizing the previous two works from a nonparametric setting on $\mathbb{R}^+$ to a setting over general spaces.
On the other hand, the exchangeability assumption fails to hold when we consider observations coming from different (though similar) populations. Consider for example a clinical study conducted on the same drug but in different hospitals: reasonably, the survival times

of the patients can be considered exchangeable within each hospital, while there may be factors specific to each hospital that express great influence on the observations. When (as in the example) we consider observations from $d > 1$ groups in which the assumption of exchangeability can be held only within each group, partial exchangeability is a more suitable assumption for the dependence of the data. This motivates the extension of Bayesian nonparametric models to the partially exchangeable setting. The BNP literature has introduced numerous models of this kind. Notable examples in this area includes nonparametric priors such as the dependent vector of hazard rates introduced in Lijoi and Nipoti (2014) and the two-dimensional extension of NTR priors proposed in Epifani and Lijoi (2010): the last paper is particularly important for our work since it introduces the idea of generalizing the NTR processes to the partially exchangeable framework. The work presented in Riva Palacio and Leisen (2018) extends Epifani and Lijoi (2010) introducing a nonparametric model for the survival functions of $d \geq 2$ groups of survival times, modeling the dependence structure via Lévy copulas. Moreover, in this context, hierarchical processes are hugely popular Bayesian nonparametric models since they are ideally suited to model relationships across multiple samples which may show some kind of dependence; for example, they may share the values of distinct observations. More precisely, the general structure of a hierarchical process is

$$\begin{aligned}(\tilde{p}_1, \ldots, \tilde{p}_d) \mid \tilde{p}_0 &\overset{\text{i.i.d.}}{\sim} \tilde{\mathcal{L}}_0, \\ \tilde{p}_0 &\sim \mathcal{L}_0,\end{aligned}$$

where $\tilde{\mathcal{L}}_0$ is the probability distribution of each random probability measure $\tilde{p}_i$ such that $\mathbb{E}_{\tilde{\mathcal{L}}_0}[\tilde{p}_i \mid \tilde{p}_0] = \int p \tilde{\mathcal{L}}_0(\mathrm{d}p) = \tilde{p}_0$, whereas $\mathcal{L}_0$ is such that $\mathbb{E}_{\mathcal{L}_0}[\tilde{p}_0] = \int p \mathcal{L}_0(\mathrm{d}p) = P_0$ for some fixed non-atomic probability measure P_0. In this setting, the vector of random probability measures $(\tilde{p}_1, \ldots, \tilde{p}_d)$ defines a prior for the probability distributions of d partially exchangeable samples, while the measure $\tilde{p}_0$ induces dependence across samples. The first example of a hierarchical process is the hierarchical Dirichlet process introduced in Teh et al. (2006), and since then, the literature has extensively addressed the topic. A study on general representations and properties for this class of processes can be found in Camerlenghi et al. (2019). A notable example in the context of survival analysis can be found in Camerlenghi et al. (2021), where the authors present a class of multivariate mixtures whose distribution acts as a prior for the vector of sample-specific baseline hazard rates.

The contribution presented in this work fits within the research stream of Bayesian nonparametrics applied to survival analysis, particularly in the presence of partially exchangeable data. Following the mentioned literature, out aim is to introduce a hierarchical model that generalizes NTR processes to the partially exchangeable setting, hence naturally obtaining a suitable family of nonparametric priors for partially exchangeable and right-censored survival times. Although hierarchical processes are usually very complex, and handling them typically requires introducing a large number of latent variables, we will show how our model, on the contrary, allows for the introduction of a limited number of them while still leading to closed-form formulas.

The outline of the chapter is as follows. In Section 3.2, after defining the notation, the family of hierarchical Neutral To the Right (hNTR) processes is introduced, pointing out how they can be suitable nonparametric priors for partially exchangeable survival times. A complete analysis of the posterior behaviour of the hNTR processes is described in Section 3.3; the main result of this work, i.e., the posterior characterization of a hNTR prior, is stated and analyzed in this section. In Section 3.4 an example of hNTR process is presented; this process, called hierarchical beta-stacy, is then used as a hierarchical prior for partially exchangeable survival times, studying its properties and posterior behavior; therefore, a marginal and a conditional algorithms are proposed in Section 3.5 in order to approximate the posterior distribution of this process. Finally some applications of the hierarchical beta-stacy prior on simulated and real datasets are shown in Section 3.6. Proofs and other technical details are deferred to the Appendix.

3.2 NTR processes in Survival Analysis

The aim of this section is to extend Neutral to The Right (NTR) processes to the partially exchangeable framework. To do so, in Section 3.2.1 we recall some basic concepts regarding NTR processes as nonparametric priors on exchangeable survival times, while in Section 3.2.2 we extend them defining the family of hierarchical Neutral to the Right (hNTR) processes. Finally, Section 3.2.3 shows how hNTR processes can be used as nonparametric priors for a set of partially exchangeable survival times.

3.2.1 Background material

Exchangeability is an ubiquitous assumption in Bayesian nonparametrics, which entails homogeneity of observations. A sequence of observations $(X_n)_{n\geq 1}$ is exchangeable if its law is invariant under finite permutations of its elements. The celebrated de Finetti theorem states that this is equivalent to the existence of a random probability measure $\tilde{p}$ such that

$$X_i \mid \tilde{p} \overset{\text{i.i.d.}}{\sim} \tilde{p}$$

for any $i \geq 1$. When the X_i's are survival times, the usual choice for $\tilde{p}$ is a neutral to the right process. For further details see Doksum (1974). The reason for the popularity of these processes as nonparametric priors for exchangeable survival times is given by their conjugacy property: a NTR prior leads to a NTR posterior (even on censored data), whose closed form representation was provided in Ferguson and Phadia (1979). A useful property of NTR processes, widely used in the literature, is that they can be characterized as functionals of specific random probability measures known as completely random measures; for an exhaustive treatment of the topic, refer to Kingman (1967); Daley and Vere-Jones (2008); Lijoi and Prünster (2010). To fix the notation, let us denote by $(\Omega, \mathcal{A}, \mathbb{P})$ a probability space, and let $(\mathbb{X}, \mathcal{X})$ be a measurable space, where $\mathbb{X}$ is a Polish space with its Borel σ-algebra $\mathcal{X}$. Let us denote by $\boldsymbol{M}$ the space of boundedly finite measures on $(\mathbb{X}, \mathcal{X})$, and by $\mathcal{M}$ the corresponding Borel σ-algebra. Completely random measures (CRMs), i.e., measurable maps from $(\Omega, \mathcal{A})$ to $(\boldsymbol{M}, \mathcal{M})$ which map disjoint events in $\mathcal{X}$ to independent random variables, were introduced in Kingman (1967), and they play a key role in Bayesian Nonparametrics since they allow the construction of several nonparametric priors, as described for example in Lijoi and Prünster (2010). In this chapter we will focus on CRMs without fixed atoms and without deterministic drifts: as shown in Kingman (1967), each of these is a transformation of a marked Poisson process, therefore its distribution can be characterized by means of its Laplace functional, which equals

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

for any measurable function $f : \mathbb{X} \to \mathbb{R}$ such that $\int |f| \mathrm{d}\tilde{\mu} < \infty$ almost surely. The measure ν in the previous expression is called *Lévy intensity measure* of the CRM $\tilde{\mu}$. It is well known that the Lévy intensity characterizes the CRM: we will write $\tilde{\mu} \sim \mathrm{CRM}(\nu)$ to denote the distribution of a CRM $\tilde{\mu}$ having Lévy intensity ν.

As for the exchangeable framework, Doksum (1974) shows that a NTR process on $\mathbb{R}^+$ is completely characterized by a CRM. In particular, given a random probability measure $\tilde{p}$, the process $\{\tilde{p}((0,t]) : t \geq 0\}$ is neutral to the right if and only if there exists a completely random measure $\tilde{\mu}$ on $\mathbb{R}^+$ such that

$$\{\tilde{p}((0,t]) : t \geq 0\} \overset{\mathrm{d}}{=} \{1 - e^{-\tilde{\mu}((0,t])} : t \geq 0\}, \quad \text{where} \quad \mathbb{P}\left[\lim_{t\to\infty} \tilde{\mu}((0,t]) = \infty\right] = 1.$$

Let us write $\tilde{p} \sim \text{NTR}(\tilde{\mu})$. This one-to-one correspondence between NTR processes and CRMs is a key theoretical characterization which will be widely used in this work. Despite the conjugacy property holds for any NTR process, in applications it is often useful to consider specific NTR processes which present useful theoretical or computational features, for example a posterior distribution easy to compute. For example, the beta process introduced in Hjort (1990) has a posterior form that is still a beta process, with updated parameters; another example of conjugate NTR process is the beta-stacy process, introduced in Walker and Muliere (1997) as nonparametric prior for survival functions in the exchangeable setting. In particular, let α be a probability measure on $\mathbb{R}^+$ which is absolutely continuous with respect to the Lebesgue measure, and let $c : \mathbb{R}^+ \to \mathbb{R}^+$ be a piecewise continuous function. Let us recall that the random probability distribution $\tilde{F}$ is a beta-stacy process with parameters c and α if $\tilde{F}(t) = \tilde{p}((0,t])$ for each time $t > 0$ and $\tilde{p} \sim \text{NTR}(\tilde{\mu})$, where $\tilde{\mu}$ is a CRM whose Lévy intensity is

$$\nu(\mathrm{d}s, \mathrm{d}x) = \frac{e^{-sc(x)\alpha((x,\infty))}}{1 - e^{-s}} c(x) \mathrm{d}s \alpha(\mathrm{d}x).$$

In such a case we say that the CRM $\tilde{\mu}$ is a log-Beta measure with parameters c and α. In the sequel we write $\tilde{F} \sim \text{Beta-Stacy}(c, \alpha)$ and $\tilde{\mu} \sim \text{log-Beta}(c, \alpha)$. It is known from Walker and Muliere (1997) that in a model for exchangeable survival times, the beta-stacy process (such as the beta process) is a conjugate nonparametric prior, i.e., assuming a beta-stacy prior the posterior is a beta-stacy process with updated parameters.

3.2.2 Hierarchical neutral to the right processes

Exchangeability is too restrictive when the data are organized in groups, for example when data are survival times of patients with the same pathology who are undergoing different treatments. In these cases the most appropriate assumption is the partial exchangeability. Therefore, let us assume that the data are d groups of $\mathbb{X}$-valued observations, where we denote by $X_{i,j}$ the ith observation of group j, for $j = 1, \ldots, d$, where N_j is the number of observations in group j; each observation $X_{i,j}$ is a $\mathbb{X}$-valued random element defined on the common probability space $(\Omega, \mathcal{A}, \mathbb{P})$. The partial exchangeability of the observations is equivalent to the existence of a vector of dependent random probabilty measures, thanks to the de Finetti representation theorem. More precisely, let us denote by $\boldsymbol{P}$ the space of all probability measures over $(\mathbb{X}, \mathcal{X})$, and by $\boldsymbol{P}^d$ the corresponding d-dimensional product space. Then, the sequences $(X_{i,j})_{i \geq 1}$ are partially exchangeable if and only if there exists a probability measure Q_d over $\boldsymbol{P}^d$ such that

$$\mathbb{P}\left[\bigcap_{j=1}^{d}\bigcap_{i=1}^{N_j}\{X_{i,j}\in A_{i,j}\}\right]=\int_{\boldsymbol{P}^d}\prod_{j=1}^{d}\prod_{i=1}^{N_j}p_i(A_{i,j})Q_d(\mathrm{d}p_1,..,\mathrm{d}p_d),$$

for any $(N_1,\dots,N_d)\in\mathbb{N}^d$ and for any collection of Borel sets $A_{i,j}\in\mathcal{X}$, as $j=1,\dots,d$ and $i=1,\dots,N_j$. The measure Q_d, called de Finetti measure, works as a prior distribution. The choice of an appropriate de Finetti measure for the joint modelling of $d\geq 1$ groups of partially exchangeable survival times is addressed in various ways in the literature; our proposal is to introduce a hierarchical model that represents a natural extension of the NTR processes introduced in Doksum (1974). In particular, the aim of this section is to introduce a new family of nonparametric priors called *hierarchical Neutral to the Right processes*. Let $\tilde{\mu}_0\sim\mathrm{CRM}(\nu)$ be a CRM having Lévy intensity $\nu=\rho(s\mid x)\mathrm{d}s\alpha(\mathrm{d}x)$, where α is a measure on $(\mathbb{X},\mathcal{X})$ and ρ is a transition kernel on $\mathbb{X}\times\mathcal{B}(\mathbb{R}^+)$. It is well known from Kingman (1967) that the CRM $\tilde{\mu}_0$ can be expressed as functional of the points $(\tilde{h}_{0,k},\tilde{x}_k)_{k\geq 1}$ of a marked Poisson point process on $\mathbb{R}^+\times\mathbb{X}$ as follows:

$$\tilde{\mu}_0=\sum_{k\geq 1}\tilde{h}_{0,k}\delta_{\tilde{x}_k}. \tag{3.1}$$

Following Masoero et al. (2018); Camerlenghi et al. (2021), let us now define a hierarchical extension of completely random measures.

Definition 3.2.1 (hCRM)**.** A vector of random measures $(\tilde{\mu}_1,\dots,\tilde{\mu}_d)$ is said to be a vector of *hierarchical Completely Random Measures (hCRMs)* if there exists a CRM $\tilde{\mu}_0$ as in (3.1) such that

$$\tilde{\mu}_j\mid\tilde{\mu}_0\overset{\mathrm{d}}{=}\sum_{k\geq 1}\tilde{h}_{j,k}\delta_{\tilde{x}_k}, \tag{3.2}$$

where $\tilde{h}_{j,k}$ are independent conditionally on $\tilde{\mu}_0$ with probability density function on $\mathbb{R}^+$ given by $f_j(h\mid\tilde{h}_{0,k},\tilde{x}_k,b_j)$, for each $j=1,\dots,d$. Here we denoted by b_j any additional parameter that can be defined in order to specify the jth measure of the hCRM vector. Moreover, $\tilde{\mu}_0$ is called *base measure*.

Remark 3.2.1. The measures $\tilde{\mu}_j$'s are independent conditional on the base measure $\tilde{\mu}_0$. Note that for any $j=1,\dots,d$ the points $\tilde{x}_k$'s are the same atoms of the base measure $\tilde{\mu}_0$ and $\tilde{h}_{j,k}$ represents the kth non-negative jump of the jth random measure $\tilde{\mu}_j$.

As recalled before in this section, in the exchangeable framework there is a one-to-one correspondence between a CRM and a NTR process on $\mathbb{X}=\mathbb{R}^+$; it is therefore natural that in a d-dimensional partially exchangeable framework a vector of hierarchical CRMs as in Definition 3.2.1 can be related one-to-one to a vector of NTR processes. Note that, in general, a d-dimensional vector of CRMs $(\tilde{\mu}_1, \dots, \tilde{\mu}_d)$ on $\mathbb{R}^+$ is related to the corresponding vector of random distributions $(\tilde{F}_1, \dots, \tilde{F}_d)$ such that

$$\tilde{F}_j(t) = 1 - e^{-\tilde{\mu}_j(0,t]},$$

for any $j = 1, \dots, d$ and $t > 0$, and

$$\mathbb{P}\left[\lim_{t\to\infty} \tilde{\mu_j}(0,t] = \infty\right] = 1,$$

for any $j = 1, \dots, d$. Let us now exploit Definition 3.2.1 to extend the definition of neutral to the right processes as follows.

Definition 3.2.2 (hNTR)**.** Let us consider a d-dimensional vector of hCRMs $(\tilde{\mu}_1, \dots, \tilde{\mu}_d)$ with base measure $\tilde{\mu}_0$ as in Definition 3.2.1. The vector of NTR processes $(\tilde{F}_1, \dots, \tilde{F}_d)$ such that

$$\tilde{F}_j(t) = 1 - e^{-(\tilde{\mu}_j|\tilde{\mu}_0)(0,t]} \quad \text{and} \quad \mathbb{P}\left[\lim_{t\to\infty} (\tilde{\mu_j} \mid \tilde{\mu}_0)(0,t] = \infty\right] = 1$$

for any $j = 1, \dots, d$, and $t > 0$, is called a vector of hierarchical Neutral to the Right (hNTR) processes.

Remark 3.2.2. Note that given a the vector of hCRMs $(\tilde{\mu}_1, \dots, \tilde{\mu}_d)$ with base measure $\tilde{\mu}_0$, since the vector $(\tilde{\mu}_1 \mid \tilde{\mu}_0, \dots, \tilde{\mu}_d \mid \tilde{\mu}_0)$ is composed by CRMs on $\mathbb{R}^+$, each distribution $\tilde{F}_j$ as in Definition 3.2.2 is a NTR process. Note also that Definition 3.2.2 introduces a *family of hierarchical processes* whose elements are vectors of hNTR processes which, according to Definition 3.2.1, are completely specified choosing a base measure $\tilde{\mu}_0$ and the law of the conditional jumps $\tilde{h}_{j,k} \sim f_j(h \mid \tilde{h}_{0,k}, \tilde{x}_k, b_j)$ for each $j = 1, \dots, d$.

Therefore, this family is a hierarchical extension of the neutral to the right processes, characterized by a hierarchical extension of the completely random measures. This makes it theoretically interesting as a family of nonparametric priors for partially exchangeable survival times.

3.2.3 hNTR priors for Survival Analysis

We now discuss how to use hNTR processes in presence of partially exchangeable survival times. We also introduce the main objects of interest for our inferential goals (survival functions). Therefore, let us assume that $\mathbb{X} = \mathbb{R}^+$ and let us assume to be provided with d

groups of survival times $(T_{1,1}, \dots, T_{N_1,1})$, ..., $(T_{1,d}, \dots, T_{N_d,d})$, along with the d groups of corresponding right censoring variables $(\Delta_{1,1}, \dots, \Delta_{N_1,1})$, ..., $(\Delta_{1,d}, \dots, \Delta_{N_d,d})$, where

$$\Delta_{i,j} = \begin{cases} 1 & \text{if } T_{i,j} \leq C_{i,j} \\ 0 & \text{otherwise} \end{cases},$$

and let us define as $X_{i,j} = \min(T_{i,j}, \Delta_{i,j})$ the ith observation of the jth group. In the sequel let us denote by $\boldsymbol{D}$ the vector containing all the observations $X_{i,j}$ and the corresponding variables $\Delta_{i,j}$, namely

$$\boldsymbol{D} := \{(X_{i,j}, \Delta_{i,j}) : \ j = 1, \dots, d, \, i = 1, \dots, N_j\},$$

where N_j is the number of observations from the jth group. The observations, which are $\mathbb{R}^+$-valued survival times, are assumed to come from an infinite array of partially exchangeable random variables. In particular, the complete survival model is

$$\begin{aligned} (X_{i_1,1}, \dots, X_{i_d,d}) \mid (\tilde{p}_1, \dots, \tilde{p}_d) &\overset{\text{ind}}{\sim} \tilde{p}_1 \times \dots \times \tilde{p}_d \quad (i_1, \dots, i_d) \in \mathbb{N}^d \\ (\tilde{p}_1, \dots, \tilde{p}_d) &\sim \mathscr{Q}_d, \end{aligned} \tag{3.3}$$

where $\mathscr{Q}_d$ is the joint distribution of the vector of random probability measures. Our aim is to consider hierarchical processes from the hNTR family described in Section 3.2.2 as prior distributions in the partially exchangeable model (3.3), thus extending the NTR priors as nonparametric priors for exchangeable survival times. Let us therefore assume a hNTR prior in (3.3), that is, the distribution $\mathscr{Q}_d$ is the law of the hNTR vector $(\tilde{p}_1, \dots, \tilde{p}_d)$. Therefore, thanks to the representation (3.2), the jth random measure in (3.3) under a hNTR prior is

$$\tilde{p}_j(0, t] = 1 - e^{-\sum_{k\geq 1} \tilde{h}_{j,k} \mathbb{1}_{(0,t)}(\tilde{x}_k)} = 1 - \prod_{k \geq 1} e^{-\tilde{h}_{j,k} \mathbb{1}_{(0,t)}(\tilde{x}_k)},$$

for any $j = 1, \dots, d$ and for any $t > 0$.

Let us recall the definition of survival function. Assuming that the survival times of the jth group of observations are sampled i.i.d from a random variable T_j with distribution $H_j(t)$, the survival function for the jth group is

$$S_j(t) = \mathbb{P}\left[T_j > 1\right] = 1 - H_j(t),$$

i.e., $S_j(t)$ is the probability for a subject from group jth of being still alive at time t. As a consequence the random survival function under a hNTR prior in (3.3) is

$$\tilde{S}_j(t) = 1 - \tilde{F}_j(t) = \prod_{k\geq 1} e^{-\tilde{h}_{j,k}\mathbb{1}_{(0,t)}(\tilde{x}_k)},$$

for any $j = 1, \dots, d$ and for any $t > 0$.
Note that, following this approach, each group of partially exchangeable observations is modeled by a different element of the hCRM vector $(\tilde{\mu}_1, \dots, \tilde{\mu}_d)$, while the base measure $\tilde{\mu}_0$ induces (prior) dependence between the groups; this induces a sharing of information between the random survival functions of the different groups. Note for example that the locations $\tilde{x}_k$'s of the jumps of the CRMs are shared and that they coincide with the locations of the jumps of the base measure $\tilde{\mu}_0$; furthermore, the law of the conditional jumps $\tilde{h}_{j,k} \sim f_j(h \mid \tilde{h}_{0,k}, \tilde{x}_k, b_j)$, for each $j = 1, \dots, d$, depends both on the locations $\tilde{x}_k$'s and the jumps $\tilde{h}_{0,k}$ of the base measure. This sharing of information induced by the base measure is known in the literature as *borrowing of information*, or *borrowing of strength*; therefore, the estimate of the survival functions depend on both the group-specific parameters and the estimates of the other groups.

3.3 Posterior Analysis

The previous section introduced the family of hNTR processes and showed how they represent a natural extension of NTR processes as nonparametric priors for partially exchangeable survival times. The aim of this section is to study the marginal and posterior distribution of the survival model (3.3) under a hNTR prior, as introduced in Section 3.2.3. The results presented in this section provides theoretical tools to study the marginal and posterior distributions of a generic hNTR prior; these results will be exploited in the next sections to define sampling algorithms for the marginal and conditional estimation of the posterior survival functions under a hNTR prior. In the Section 3.4 an example of hNTR prior will be introduced and these results will be applied to study its marginal and posterior behaviour.
Before stating the main results of this section, let us introduce some useful notation. Note that a hNTR process is a discrete prior, hence there may be ties among the observations sampled from our model; therefore, let us denote by $X_1^*, \dots, X_K^*$ the K distinct observations out of all the samples. Moreover, for any $j = 1, \dots, d$ and for any $r = 1, \dots, K$ let us define

$$n_{r,j} = \sum_{i=1}^{N_j} \mathbb{1}_{(X_{i,j}=X_r^*, \Delta_{i,j}=1)} \quad \text{and}$$
$$n_{r,j}^c = \sum_{i=1}^{N_j} \mathbb{1}_{(X_{i,j}=X_r^*, \Delta_{i,j}=0)},$$

i.e., $n_{r,j}$ is the number of exact observations in group j which coincide with the rth distinct value X_r^*, while $n_{r,j}^c$ is the number of censored observations in group j which coincide with the rth distinct value X_r^*. Moreover, for any $r = 1, \ldots, K$, we define the variable

$$\Delta_r^* := \max_{(i,j)\colon X_{i,j}=X_r^*} \Delta_{i,j},$$

which is equal to 1 if and only if there exists an exact observation coinciding with X_r^*. Then, we introduce the at-risk processes referring to population $j \in \{1, \ldots, d\}$:

$$\bar{N}_j(x) := \sum_{i=1}^{N_j} \mathbb{1}_{[x,+\infty)}(X_{i,j}) \mathbb{1}_{\{1\}}(\Delta_{i,j}),$$
$$\tilde{N}_j^c(x) := \sum_{i=1}^{N_j} \mathbb{1}_{[x,+\infty)}(X_{i,j}) \mathbb{1}_{\{0\}}(\Delta_{i,j}).$$

In particular, $\bar{N}_j(x)$ counts the number of exact observations which are at risk after time x, while $\tilde{N}_j^c(x)$ counts the number of censored observations which are at risk after time x. Let us further define the general at-risk process as $N_j(x) = \bar{N}_j(x) + \tilde{N}_j^c(x)$.
The first result we present is a closed form representation of likelihood function assuming a hNTR prior. The following theorem states this result; the proof is reported in Section 3.A.

Theorem 3.3.1. *Consider the model described in* (3.3) *with a hNTR prior. The joint probability distribution of the vector of distinct observations* $(X_1^*, \ldots, X_K^*)$ *and the data* $\boldsymbol{D}$ *is absolutely continuous with respect to the measure* $\prod_{r=1\ \Delta_r^*=1}^{K} \alpha(\mathrm{d}X_r^*)$*, and its Radon–Nikodym* *derivative is given by*

$$
\begin{aligned}
&P_K(X_1^*, \ldots, X_K^*, \boldsymbol{D}) = \\
&\prod_{r=1}^{K}\prod_{\Delta_r^*=1} \left[\int_0^\infty \prod_{j=1}^{d} \int_0^\infty e^{-h(\bar{n}_{r+1,j}+\tilde{n}^c_{r,j})}(1-e^{-h})^{n_{r,j}} f_j(h \mid s, X_r^*, b_j)\mathrm{d}h \rho(s \mid X_r^*)\mathrm{d}s\right] \\
&\times \exp\left\{-\int_{\mathbb{X}}\int_0^\infty \left[1-\prod_{j=1}^{d}\int_0^\infty e^{-hN_j(x)} f_j(h \mid s, x, b_j)\mathrm{d}h\right]\rho(s \mid x)\mathrm{d}s\alpha(\mathrm{d}x)\right\},
\end{aligned} \tag{3.4}
$$

where $\bar{n}_{r+1,j}$ *is the number of exact observations from the* $(r+1)$*th distinct value in group* j*, and* $\tilde{n}^c_{r,j}$ *is the number of censored observations from the* r*th distinct value in group* j*. Moreover,* $f_j(h \mid s, X_r^*, b_j)$ *is the law of the conditional jumps of the hierarchical NTR prior, as introduced in Remark 3.2.2.*

Remark 3.3.1. Note that the expression of the joint distribution reported in (3.4) can be simplified introducing a vector of latent jumps $\boldsymbol{s} = \{s_r : r = 1, \ldots, K \text{ s.t. } \Delta_r^* = 1\}$ to avoid the first integral. These latent jumps are associated to each distinct and exact observation, and they lead to the following augmented version of the likelihood:

$$
\begin{aligned}
&P_K(X_1^*, \ldots, X_K^*, \boldsymbol{D}, \boldsymbol{s}) = \\
&\prod_{r=1}^{K}\prod_{\Delta_r^*=1} \left\{\prod_{j=1}^{d} \int_0^\infty e^{-h(\bar{n}_{r+1,j}+\tilde{n}^c_{r,j})}(1-e^{-h})^{n_{r,j}} f_j(h \mid s_r, X_r^*, b_j)\mathrm{d}h \rho(s_r \mid X_r^*)\mathrm{d}s_r\right\} \\
&\times \exp\left\{-\int_{\mathbb{X}}\int_0^\infty \left[1-\prod_{j=1}^{d}\int_0^\infty e^{-hN_j(x)} f_j(h \mid s, x, b_j)\mathrm{d}h\right]\rho(s \mid x)\mathrm{d}s\alpha(\mathrm{d}x)\right\}.
\end{aligned} \tag{3.5}
$$

The expression of the likelihood reported in (3.5) will be assumed for the rest of the chapter and in the appendix.

The importance of Theorem 3.3.1, together with Remark 3.3.1, in the present work is twofold. First of all the expression of the likelihood allows characterizing the posterior distribution of the generic hNTR prior, as will be shown in the next result. On the other hand, we are able to recover the full conditional distributions of the model parameters from the expression (3.5).
We now move to the second result of the section. The following theorem provides a closed form characterization of the posterior distribution assuming a hNTR prior, even in presence of right-censored survival times; the proof of the theorem is provided in Section 3.A.

Theorem 3.3.2. *Consider the survival model introduced in* (3.3)*, let* $(\tilde{\mu}_1, \dots, \tilde{\mu}_d)$ *be a vector of hierarchical completely random measures as in Definition 3.2.1 and let us assume that* $\boldsymbol{s}$ *is the vector of latent jumps introduced in* (3.5)*. The posterior distribution of the vector of hierarchical completely random measures can be described as follows:*

$$(\tilde{\mu}_1, \dots, \tilde{\mu}_d) \mid \boldsymbol{D}, \boldsymbol{s} \overset{d}{=} (\tilde{\mu}'_1, \dots, \tilde{\mu}'_d) + \left(\sum_{r=1\ \Delta^*_r=1}^{K} J_{1,r}\delta_{X^*_r}, \dots, \sum_{r=1\ \Delta^*_r=1}^{K} J_{d,r}\delta_{X^*_r} \right), \tag{3.6}$$

where all the elements of this representation are defined below.

i. $(\tilde{\mu}'_1, \dots, \tilde{\mu}'_d)$ *is a vector of hierarchical completely random measures that are independent conditionally on a random measure* $\tilde{\mu}'_0$*, where*

$$\tilde{\mu}'_0 = \sum_{k \geq 1} \tilde{h}'_{0,k}\delta_{\tilde{x}'_k}, \quad \tilde{\mu}'_j \mid \tilde{\mu}'_0 \overset{d}{=} \sum_{k \geq 1} \tilde{h}'_{j,k}\delta_{\tilde{x}'_k}.$$

The posterior jumps $\tilde{h}'_{j,k}$ *are independent conditionally on* $\tilde{\mu}'_0$ *with density*

$$f'_j(h \mid \tilde{h}'_{0,k}, \tilde{x}'_k, b_j) \propto e^{-hN_j(x)} f_j(h \mid s, x, b_j),$$

where $N_j(x)$ *is the general at-risk process for group* j *already defined in this section. The base measure* $\tilde{\mu}'_0 \sim CRM(\nu')$ *has updated Lévy intensity as follows:*

$$\nu'(\mathrm{d}s, \mathrm{d}x) = \left[\prod_{j=1}^{d} \int_0^\infty e^{-hN_j(x)} f_j(h \mid s, x, b_j)\mathrm{d}h \right] \rho(s \mid x)\mathrm{d}s\alpha(\mathrm{d}x).$$

ii. The random variables $J_{j,r}$*'s are independent jumps, as well as independent from the hierarchical completely random measures described in the previous point, and they have a density function on* $\mathbb{R}^+$ *proportional to*

$$(1 - e^{-h})^{n_{r,j}} e^{-h(\bar{n}_{r+1,j} + \tilde{n}^c_{r,j})} f_j(h \mid s_r, X^*_r, b_j).$$

Theorem 3.3.2 describes the posterior distribution of the vector of hierarchical CRMs, hence providing a representation of the posterior distribution of a hNTR prior. Note that the representation in (3.6) shows a "structural conjugacy" of the hNTR prior. In fact, the posterior distribution is obtained combining a hNTR process (deriving from the vector of hCRMs $(\tilde{\mu}'_1, \dots, \tilde{\mu}'_d)$ described in (*i*)) and a finite and discrete part located on the

atoms X_r^*'s (whose jumps follows the law described in (ii)). Moreover, note that the second component of the posterior vector in (3.6) induces borrowing of information between the groups: all the locations (the observations X_r^*'s) are shared among the d groups, so whenever an observation is detected in one group an not detected in all the others, the posterior distribution for that group still has a jump located on this observation; obviously, the size of the jumps will depend on the data and on the group-specific parameters, according to (ii). Moreover, the component $\tilde{\mu}_j'$ conditionally on the CRM $\tilde{\mu}_0'$ is a CRM whose expression is completely defined in Theorem 3.3.2. The posterior base measure carries information between the groups modelling the posterior dependence of the observations, since it contributes to defining all the CRMs of the absolutely continuous component in (3.6). The measure $\tilde{\mu}_0'$ is provided in Theorem 3.3.2 via the usual correspondence with its Lévy intensity $\nu'(\mathrm{d}s, \mathrm{d}x)$: this representation allows to sample posterior trajectories exploiting suitable algorithms, such as the one introduced in Wolpert and Ickstadt (1998).
The next section will be focused on an example of a hierarchical NTR process and on the description of its structure and posterior distribution.

3.4 Hierarchical beta-stacy process

As discussed before, the appealing feature of NTR processes in survival analysis is the availability of closed-form posterior distribution, provided in Ferguson and Phadia (1979). However, as we discussed previously, it is useful to consider NTR priors that have a conjugate posterior, such as the beta-stacy process introduced in Walker and Muliere (1997) and whose definition is reported in Section 3.2.1. Let us now consider the hierarchical model defined in Section 3.2.2: the aim of this section is to introduce an example of hNTR process which could be seen as an extension of the beta-stacy process, and to study its posterior behaviour exploiting the results discussed in Section 3.3.

3.4.1 Hierarchical beta-stacy prior

A possible extension of the beta-stacy process to the partially exchangeable framework can be obtained by choosing a log-Beta CRM as the base measure $\tilde{\mu}_0$ in (3.1), as defined in Section 3.2.1, and specifying jumps conditionally distributed as transformations of Beta variables in (3.2). Therefore, let us introduce the following definition.

Definition 3.4.1. (hierarchical beta-stacy) Consider a vector of hCRMs $(\tilde{\mu}_1, \ldots, \tilde{\mu}_d)$ with base measure $\tilde{\mu}_0$ as in Definition 3.2.1, i.e.,

$$\tilde{\mu}_j \mid \tilde{\mu}_0 = \sum_{k\geq 1} \tilde{h}_{j,k}\delta_{\tilde{x}_k} \quad \text{for any } j = 1,\dots,d,$$
$$\tilde{\mu}_0 = \sum_{k\geq 1} \tilde{h}_{0,k}\delta_{\tilde{x}_k}.$$

Let α be a probability measure on $\mathbb{R}^+$ which is absolutely continuous with respect to the Lebesgue measure, and let $c : \mathbb{R}^+ \to \mathbb{R}^+$ be a piecewise continuous function. The *hierarchical beta-stacy process* is a hNTR process whose vector of hCRMs is completely defined as follows:

$$\begin{aligned} \tilde{\mu}_0 &\sim \text{log-Beta}(c,\alpha), \\ 1 - e^{-\tilde{h}_{j,k}} \mid \tilde{\mu}_0 &\sim \text{Beta}(c_j(\tilde{x}_k)F_j(\tilde{h}_{0,k}), c_j(\tilde{x}_k)(1 - F_j(\tilde{h}_{0,k}))), \end{aligned} \tag{3.7}$$

where $c_j : \mathbb{R}^+ \to \mathbb{R}^+$ is a function referring to group $j \in \{1,\dots,d\}$, while F_j is c.d.f. which should encapsulate our prior opinion on the hazard function for the jth group.

Note that Definition 3.4.1 is well-posed: from Remark 3.2.2, to completely define a hNTR process it is sufficient to specify the base measure $\tilde{\mu}_0$ and the conditional laws of the jumps $f_j(\cdot)$, and (3.7) provides these specifications. Note also that the group-specific variables, in the example of the hierarchical beta-stacy process, are the functions c_j and the functions F_j introduced in Definition 3.4.1.
Let us now consider the prior expected value $\mathbb{E}\left[e^{-\tilde{\mu}_j(0,t]}\right]$ of each survival function S_j, as $j = 1,...,d$ and for $t > 0$. Exploiting what has been recalled in Section 3.2.1 and Definition 3.4.1, these value can be analytically computed, as summarized in the following proposition which will be proved in Section 3.B.

Proposition 3.4.1. *The prior expected value for the jth group in the hierarchical Beta Stacy model is equal to*

$$\mathbb{E}\left[e^{-\tilde{\mu}_j(0,t]}\right] = exp\left[-\int_0^t \int_{\mathbb{R}^+} F_j(h)\, \frac{e^{-hc(x)\alpha((x,+\infty))}}{1-e^{-h}} c(x)dh\alpha(dx)\right], \tag{3.8}$$

for any $j = 1,\dots,d$ and any time $t > 0$.

The following proposition provides sufficient conditions to ensure that each survival function $\tilde{S}_j(\cdot)$ induced by a hNTR prior is proper. Let us recall that a survival function $S(t)$ is said to be *proper* if $S(t) \to 0$ as $t \to \infty$. The following results is proved in Section 3.B.

Proposition 3.4.2. *Let $\tilde{S}_j(t)$ be the prior survival function of group j assuming hierarchical beta-stacy prior. Let us assume that for each $j = 1, ..., d$ there exists a positive constant $A_j > 0$ such that*

$$\frac{F_j(s)}{1 - e^{-s}} \geq A_j. \tag{3.9}$$

Then, if the measure α is a uniform measure or the standard exponential measure, the function $\tilde{S}_j(t)$ is proper, i.e.,

$$\lim_{t \to \infty} \tilde{S}_j(t) = 0 \quad \textit{almost surely, for any } j = 1, \ldots, d \textit{ and } t > 0.$$

Remark 3.4.1. Note that the assumption (3.9) in Proposition 3.4.2 is satisfied, for example, for functions F_j's as follows:

$$F_j(s) = 1 - e^{-a_j s}, \quad \text{with} \quad a_j \in \mathbb{N}.$$

In fact, since

$$\frac{1 - e^{-a_j s}}{1 - e^{-s}} = 1 + e^{-s} + e^{-2s} + ... + e^{-(a_j - 1)s} \geq 1$$

for each $a_j \in \mathbb{N}$, it follows that (3.9) is satisfied with $A_j = 1$.

Ultimately, the hierarchical beta-stacy process can be used as nonparametric prior for survival functions of partially exchangeable survival times. The results stated in this section provide a closed form representation for the expected values under a hierarchical beta-stacy prior, as well as sufficient conditions to ensure that the prior survival functions under this model are proper. The next section will be focused on the analysis of the posterior and marginal distributions of the hierarchical beta-stacy process, exploiting the general results for hNTR processes discussed in Section 3.3.

3.4.2 Posterior representations

Let us consider the partially exchangeable model defined in (3.3), where $\tilde{p}_j = 1 - e^{-\tilde{\mu}_j}$ and $(\tilde{\mu}_1, \ldots, \tilde{\mu}_d)$ is the vector of hCRMs of a hierarchical beta-stacy process defined in Definition 3.4.1. Exploiting Theorem 3.3.1 and Theorem 3.3.2 it is possible to obtain the marginal and posterior distributions under a hierarchical beta-stacy prior. In particular, the following result provides the marginal distribution assuming the survival model in (3.3) under a hierarchical beta-stacy prior; the proof is an immediate application of Theorem 3.3.1 and can be found in Section 3.B.

Corollary 3.4.1. *Let us consider the model introduced in* (3.3) *under a hierarchical beta-stacy prior. The joint probability distribution of the vector of distinct observations* $(X_1^*, \dots, X_K^*)$*, the data* $\boldsymbol{D}$ *and the vector* $\boldsymbol{s} = \{s_r : r = 1, \dots, K \text{ s.t. } \Delta_r^* = 1\}$ *of latent jumps introduced in* (3.5) *is absolutely continuous with respect to the measure* $\prod_{r=1\ \Delta_r^*=1}^{K} \alpha(\mathrm{d}X_r^*)$*, and its Radon–Nikodym derivative is given by*

$$P_K(X_1^*, \dots, X_K^*, \boldsymbol{D}, \boldsymbol{s}) =$$
$$\prod_{r=1\ \Delta_r^*=1}^{K} \left\{ \prod_{j=1}^{d} \int_0^\infty e^{-h(\bar{n}_{r+1,j} + \tilde{n}_{r,j}^c)} (1 - e^{-h})^{n_{r,j}} f_j(h \mid s_r, X_r^*, c_j, F_j) \mathrm{d}h \rho(s_r \mid X_r^*) \mathrm{d}s_r \right\}$$
$$\times \exp\left\{ - \int_{\mathbb{X}} \int_0^\infty \left[1 - \prod_{j=1}^{d} \int_0^\infty e^{-hN_j(x)} f_j(h \mid s, x, c_j, F_j) \mathrm{d}h \right] \rho(s \mid x) \mathrm{d}s \alpha(\mathrm{d}x) \right\},$$

where $\bar{n}_{r+1,j}$ *is the number of exact observations from the* $(r+1)$*th distinct value in group* j*, and* $\tilde{n}_{r,j}^c$ *is the number of censored observations from the* r*th distinct value in group* j*. In particular,*

$$\rho(s \mid x) = \frac{e^{-sc(x)\alpha((x,\infty))}}{1 - e^{-s}} c(x)$$

is the kernel density of the Beta-Stacy(c, α) *process and*

$$f_j(h \mid s, x, c_j, F_j) = \frac{\left(1 - e^{-h}\right)^{c_j(x)F_j(s) - 1} e^{-hc_j(x)(1 - F_j(s))}}{Be(c_j(x)F_j(s), c_j(x)(1 - F_j(s)))}$$

is the conditional density of the jumps $\tilde{h}_{j,k}$*'s, where* $Be(\cdot, \cdot)$ *is the beta function.*

Moreover, thanks to Theorem 3.3.2 it is possible to describe the posterior distribution of the vector of hCRMs $(\tilde{\mu}_1, \dots, \tilde{\mu}_d)$ of a hierarchical beta-stacy prior; as previously discussed, this is enough to characterize the posterior distribution of the process. This result is summarized in the following corollary to Theorem 3.3.2, which is proved in Section 3.B.

Corollary 3.4.2. *Consider the partially exchangeable model introduced in* (3.3) *under a hierarchical beta-stacy prior, let* $(\tilde{\mu}_1, \dots, \tilde{\mu}_d)$ *be the prior vector of hCRMs as in Definition 3.4.1 and let us assume that* $\boldsymbol{s}$ *is the vector of latent jumps introduced in* (3.5)*. The posterior distribution of the vector of hierarchical completely random measures can be described as follows:*

$$(\tilde{\mu}_1, \ldots, \tilde{\mu}_d) \mid \boldsymbol{D}, \boldsymbol{s} \stackrel{d}{=} (\tilde{\mu}'_1, \ldots, \tilde{\mu}'_d) + \left(\sum_{r=1}^{K} \sum_{\Delta^*_r = 1} J_{1,r} \delta_{X^*_r}, \ldots, \sum_{r=1}^{K} \sum_{\Delta^*_r = 1} J_{d,r} \delta_{X^*_r} \right),$$

where all the elements of this representation are defined below.

i. $(\tilde{\mu}'_1, \ldots, \tilde{\mu}'_d)$ *is a vector of hierarchical completely random measures that are independent conditionally on a base measure* $\tilde{\mu}'_0$*, where*

$$\tilde{\mu}'_0 = \sum_{k \geq 1} \tilde{h}'_{0,k} \delta_{\tilde{x}'_k}, \quad \tilde{\mu}'_j \mid \tilde{\mu}'_0 \stackrel{d}{=} \sum_{k \geq 1} \tilde{h}'_{j,k} \delta_{\tilde{x}'_k}.$$

The base measure $\tilde{\mu}'_0 \sim CRM(\nu')$ *has Lévy intensity as follows:*

$$\begin{aligned} \nu'(\mathrm{d}s, \mathrm{d}x) = \prod_{j=1}^{d} & \left[\frac{Be(c_j(x)F_j(s), N_j(x) + c_j(x)(1 - F_j(s)))}{Be(c_j(x)F_j(s), c_j(x)(1 - F_j(s)))} \right] \\ & \times \left[\frac{e^{-sc(x)\alpha((x,+\infty))}}{1 - e^{-s}} c(x) \right] ds\alpha(dx), \end{aligned} \tag{3.10}$$

where $Be(\cdot, \cdot)$ *is the beta function. Moreover, the conditional distribution of each jump* $\tilde{h}'_{j,k} \mid \tilde{\mu}'_0$ *is the transformation of a Beta random variable:*

$$1 - e^{-\tilde{h}'_{j,k}} \mid \tilde{\mu}'_0 \sim \mathrm{Beta}(c_j(\tilde{x}'_k)F_j(\tilde{h}'_{0,k}), c_j(\tilde{x}'_k)(1 - F_j(\tilde{h}'_{0,k})) + N_j(\tilde{x}'_k)), \tag{3.11}$$

where $N_j(x)$ *is the general at-risk process for group* j*.*

ii. The random variables $J_{j,r}$*'s are independent jumps, as well as independent from the hierarchical completely random measures described in the previous point, and each one of them is the transformation of a Beta random variable:*

$$1 - e^{-J_{j,r}} \sim \mathrm{Beta}(n_{r,j} + c_j(X^*_r)F_j(s_r), \bar{n}_{r+1,j} + \tilde{n}^c_{r,j} + c_j(X^*_r)(1 - F_j(s_r))). \tag{3.12}$$

Tracing the parallels with the classical beta-stacy process, note that hierarchical beta-stacy is not a conjugate prior. However, Corollary 3.4.2 completely characterizes its posterior distribution, when used as a nonparametric prior in model (3.3). It is possible to derive the posterior estimator of the survival functions under a quadratic loss analytically for each group. The following proposition summarizes this result, which is proved in Section 3.B.

Proposition 3.4.3. *For any $t \geq 0$ the posterior estimator of the jth survival function under a quadratic loss is*

$$\begin{aligned}\mathbb{E}\left[\tilde{S}_j(t) \mid \boldsymbol{D}, \boldsymbol{s}\right] = exp&\left[-\int_0^t \int_{\mathbb{R}+} \frac{c_j\left(x\right) F_j\left(s\right)}{c_j\left(x\right) + N_j\left(x\right)} \nu'\left(dx, ds\right)\right] \times \\ &\times \prod_{r:\Delta_r^*=1} \left(-\frac{n_{r,j} + c_j\left(X_r^*\right) F_j(s_r)}{n_{r,j} + \overline{n}_{r+1,j} + \tilde{n}_{r,j} + c_j\left(X_r^*\right)} \mathbb{1}_{(0,t]}(X_r^*) + 1\right).\end{aligned} \tag{3.13}$$

Note that while Corollary 3.4.2 allows to find trajectories to generate the posterior survival function under a hierarchical beta-stacy model, Proposition 3.4.3 integrates the trajectories out providing the marginal expression (3.13). A conditional algorithm will be discussed in the next section.

3.5 Sampling Algorithm

The aim of this section is to describe a sampling algorithm that allows to sample from the posterior distribution of the survival model (3.3) under a hierarchical beta-stacy prior; this can be used, for instance, to obtain posterior estimates of the survival functions. For this purpose there are two possible strategies: a marginal algorithm, which obtains dependent survival functions based on Proposition 3.4.3, hence marginalizing out the CRMs; and a conditional algorithm, which allows to simulate the trajectories of the posterior survival functions $\tilde{S}_j(\cdot) \mid \boldsymbol{D}, \boldsymbol{s}$ based on Corollary 3.4.2.
To fully define the beta-stacy prior as introduced in Definition 3.4.1, it is necessary to choose the model parameters. In particular, for computational reasons we choose a log-Beta(c, α) CRM as base measure, where c is a positive constant and α is a uniform distribution between 0 and a positive value τ, i.e.,

$$\begin{aligned} c(\cdot) &\equiv c \in \mathbb{R}^+ \text{ and} \\ \alpha &\sim \text{Uniform}(0, \tau), \quad \tau \in \mathbb{R}^+. \end{aligned}$$

Regarding the group-specific parameters introduced in Remark 3.2.2, our prior assumptions are the same for each group; more specifically, we assume that

$$c_j(\cdot) \equiv c_* \in \mathbb{R}^+ \text{ for any } j = 1, \dots, d, \quad \text{and}$$
$$F_j(s) = F(s) = 1 - e^{-s} \text{ for any } j = 1, \dots, d.$$

Finally, we assume a Gamma(κ, θ) prior on the parameter c and a Gamma(κ_*, θ_*) prior on the parameter c_*.

Remark 3.5.1. Note that the measure α is uniformly distributed and the functions $F_j(\cdot)$ satisfy the sufficient condition introduced in Remark 3.4.1 with $a_j = 1$ for each j. Therefore, the hypotheses of the Proposition 3.4.2 are satisfied by our parameters specification, so it follows that the survival functions derived from a hierarchical beta-stacy model with the previous specifications are proper.

3.5.1 Marginal algorithm

A marginal algorithm can be obtained by specializing the posterior estimator in (3.13) with the choices on the model parameters described above. Therefore the posterior estimator for the jth survival function under our assumptions is

$$\begin{aligned}
&\mathbb{E}\left[\tilde{S}_j(t) \mid \boldsymbol{D}, \boldsymbol{s}\right] = \\
&= exp\left[-\frac{c}{\tau}\int_0^t \int_{\mathbb{R}^+} \frac{c_* e^{-sc\left(1-\frac{x}{\tau}\right)}}{c_* + N_j\left(x\right)} \prod_{j=1}^d \left[\frac{Be(c_*F(s), N_j(x) + c_*(1-F(s)))}{Be(c_*F(s), c_*(1-F(s)))}\right] \mathrm{d}s\mathrm{d}x\right] \qquad (3.14)\\
&\times \prod_{r:\Delta_r^*=1} \left(-\frac{n_{r,j} + c_*F(s_r)}{n_{r,j} + \overline{n}_{r+1,j} + \tilde{n}_{r,j} + c_*} \mathbb{1}_{(0,t]}(X_r^*) + 1\right).
\end{aligned}$$

A numerical approximation of each estimator $\mathbb{E}\left[\tilde{S}_j(t) \mid \boldsymbol{D}, \boldsymbol{s}\right]$ can be obtained via a suitable MCMC algorithm; to this end, the full conditional distributions of the latent variables and of the model parameters must be identified. The calculation of the full-conditional distributions is reported in Section 3.C. At each step of the marginal algorithm the latent variable $\boldsymbol{s}$ and the parameters c and c_* are sampled according to their full-conditional distribution; then, the corresponding set of dependent estimated survival functions can be estimated on a time-grid according to (3.14). Finally, each posterior estimator can be obtained via a Monte-Carlo approximation exploiting the samples obtained with the marginal algorithm.

3.5.2 Conditional algorithm

The marginal algorithm discussed in the previous section is useful when it comes to estimate the survival functions. On the other hand, the aim of this section is to describe a conditional algorithm that generates trajectories from the posterior distribution of the vector of hCRMs under a hierarchical beta-stacy prior, with the assumption on the model parameters described before. This can be achieved by exploiting Corollary 3.4.2 and specializing this result with our choices on the model parameters. This sampler is useful since it allows to estimate the actual posterior distribution of the survival functions under our model, as well as credible intervals for the estimated quantities.
In order to obtain a trajectory of the posterior vector of CRMs under our prior we apply a MCMC procedure that samples two components at each step. First of all, we must sample a trajectory of the vector of hCRMs $(\tilde{\mu}'_1, \ldots, \tilde{\mu}'_d)$ exploiting its characterization reported in point (i) of Corollary 3.4.2. To do so, we approximate each CRM $\tilde{\mu}'_j \mid \tilde{\mu}'_0$ relying on the algorithm described in Wolpert and Ickstadt (1998), which is a suitable choice for approximation of completely random measure with non-homogeneous Lévy intensity such as the one defined in (3.10). In particular, given a tolerance ϵ and a maximum number of approximation steps M, this algorithm allows to approximate the CRM $\tilde{\mu}'_0$ described in Corollary 3.4.2, i.e.,

$$\tilde{\mu}'_0 = \sum_{k \geq 1} \tilde{h}'_{0,k} \delta_{\tilde{x}'_k},$$

as the finite sum

$$\tilde{\mu}'^M_0 := \sum_{k=1}^{M} \hat{h}'_{0,k} \delta_{\hat{x}'_k},$$

where the set of locations $(\hat{x}'_k)_{k=1}^M$ is independently sampled from the base measure $\alpha \sim \text{Uniform}(0, \tau)$. Moreover, the approximated jumps $(\hat{h}'_{0,k})_{k=1}^M$ can be sampled as follows. Following Wolpert and Ickstadt (1998) let us rewrite the Lévy intensity $\nu'(\mathrm{d}s, \mathrm{d}x)$ in (3.10) as

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

where

$$\nu'(s,x) = \prod_{j=1}^{d} \left[\frac{Be(c_j(x)F_j(s), N_j(x) + c_j(x)(1-F_j(s)))}{Be(c_j(x)F_j(s), c_j(x)(1-F_j(s)))} \right] \cdot \left[\frac{e^{-sc(x)\alpha((x,+\infty))}}{1-e^{-s}} c(x) \right].$$

As $k = 1, \ldots, M$, the kth approximated jump $\hat{h}'_{0,k}$ is obtained as the zero of the function

$$\xi(s) = \int_s^{\infty} \nu'(v, \hat{x}'_k) \mathrm{d}v - \sigma_k,$$

where $\sigma_k = \sum_{b=1}^{k} E_b$, $E_k \sim$ Exponential(1) for any $k = 1, \ldots, M$. Note that ξ is a decreasing function, so if $\xi(\epsilon) < 0$ then we simply assume $\hat{h}'_{0,k} = 0$. The approximation of the CRM $\tilde{\mu}'_0$ allows to sample trajectories from the vector of hCRMs $(\tilde{\mu}'_1, \ldots, \tilde{\mu}'_d)$ via (3.11). On the other hand, we must sample a trajectory from the finite and discrete component of the posterior vector of CRMs located on the atoms X_r's. This is straightforward exploiting (3.12). Note that the sampling of the latent variable $\boldsymbol{s}$ and of the parameters c and c_* is performed according to their full-conditional distribution as described in Section 3.5.1; the calculation of the full-conditional distributions is reported in Section 3.C. The complete pseudo-code that implements the conditional algorithm just described is reported in Section 3.D.

3.6 Illustrations

To concretely study the estimation of survival functions assuming a hierarchical beta-stacy prior on partially exchangeable survival times, the aim of this section is to discuss a set of simulation studies and an application to a real dataset, applying the conditional algorithm described in Section 3.5.2. The goal of this analysis is two-fold: first of all, to compare the estimated posterior survival functions obtained assuming a hierarchical beta-stacy prior with the survival functions obtained following independent model over the different groups; then, to explore the effect of borrowing of information in our model. The natural comparisons for the hierarchical beta-stacy model, among the models suitable for exchangeable survival times, are the Kaplan-Meier estimator and a nonparametric NTR prior. In particular, we focus on the beta-stacy process as described in Section 3.2.1.
Here we explain the actual law of survival times, used for simulations. Let us consider $d = 2$ groups of survival times, sampled from a group of $K = 5$ distinct observations with different probabilities. Table 3.1 summarizes this setting: the column T_r^* shows the distinct

Table 3.1: *Simulated survival times and their probabilities in each group*

r	$p_{r,1}$	$p_{r,2}$	T_r^*
1	0.1	0.4	2.2685
2	0.4	0.1	2.8614
3	0.0	0.3	5.5131
4	0.2	0.2	6.9647
5	0.3	0.0	7.1947

observed survival times, while the columns $p_{r,1}$ and $p_{r,2}$ shows the different probabilities, for group 1 and 2 respectively, of assuming the value T_r^* for any $r = 1, \dots, K$.

Let us consider two different settings: in the first one, we sample the same number of *uncensored observations* for each group according to the probabilities provided in Table 3.1, while in the second one we *censor a portion of observations in each group*. In both cases, we first consider a dataset where both groups have 20 observations and then a second example where both groups have 100 observations. Following the notation introduced in Section 3.2.3, in the first example we consider $N_1 = N_2 = 20$, while in the second one we consider $N_1 = N_2 = 100$.

The assumptions on the parameters of the hierarchical beta-stacy model are the ones described in Section 3.5; moreover, we select as the left bound of the Uniform base measure $\alpha \sim \text{Uniform}(0, \tau)$ the value τ is large enough such that the interval $(0, \tau)$ includes the whole sample.

As described in Section 3.5, we set two Gamma priors on the parameters c and c_*; in particular, we assume the same Gamma(2, 2) distributions:

$$c \sim \text{Gamma}(2,2) \quad \text{and} \quad c_* \sim \text{Gamma}(2,2).$$

The survival functions from the posterior distribution of our model are sampled running 20.000 steps of the MCMC algorithm discussed in Algorithm 3.5.2, where the number of iterations is 10.000, after further 10.000 steps of burn-in.

We then compare the results of our model with the Kaplan-Meier estimators for both groups. Finally, we also compute the posterior estimation of the survival functions under a beta-stacy prior for each group, i.e.,

$$\begin{aligned}(X_{1,j},\ldots,X_{N_j,j}) \mid \tilde{F} &\sim \tilde{F}\\ \tilde{F} &= 1-e^{-\tilde{\mu}}\\ \tilde{\mu} &\sim logBeta(c,\alpha),\end{aligned}$$

as $j = 1, 2$, where $X_{i,j}$ is the ith observation of group j. For each beta-stacy prior we assume the same parameters for the base-measure $\tilde{\mu}_0$ of our model; in particular, they are both NTR processes whose CRM is a log-Beta(α, c), where $\alpha \sim$ Uniform$(0, \tau)$ and c is a constant on which we put a Gamma$(2, 2)$ prior. We run a MCMC conditional algorithm in order to estimate the actual posterior distribution of both survival functions under the beta-stacy priors just described. Each of the 3 estimators of the survival functions for each group (the hierarchical beta-stacy, the Kaplan-Meier and the beta-stacy estimators) is then compared to the real survival functions by calculating the Wasserstein-1 distances between the real and the estimated curves. Let us recall that the Wasserstein-1 distance, or Wasserstein distance of order 1, between two probability distributions P_1 and P_2 over $\mathbb{R}^+$ is defined as

$$W(P_1, P_2) = \inf_{\gamma\in\Gamma(P_1,P_2)} \int_{\mathbb{R}^+\times\mathbb{R}^+} \mid u - v \mid \mathrm{d}\gamma(u, v),$$

where $\Gamma(P_1, P_2)$ is the set of all joint distributions that have P_1 and P_2 as marginals.

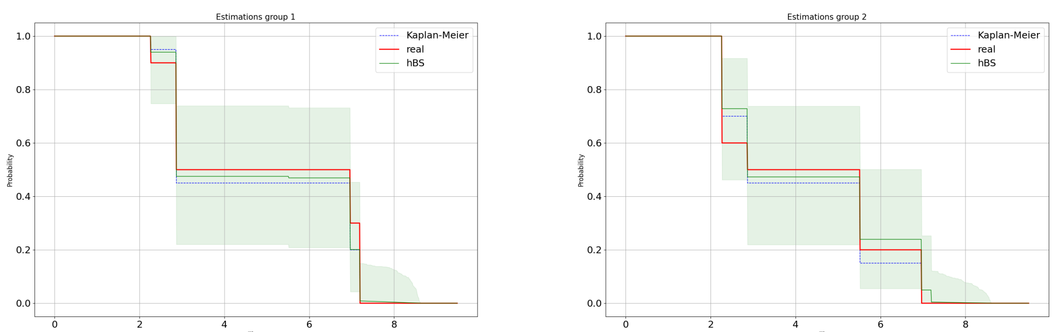
3.6.1 Simulation study: results

The results of the simulation study are summarized in Table 3.2, which shows the Wasserstein distances between the real survival curve and each estimation for each group, in the different examples; in particular, the column **hBS** shows the distance between the real survival curve and the hierarchical beta-stacy estimations, while the columns **KM** and **BS** show the distances between the real curve and the Kaplan-Meier and beta-stacy estimation respectively. Note that the posterior estimator under our prior, as well as under the independent beta-stacy prior, is approximated as an average of the survival functions sampled from the posterior distribution and evaluated on a grid; on the other hand, it is also possible to construct punctual confidence bands by calculating the quantiles of the posterior distributions.

Let us consider the first two experiments, whose datasets are composed by uncensored observations. As the number of observations increases the precision of the estimates (in

Table 3.2: *Wasserstein distances between real and estimated survival curves*

Observations per group	Censorship	Group	hBS	KM	BS
20	No	1	0.0173	0.0272	0.0809
20	No	2	0.0233	0.0279	0.0439
100	No	1	0.0108	0.0088	0.0990
100	No	2	0.0173	0.0139	0.0255
20	Yes	1	0.0368	∞	0.0785
20	Yes	2	0.0434	∞	0.0474
100	Yes	1	0.0290	∞	0.1094
100	Yes	2	0.0410	∞	0.0373



(a) *Group 1* ***(b)*** *Group 2*

Figure 3.1: *Kaplan-Meier estimates (dotted blue lines), hierarchical beta-stacy posterior estimates (green lines) on 20 uncensored observations. The green areas are the 99% confidence intervals of the hBS estimates. The real survival curves are reported as red lines.*

terms of Wasserstein distances) for both groups increases; this is particularly evident for the Kaplan-Meier estimates, since they exclusively relies on the data without any prior effect. Let us focus on the first experiment, characterized by 20 uncensored observations per group. Observing the first two rows of Table 3.2, note that the distances between the real curve and the hBS estimators are lower, for both groups, then the distances between the real curves and the other estimators. This means that with a lower number of observations the effect of the prior and the borrowing of information between groups allows the posterior estimators from a hierarchical beta-stacy prior to perform better than the posterior estimators of two independent beta-stacy priors and the Kaplan-Meier estimators. Figure 3.1 shows the comparison between the real survival curves of the two groups, plotted as red lines, the Kaplan-Meier estimates, plotted as dotted blue lines, and the posterior estimates from the hierarchical beta-stacy model, plotted as green lines. The green areas represent the 99% confidence intervals of the posterior estimates from our model; note that the confidence bands include the real curves. The jumps of the discrete components of our posterior estimations for each group are shared, since they are located on the distinct and exact observations from all the available groups, as described in (3.6). Therefore, even if one of the survival times is not observed in one of the group, this information is borrowed through the estimation of both groups. See for example the last jump of the estimated survival function of the second group in Figure 3.1: as reported in Table 3.1, the simulations were set so that the last (5th) observation appears in the second group with probability equal to zero; on the other hand, it is present in the dataset of the first group, so this information is carried forward in the estimation of both groups.
Moreover, the effect of the borrowing of information can be noted looking at Figure 3.2, which shows the comparison between the estimated survival functions from our model (plotted as green lines) and the estimated survival functions from two independent beta-stacy priors (plotted as red lines). Note how the two survival curves estimated from a hBS model are closer between each other then the two curves estimated from two independent beta-stacy priors; this is numerically confirmed by the first row of Table 3.3, that shows the Wasserstein distances between the survival curves estimated by our model (column **hBS**) and by the two independent beta-stacy models (column **BS**), in the different scenarios. The dependence of the two estimates, modeled by the base measure of the hierarchical beta-stacy prior, keeps them closer in every scenario than the two independent estimates.

Let us now consider the last two experiments, where in each group the last survival times are censored. Again, as the number of observations increases the precision of the estimates for both groups increases. The effect of the prior distributions allows the estimated curves from the posterior distribution in our model to converge to zero as the

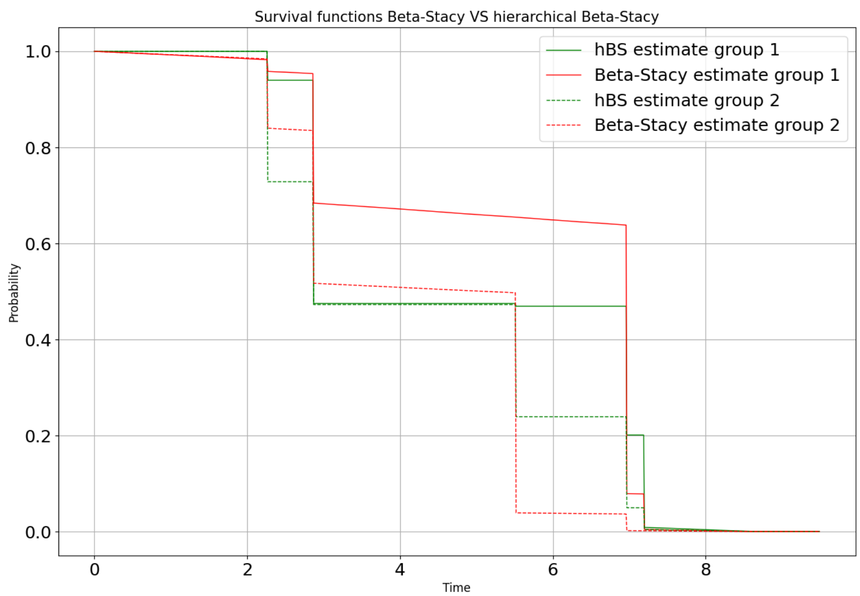
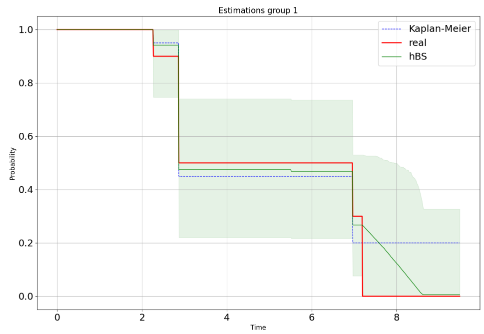


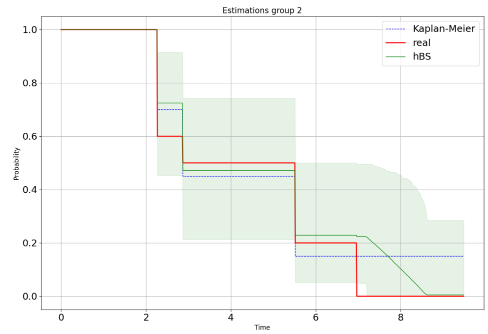
Figure 3.2: *Hierarchical beta-stacy posterior estimates (green lines) and beta-stacy posterior estimates (red lines) for the two groups on 20 uncensored observations. The continuous lines and the dotted lines represent the estimates of the survival functions from the first and the second group respectively.*

Table 3.3: *Wasserstein distances between estimated survival functions*

Observations per group	Censorship	hBS	BS
20	No	0.0531	0.1479
100	No	0.0768	0.1584
20	Yes	0.0555	0.1421
100	Yes	0.0728	0.1569



(a) *Group 1*



(b) *Group 2*

Figure 3.3: *Kaplan-Meier estimates (dotted blue lines), hierarchical beta-stacy posterior estimates (green lines) on 20 observations with censorship. The green areas are the 99% confidence intervals of the hBS estimates. The real survival curves are reported as red lines. Note how the Kaplan-Meier estimates for both groups do not converge to zero.*

time increases, as well as the estimated curves from two independent beta-stacy priors. Note that, on the other hand, the censorship of the last observations induce non-proper independent Kaplan-Meier estimators, as the Wasserstein distances between the real and the estimated survival curves diverges in these experiments. For details, see again Table 3.2.

Let us focus on the third experiment, where the simulated dataset is composed by groups of 20 observations with censorship. Figure 3.3 shows the comparison between the real survival curves of the two groups, plotted as red lines, the Kaplan-Meier estimates, plotted as dotted blue lines, and the posterior estimates from the hierarchical beta-stacy model, plotted as green lines. The green areas represent again the 99% confidence intervals of the posterior estimates from our model; note that the confidence bands include again the real curves. As discussed in the previous case, our models exhibits borrowing of information between the groups by sharing the locations of the jumps of the survival curves and by reducing the distance between the estimated survival curves of the two groups, compared to two independent models. Figure 3.4 provides again a graphical comparison between the estimated survival functions from our model (plotted as green lines) and the estimated survival functions from two independent beta-stacy priors (plotted as red lines), allowing to note how the two green curves are closer between each other then the two red curves; for the numerical comparison, see again Table 3.3 (third row). Note also how the effect of the prior distributions in the independent beta-stacy priors allows them to properly estimate the survival curves (i.e., the estimated survival curves converge to zero), but again the

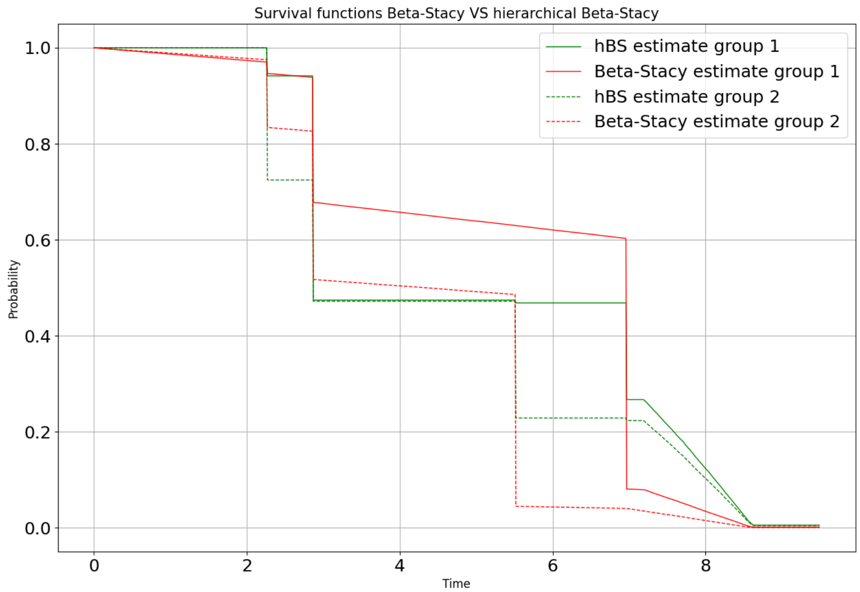


Figure 3.4: *Hierarchical beta-stacy posterior estimates (green lines) and beta-stacy posterior estimates (red lines) for the two groups on 20 observations with censorship. The continuous lines and the dotted lines represent the estimates of the survival functions from the first and the second group respectively.*

performances in terms of Wasserstein distances between the real curve and the posterior estimations shows a worse performance with respect to our model.

3.6.2 Application: leukemia dataset

Following Lijoi and Nipoti (2014), let us now consider two well-known datasets composed by two groups of leukemia remission times (in weeks). The first dataset is taken from Cox (1972) and compares the effect of a treatment versus a placebo observing two groups of 21 and 20 survival times respectively. In particular, the observations from the the treated group are

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

while the observations from the placebo group are

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

The second dataset is taken from Lawless (2003) (Example 7.1.1) and compares the effect of two different treatments (let us call A and B the two treatments), considering two groups of 20 survival times each. In particular, the observations from group A are

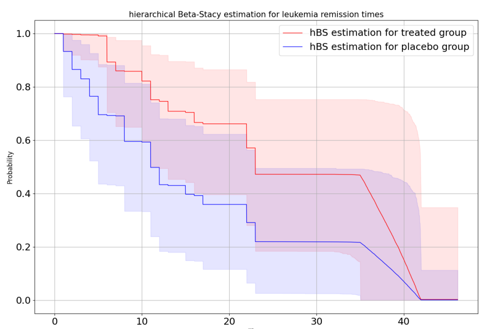
$$1, 3, 3, 6, 7, 7, 10, 12, 14, 15, 18, 19, 22, 26, 28^*, 29, 34, 40, 48^*, 49^*,$$

while the observations from group B are

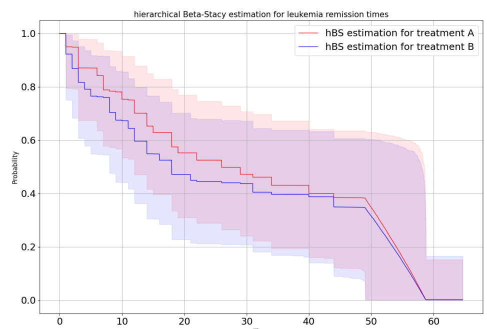
$$1, 1, 2, 2, 3, 4, 5, 8, 8, 9, 11, 12, 14, 16, 18, 21, 27^*, 31, 38^*, 44.$$

In both datasets, the asterisk over the survival time indicates the presence of a censorship. Note how both the examples are discussed also in Damien and Walker (2002).
In each dataset, the estimation of the survival functions for the two groups assuming the hierarchical beta-stacy prior keep the same specifications of the simulation study. In particular, the assumptions on the parameters of our model are the same described in Section 3.5; moreover, we select again as the left bound of the Uniform base measure $\alpha \sim \text{Uniform}(0, \tau)$ the value τ equal to 1.2 times the maximum value observed between the two groups, and we set two Gamma$(2, 2)$ priors on the parameters c and c_*. Finally, we run 20.000 steps of the MCMC algorithm described in Section 3.5.2, where the number of iterations is $R = 10.000$, after further $R_{burnin} = 10.000$ steps of burn-in.
Figure 3.5a shows the estimated survival curves for treatment (red curve) and placebo (blue curve) from the first dataset, along with the 99% confidence intervals for the estimations. Figure 3.5b shows the same results referred tor treatment A (red curve) and treatment B (blue curve) from the second dataset.

As in the simulation study, the locations of the jumps are shared between the two groups; in the second dataset, for example, the exact survival time $t = 44$ appears in group B but not in group A. Nevertheless, Figure 3.5b shows how both the estimated survival curves have a jump at $t = 44$, which is the last available observation. After that, as time increases, the data carry progressively less weight in the estimation of the survival functions, while the prior gains more influence; it can indeed be observed that the survival functions decreases toward zero starting from the last available observation onward. Analyzing the survival curves estimated by the hierarchical beta-stacy model and the respective confidence bounds, it is possible to note that while in the first example the survival curve of the treated patients differs, although not significantly, from the survival curve of the placebo patients, in the second example the survival curves of the patients in the two groups are poorly distinct; this would confirm the conclusions presented in Damien and Walker (2002) as well as in Lijoi and Nipoti (2014).



(a) *Dataset 1 (treatment versus placebo)*



(b) *Dataset 2 (treatment A versus treatment B)*

Figure 3.5: *Estimated survival curves of treatment versus placebo (a) and treatments A versus treatment B (b). Red curve and blue curve represents the estimated survival survival curves for treated group and placebo group (a) and for group A and B (b), while the red and blue areas represent the 99% confidence area for red and blue curve respectively.*

Bibliography

Camerlenghi, F., Lijoi, A., Orbanz, P., and Prünster, I. (2019). Distribution theory for hierarchical processes. *The Annals of Statistics*, 47(1):67–92.

Camerlenghi, F., Lijoi, A., and Prünster, I. (2021). Survival analysis via hierarchically dependent mixture hazards. *The Annals of Statistics*, 49(2):863–884.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2):187–220.

Daley, D. J. and Vere-Jones, D. (2008). *An introduction to the theory of point processes.* Probability and Its Applications, a Series of the Applied Probability Trust. New York: Springer, 2nd edition.

Damien, P. and Walker, S. (2002). A bayesian non-parametric comparison of two treatments. *Scandinavian Journal of Statistics*, 29(1):51–56.

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

Dykstra, R. L. and Laud, P. (1981). A bayesian nonparametric approach to reliability. *The Annals of Statistics*, 9(2):356–367.

Epifani, I. and Lijoi, A. (2010). Nonparametric priors for vectors of survival functions. *Statistica Sinica*, 20(4):1455–1484.

Ferguson, T. S. and Phadia, E. G. (1979). Bayesian nonparametric estimation based on censored data. *The Annals of Statistics*, 7(1):163–186.

Hjort, N. L. (1990). Nonparametric Bayes estimators based on beta processes in models for life history data. *The Annals of Statistics*, 18(3):1259–1294.

Ishwaran, H. and James, L. F. (2004). Computational methods for multiplicative intensity models using weighted Gamma processes. *Journal of the American Statistical Association*, 99(465):175–190.

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

Lawless, J. F. (2003). *Statistical models and methods for lifetime data.* Hoboken, N.J.: Wiley-Interscience.

Lijoi, A. and Nipoti, B. (2014). A class of hazard rate mixtures for combining survival data from different experiments. *Journal of the American Statistical Association*, 109(506):802–814.

Lijoi, A. and Prünster, I. (2010). Models beyond the Dirichlet process. In *Bayesian nonparametrics*, volume 28 of *Camb. Ser. Stat. Probab. Math.*, pages 80–136. Cambridge Univ. Press, Cambridge.

Lo, A. Y. and Chung-Sing, W. (1989). On a class of bayesian nonparametric estimates: Ii. hazard rate estimates. *Annals of the Institute of Statistical Mathematics*, 41(2):227–245.

Masoero, L., Camerlenghi, F., Favaro, S., and Broderick, T. (2018). Posterior representations of hierarchical completely random measures in trait allocation models. *BNP@NeurIPS*.

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.

Teh, Y. W., Jordan, M. I., Beal, M. J., and Blei, D. M. (2006). Hierarchical Dirichlet processes. *Journal of the American Statistical Association*, 101(476):1566–1581.

Walker, S. and Muliere, P. (1997). Beta-Stacy processes and a generalization of the Pólya-urn scheme. *The Annals of Statistics*, 25(4):1762–1780.

Wolpert, R. L. and Ickstadt, K. (1998). Simulation of lévy random fields. In *Practical Nonparametric and Semiparametric Bayesian Statistics*, volume 133 of *Lecture Notes in Statistics*, pages 227–242. Springer, New York, NY.

Appendix

This appendix is organized in different sections. In particular, Sections 3.A and 3.B are dedicated to the proofs of the results reported in Sections 3.3 and 3.4 respectively. Section 3.C is dedicated to the computation of the full conditional distributions useful for the algorithms described in Section 3.5. Finally, Section 3.D provides further details on the conditional algorithm discussed in Section 3.5.2.

3.A Proofs of Section 3.3

Proof of Theorem 3.3.1. In order to prove the theorem, we first consider a finite partition of $\mathbb{R}^+$ and a discretized version of the observations on this grid. Then we find an expression for the corresponding joint distribution of the data, and finally we conclude by taking the limit as the partition size goes to infinity. So, for any $m \geq 1$ we define a partition $\mathscr{P}_m$ of $\mathbb{R}^+$ as follows:

$$\mathscr{P}_m := \{A_{m,i} : \ i = 1, \ldots, k_m + 1\},$$

where $A_{m,i} = (t_{m,i-1}, t_{m,i}]$ for $i = 1, \ldots, k_m$ and $A_{m,k_m+1} = (t_{m,k_m}, \infty)$ with $0 = t_{m,0} < t_{m,1} < \ldots < t_{m,k_m}$, and

$$\lim_{m\to\infty} \left[\max_{1\leq i\leq k_m+1} \mathrm{diam}(A_{m,i}) \right] = 0, \quad \lim_{m\to\infty} t_{m,k_m} = \infty,$$

where

$$\mathrm{diam}(A_{m,i}) = \sup\{| x - y | : x, y \in A_{m,i}\}.$$

The discretized version of the observations is defined as follows:

$$\xi_{m,i,j} = \sum_{\ell=1}^{k_m} t_{m,\ell} \mathbb{1}_{A_{m,\ell}}(X_{i,j}) + t_{m,k_m+1} \mathbb{1}_{A_{m,k_m+1}}(X_{i,j}), \tag{3.15}$$

and t_{m,k_m+1} is a whichever point of the interval $A_{m,k_m+1} \in \mathscr{P}_m$. Let us define the set $\boldsymbol{D}_m := ((\xi_{m,i,j}, \Delta_{i,j}) : \ i = 1, \ldots, N_j, \, j = 1, \ldots, d)$, i.e., the vector containing all the discretized data points. We further suppose that the distinct observations are ordered, i.e., $X_1^* < X_2^* < \ldots < X_K^*$, and let us denote by A_{m,i_r} the set of the partition $\mathscr{P}_m$ containing the rth distinct value. The goal is to evaluate the joint distribution of the data $\boldsymbol{D}_m$, i.e.,

$$p^{(m)}(\boldsymbol{D}_m) := \mathbb{P}\left[\bigcap_{j=1}^{d}(\xi_{m,1,j},\ldots,\xi_{m,N_j,j}) \in \times_{r=1}^{K} A_{m,i_r}^{n_{r,j}+n_{r,j}^c}\right]. \tag{3.16}$$

Note that (3.16) is the evaluation of the probability that the vector $(\xi_{m,1,j},\ldots,\xi_{m,N_j,j})$ contains $n_{r,j}$ exact observations and $n_{r,j}^c$ censored observations coinciding with X_r^*, as $r = 1,\ldots,K$ and $j = 1,\ldots,d$. Moreover, thanks to the partial exchangeability, (3.16) can be written as

$$\begin{aligned}
p^{(m)}(\boldsymbol{D}_m) &= \mathbb{E}\left[\prod_{j=1}^{d}\prod_{r=1}^{K}\left(\tilde{p}_j^{n_{r,j}}(A_{m,i_r})(1-\tilde{p}_j(0,t_{m,i_r}])^{n_{r,j}^c}\right)\right] \\
&= \mathbb{E}\left[\prod_{j=1}^{d}\prod_{r=1}^{K}(\tilde{F}_j(t_{m,i_r})-\tilde{F}_j(t_{m,i_r-1}))^{n_{r,j}}(1-\tilde{F}_j(t_{m,i_r}))^{n_{r,j}^c}\right] \\
&= \mathbb{E}\left[\prod_{j=1}^{d}\prod_{r=1}^{K}(e^{-\tilde{\mu}_j(0,t_{m,i_r-1}]}-e^{-\tilde{\mu}_j(0,t_{m,i_r}]})^{n_{r,j}}e^{-n_{r,j}^c\tilde{\mu}_j(0,t_{m,i_r}]}\right] \\
&= \mathbb{E}\left[\prod_{j=1}^{d}\prod_{r=1}^{K}(1-e^{-\tilde{\mu}_j(t_{m,i_r-1},t_{m,i_r}]})^{n_{r,j}}e^{-n_{r,j}\tilde{\mu}_j(0,t_{m,i_r-1}]}e^{-n_{r,j}^c\tilde{\mu}_j(0,t_{m,i_r}]}\right],
\end{aligned}$$

which can also be written as

$$\begin{aligned}
p^{(m)}(\boldsymbol{D}_m) = \mathbb{E}\Big[\prod_{j=1}^{d}\exp\Big\{-\sum_{r=1}^{K}n_{r,j}\tilde{\mu}_j(0,t_{m,i_r-1}]-\sum_{r=1}^{K}n_{r,j}^c\tilde{\mu}_j(0,t_{m,i_r}]\Big\} \\
\times\prod_{j=1}^{d}\prod_{r=1}^{K}(1-e^{-\tilde{\mu}_j(t_{m,i_r-1},t_{m,i_r}]})^{n_{r,j}}\Big].
\end{aligned} \tag{3.17}$$

The first sum over r in (3.17) can be written as

$$\begin{aligned}
\sum_{r=1}^{K}n_{r,j}\tilde{\mu}_j(0,t_{m,i_r-1}] &= \sum_{r=1}^{K}n_{r,j}\sum_{i=1}^{i_r-1}\tilde{\mu}_j(t_{m,i-1},t_{m,i}] \\
&= \sum_{i=1}^{i_K-1}\tilde{\mu}_j(t_{m,i-1},t_{m,i}]\sum_{r:\,i_r\geq i+1}n_{r,j},
\end{aligned}$$

where the last equation is obtained exchanging the two sums. Let us further introduce the at risk processes for the discretized observations, which will be denoted by $\bar{N}_{m,j}$ and $\tilde{N}^c_{m,j}$; therefore, note that

$$\sum_{r=1}^{K} n_{r,j}\tilde{\mu}_j(0, t_{m,i_r-1}] = \sum_{i=1}^{k_m} \bar{N}_{m,j}(t^+_{m,i})\tilde{\mu}_j(t_{m,i-1}, t_{m,i}], \tag{3.18}$$

where $\bar{N}_{m,j}(t^+_{m,i})$ is the number of exact observations which are at risk after time $t_{m,i}$, excluding time $t_{m,i}$, for population j. Observe that the sum has been extended until k_m since $\bar{N}_{m,j}(t^+_{m,i}) = 0$ if $i \geq i_K$.
Analogously the second sum over r in (3.17) can be written as

$$\sum_{r=1}^{K} n^c_{r,j}\tilde{\mu}_j(0, t_{m,i_r}] = \sum_{i=1}^{k_m} \tilde{N}^c_{m,j}(t_{m,i})\tilde{\mu}_j(t_{m,i-1}, t_{m,i}]. \tag{3.19}$$

Exploiting (3.18) and (3.19), (3.17) can be written as

$$\begin{aligned} p^{(m)}(\boldsymbol{D}_m) = \mathbb{E}\Big[&\prod_{j=1}^{d}\prod_{r=1}^{K}(1 - e^{-\tilde{\mu}_j(t_{m,i_r-1}, t_{m,i_r}]})^{n_{r,j}} \\ &\times \prod_{j=1}^{d} \exp\Big\{ -\sum_{i=1}^{k_m} \bar{N}_{m,j}(t^+_{m,i})\tilde{\mu}_j(t_{m,i-1}, t_{m,i}] - \sum_{i=1}^{k_m} \tilde{N}^c_{m,j}(t_{m,i})\tilde{\mu}_j(t_{m,i-1}, t_{m,i}]\Big\}\Big]. \end{aligned} \tag{3.20}$$

Note that defining the set

$$\mathbb{X}^* := \mathbb{X} \setminus \bigcup_{r=1\ \Delta^*_r=1}^{K} A_{m,i_r},$$

(3.20) can be written as

$$\begin{aligned} p^{(m)}(\boldsymbol{D}_m) = \mathbb{E}\Big[&\prod_{j=1}^{d} \prod_{r=1\ \Delta^*_r=1}^{K} (1 - e^{-\tilde{\mu}_j(t_{m,i_r-1}, t_{m,i_r}]})^{n_{r,j}} e^{-\tilde{\mu}_j(t_{m,i_r-1}, t_{m,i_r}](\bar{N}_{m,j}(t^+_{m,i_r}) + \tilde{N}^c_{m,j}(t_{m,i_r}))} \\ &\times \prod_{j=1}^{d} \exp\Big\{ -\int_{\mathbb{X}^*} (\bar{N}_{m,j}(x) + \tilde{N}^c_{m,j}(x))\mathrm{d}\tilde{\mu}_j(x)\Big\}\Big]. \end{aligned} \tag{3.21}$$

The expected value in (3.21) can be evaluated conditioning on $\tilde{\mu}_0$ and using the independence to get

$$
\begin{aligned}
p^{(m)}(\boldsymbol{D}_m) = \mathbb{E}\Big[\prod_{j=1}^{d} \prod_{r=1}^{K}\prod_{\Delta_r^*=1} & \mathbb{E}\Big((1-e^{-\tilde{\mu}_j(t_{m,i_r-1},t_{m,i_r}]})^{n_{r,j}} \\
& \times \exp\Big\{-\tilde{\mu}_j(t_{m,i_r-1},t_{m,i_r}](\bar{N}_{m,j}(t_{m,i_r}^+)+\tilde{N}_{m,j}^c(t_{m,i_r}))\Big\} \mid \tilde{\mu}_0\Big) \\
& \times \prod_{j=1}^{d}\mathbb{E}\Big(\exp\Big\{-\int_{\mathbb{X}^*}(\bar{N}_{m,j}(x)+\tilde{N}_{m,j}^c(x))\mathrm{d}\tilde{\mu}_j(x)\Big\} \mid \tilde{\mu}_0\Big)\Big].
\end{aligned} \tag{3.22}
$$

For notational convenience let us set $\eta_{r,j} := (\bar{N}_{m,j}(t_{m,i_r}^+) + \tilde{N}_{m,j}^c(t_{m,i_r}))$, so that for any r and j the first conditional expected value in (3.22) may be evaluated as follows:

$$
\begin{aligned}
\mathbb{E}&\left[(1-e^{-\tilde{\mu}_j(A_{m,i_r})})^{n_{r,j}}e^{-\eta_{r,j}\tilde{\mu}_j(A_{m,i_r})} \mid \tilde{\mu}_0\right] = \\
&= \sum_{v=0}^{n_{r,j}}\binom{n_{r,j}}{v}(-1)^v\mathbb{E}\Big(e^{-(v+\eta_{r,j})\tilde{\mu}_j(A_{m,i_r})} \mid \tilde{\mu}_0\Big) \\
&= \sum_{v=0}^{n_{r,j}}\binom{n_{r,j}}{v}(-1)^v\mathbb{E}\left[\prod_{\ell\geq 1}e^{-(v+\eta_{r,j})\tilde{h}_{j,\ell}\delta_{\tilde{x}_\ell}(A_{m,i_r})} \mid \tilde{\mu}_0\right] \\
&= \sum_{v=0}^{n_{r,j}}\binom{n_{r,j}}{v}(-1)^v\prod_{\ell\geq 1}\mathbb{E}\Big[e^{-(v+\eta_{r,j})\tilde{h}_{j,\ell}\delta_{\tilde{x}_\ell}(A_{m,i_r})} \mid \tilde{\mu}_0\Big] \\
&= \sum_{v=0}^{n_{r,j}}\binom{n_{r,j}}{v}(-1)^v\prod_{\ell\geq 1}\int_0^{\infty}e^{-(v+\eta_{r,j})h\delta_{\tilde{x}_\ell}(A_{m,i_r})}f_j(h \mid \tilde{h}_{0,\ell},\tilde{x}_\ell,b_j)\mathrm{d}h.
\end{aligned}
$$

Note that the first equality is justified by Newton formula, which holds true also when $n_{r,j}=0$, indeed in such a situation the only term of the sum is the one for $v=0$.
The second conditional expectation in (3.22), by setting $N_{m,j}(x) := \bar{N}_{m,j}(x) + \tilde{N}_{m,j}^c(x)$, can be evaluated as follows:

$$\mathbb{E}\left[\exp\left\{-\int_{\mathbb{X}^*} N_{m,j}(x)\mathrm{d}\tilde{\mu}_j(x)\right\} \mid \tilde{\mu}_0\right] =$$
$$= \mathbb{E}\left[\exp\left\{-\int_{\mathbb{X}^*} N_{m,j}(x)\sum_{\ell=1}^{\infty}\tilde{h}_{j,\ell}\delta_{\tilde{x}_\ell}(\mathrm{d}x)\right\} \mid \tilde{\mu}_0\right]$$
$$= \prod_{\ell\geq 1}\int_0^\infty e^{-hN_{m,j}(\tilde{x}_\ell)\delta_{\tilde{x}_\ell}(\mathbb{X}^*)} f_j(h \mid \tilde{h}_{0,\ell}, \tilde{x}_\ell, b_j)\mathrm{d}h.$$

Substituting the previous expressions in (3.22) it can be written as

$$p^{(m)}(\boldsymbol{D}_m) =$$
$$\mathbb{E}\Big[\prod_{j=1}^{d}\prod_{r=1}^{K}\prod_{\Delta_r^*=1}\sum_{v=0}^{n_{r,j}}\binom{n_{r,j}}{v}(-1)^v\prod_{\ell\geq 1}\int_0^\infty e^{-(v+\eta_{r,j})h\delta_{\tilde{x}_\ell}(A_{m,i_r})} f_j(h \mid \tilde{h}_{0,\ell}, \tilde{x}_\ell, b_j)\mathrm{d}h$$
$$\times \prod_{j=1}^{d}\prod_{\ell\geq 1}\int_0^\infty e^{-hN_{m,j}(\tilde{x}_\ell)\delta_{\tilde{x}_\ell}(\mathbb{X}^*)} f_j(h \mid \tilde{h}_{0,\ell}, \tilde{x}_\ell, b_j)\mathrm{d}h\Big],$$

where the expectation is made with respect to $\tilde{\mu}_0$. Exploiting the independence properties of the CRM $\tilde{\mu}_0$, the previous expression can be written as

$$p^{(m)}(\boldsymbol{D}_m) =$$
$$\prod_{r=1}^{K}\prod_{\Delta_r^*=1}\mathbb{E}\left[\prod_{j=1}^{d}\sum_{v=0}^{n_{r,j}}\binom{n_{r,j}}{v}(-1)^v\prod_{\ell\geq 1}\int_0^\infty e^{-(v+\eta_{r,j})h\delta_{\tilde{x}_\ell}(A_{m,i_r})} f_j(h \mid \tilde{h}_{0,\ell}, \tilde{x}_\ell, b_j)\mathrm{d}h\right] \tag{3.23}$$
$$\times\, \mathbb{E}\left[\prod_{\ell\geq 1}\prod_{j=1}^{d}\int_0^\infty e^{-hN_{m,j}(\tilde{x}_\ell)\delta_{\tilde{x}_\ell}(\mathbb{X}^*)} f_j(h \mid \tilde{h}_{0,\ell}, \tilde{x}_\ell, b_j)\mathrm{d}h\right].$$

The second expected value in (3.23) can be computed as follows:

$$
\begin{aligned}
&\mathbb{E}\left[\prod_{\ell\geq 1}\prod_{j=1}^{d}\int_0^{\infty} e^{-hN_{m,j}(\tilde{x}_\ell)\delta_{\tilde{x}_\ell}(\mathbb{X}^*)} f_j(h \mid \tilde{h}_{0,\ell},\tilde{x}_\ell,b_j)\mathrm{d}h\right] = \\
&= \mathbb{E}\left[\exp\left\{\sum_{\ell\geq 1}\log\Big(\prod_{j=1}^{d}\int_0^{\infty} e^{-hN_{m,j}(\tilde{x}_\ell)\delta_{\tilde{x}_\ell}(\mathbb{X}^*)} f_j(h \mid \tilde{h}_{0,\ell},\tilde{x}_\ell,b_j)\mathrm{d}h\Big)\right\}\right] \qquad (3.24)\\
&= \exp\left\{-\int_{\mathbb{X}^*}\int_0^{\infty}\Big(1-\prod_{j=1}^{d}\int_0^{\infty} e^{-hN_{m,j}(x)} f_j(h \mid s,x,b_j)\mathrm{d}h\Big)\rho(s\mid x)\mathrm{d}s\alpha(\mathrm{d}x)\right\}.
\end{aligned}
$$

Note that the last equality exploits the availability of the Laplace functional of the CRM $\tilde{\mu}_0$.

As for the first expected value appearing in (3.23), note that

$$
\begin{aligned}
&\mathbb{E}\left[\prod_{j=1}^{d}\sum_{v=0}^{n_{r,j}}\binom{n_{r,j}}{v}(-1)^v\prod_{\ell\geq 1}\int_0^{\infty} e^{-(v+\eta_{r,j})h\delta_{\tilde{x}_\ell}(A_{m,i_r})} f_j(h \mid \tilde{h}_{0,\ell},\tilde{x}_\ell,b_j)\mathrm{d}h\right] = \\
&= \sum_{v_{r,j}:\, j=1,\dots,d}\prod_{j=1}^{d}\binom{n_{r,j}}{v_{r,j}}(-1)^{v_{r,j}}\mathbb{E}\left[\prod_{j=1}^{d}\prod_{\ell\geq 1}\int_0^{\infty} e^{-(v_{r,j}+\eta_{r,j})h\delta_{\tilde{x}_\ell}(A_{m,i_r})} f_j(h \mid \tilde{h}_{0,\ell},\tilde{x}_\ell,b_j)\mathrm{d}h\right].
\end{aligned}
$$

The expected value appearing in the previous expression may be evaluated as follows:

$$
\begin{aligned}
&\mathbb{E}\left[\prod_{j=1}^{d}\sum_{v=0}^{n_{r,j}}\binom{n_{r,j}}{v}(-1)^v\prod_{\ell\geq 1}\int_0^{\infty} e^{-(v+\eta_{r,j})h\delta_{\tilde{x}_\ell}(A_{m,i_r})} f_j(h \mid \tilde{h}_{0,\ell},\tilde{x}_\ell,b_j)\mathrm{d}h\right] = \\
&= \sum_{v_{r,j}:\, j=1,\dots,d}\prod_{j=1}^{d}\binom{n_{r,j}}{v_{r,j}}(-1)^{v_{r,j}} \\
&\qquad \times \exp\left\{-\int_{A_{m,i_r}}\int_0^{\infty}\Big(1-\prod_{j=1}^{d}\int_0^{\infty} e^{-h(v_{r,j}+\eta_{r,j})} f_j(h \mid s,x,b_j)\mathrm{d}h\Big)\rho(s\mid x)\mathrm{d}s\alpha(\mathrm{d}x)\right\}.
\end{aligned}
$$

Since the diameter of A_{m,i_r} goes to zero as $m\to\infty$, the argument of the exponential function goes to zero as $m\to\infty$, hence using an expansion of the exponential the previous expression is equal to

$$\mathbb{E}\left[\prod_{j=1}^{d}\sum_{v=0}^{n_{r,j}}\binom{n_{r,j}}{v}(-1)^v\prod_{\ell\geq 1}\int_0^\infty e^{-(v+\eta_{r,j})h\delta_{\tilde{x}_\ell}(A_{m,i_r})}f_j(h\mid \tilde{h}_{0,\ell},\tilde{x}_\ell,b_j)\mathrm{d}h\right]=$$
$$=\sum_{v_{r,j}\colon j=1,\dots,d}\prod_{j=1}^{d}\binom{n_{r,j}}{v_{r,j}}(-1)^{v_{r,j}}\Big\{1-\int_{A_{m,i_r}}\int_0^\infty$$
$$\Big(1-\prod_{j=1}^{d}\int_0^\infty e^{-h(v_{r,j}+\eta_{r,j})}f_j(h\mid s,x,b_j)\mathrm{d}h\Big)\rho(s\mid x)\mathrm{d}s\alpha(\mathrm{d}x)+o\Big(P_0(A_{m,i_r})\Big)\Big\}.$$

Observe that if $n_{r,j}=0$ there is no sum over $v_{r,j}$ but the expression can be obtain simply substituting $v_{r,j}=0$ in it, while if $n_{r,j}>0$ the sum over the corresponding $v_{r,j}$ is proper and since

$$\sum_{v=0}^{n}\binom{n}{v}(-1)^v=(-1+1)^n=0\quad\text{if } n>0,$$

therefore

$$\mathbb{E}\left[\prod_{j=1}^{d}\sum_{v=0}^{n_{r,j}}\binom{n_{r,j}}{v}(-1)^v\prod_{\ell\geq 1}\int_0^\infty e^{-(v+\eta_{r,j})h\delta_{\tilde{x}_\ell}(A_{m,i_r})}f_j(h\mid \tilde{h}_{0,\ell},\tilde{x}_\ell,b_j)\mathrm{d}h\right]$$
$$=\int_{A_{m,i_r}}\int_0^\infty\Big(\prod_{j=1}^{d}\int_0^\infty e^{-h\eta_{r,j}}(1-e^{-h})^{n_{r,j}}f_j(h\mid s,x,b_j)\mathrm{d}h\Big)\rho(s\mid x)\mathrm{d}s\alpha(\mathrm{d}x)$$
$$+o\Big(P_0(A_{m,i_r})\Big).$$

Recalling that $A_{m,i_r}\downarrow X_r^*$ as $m\to\infty$, thanks to the Lebesgue–Besicovitch derivation theorem the expected value can be written as

$$\mathbb{E}\left[\prod_{j=1}^{d}\sum_{v=0}^{n_{r,j}}\binom{n_{r,j}}{v}(-1)^v\prod_{\ell\geq 1}\int_0^\infty e^{-(v+\eta_{r,j})h\delta_{\tilde{x}_\ell}(A_{m,i_r})}f_j(h\mid \tilde{h}_{0,\ell},\tilde{x}_\ell,b_j)\mathrm{d}h\right]$$
$$=\alpha(A_{m,i_r})\int_0^\infty\Big(\prod_{j=1}^{d}\int_0^\infty e^{-h\eta_{r,j}}(1-e^{-h})^{n_{r,j}}f_j(h\mid s,x,b_j)\mathrm{d}h\Big)\rho(s\mid X_r^*)\mathrm{d}s \tag{3.25}$$
$$+o\Big(P_0(A_{m,i_r})\Big).$$

Exploiting (3.24) and (3.25), expression (3.23) reduces to the following:

$$
\begin{aligned}
&p^{(m)}(\boldsymbol{D}_m) = \\
&\prod_{r=1}^{K}\prod_{\Delta_r^*=1}\Bigg\{\alpha(A_{m,i_r})\int_0^\infty\prod_{j=1}^{d} \\
&\qquad\int_0^\infty e^{-h\eta_{r,j}}(1-e^{-h})^{n_{r,j}}f_j(h\mid s,X_r^*,b_j)\mathrm{d}h\rho(s\mid X_r^*)\mathrm{d}s\Bigg\} \\
&\times\exp\Bigg\{-\int_{\mathbb{X}^*}\int_0^\infty\Big[1-\prod_{j=1}^{d}\int_0^\infty e^{-hN_{m,j}(x)}f_j(h\mid s,x,b_j)\mathrm{d}h\Big]\rho(s\mid x)\mathrm{d}s\alpha(\mathrm{d}x)\Bigg\} \\
&\qquad+o\Big(\prod_{r=1}^{K}\prod_{\Delta_r^*=1}P_0(A_{m,i_r})\Big).
\end{aligned}
\tag{3.26}
$$

The thesis follows observing that $\bar{N}_{m,j}\to\bar{N}_j(x)$ and $\tilde{N}^c_{m,j}(x)\to\tilde{N}^c_j(x)$ for any x as $m\to\infty$. □

In order to prove Theorem 3.3.2, let us state and prove the following proposition.

Proposition 3.A.1. *Let $(\tilde{\mu}_1,\ldots,\tilde{\mu}_d)$ be a vector of hierarchical CRMs. Then the joint Laplace functional a priori is equal to*

$$
\begin{aligned}
&L_{(\tilde{\mu}_1,\ldots,\tilde{\mu}_d)}(g_1,\ldots,g_d)=\mathbb{E}[e^{-\tilde{\mu}_1(g_1)-\ldots-\tilde{\mu}_d(g_d)}]= \\
&=\exp\left\{-\int_{\mathbb{X}}\int_0^\infty\left(1-\prod_{j=1}^{d}\int_0^\infty e^{-hg_j(x)}f_j(h\mid r,x,b_j)\mathrm{d}h\right)\rho(r\mid x)\mathrm{d}r\alpha(\mathrm{d}x)\right\},
\end{aligned}
\tag{3.27}
$$

for any measurable $g_j:\mathbb{X}\to\mathbb{R}^+$ such that $\tilde{\mu}_j(g_j):=\int_{\mathbb{X}}g_j(x)\tilde{\mu}_j(\mathrm{d}x)$, for $j=1,\ldots,d$.

Proof. Note that

$$
\begin{aligned}
L_{(\tilde{\mu}_1,\ldots,\tilde{\mu}_d)}(g_1,\ldots,g_d) &= \mathbb{E}\left[\prod_{j=1}^d \mathbb{E}(e^{-\tilde{\mu}_j(g_j)} \mid \tilde{\mu}_0)\right] = \mathbb{E}\left[\prod_{j=1}^d \prod_{k\geq 1} \mathbb{E}\left(\exp\left\{-\tilde{h}_{j,k} g_j(\tilde{x}_k)\right\}\right)\right] \\
&= \mathbb{E}\left[\prod_{k\geq 1}\prod_{j=1}^d \int_0^\infty e^{-h g_j(\tilde{x}_k)} f_j(h \mid \tilde{h}_{0,k}, \tilde{x}_k, b_j)\mathrm{d}h\right] \\
&= \mathbb{E}\left[\exp\left\{\sum_{k\geq 1}\log\left(\prod_{j=1}^d \int_0^\infty e^{-h g_j(\tilde{x}_k)} f_j(h \mid \tilde{h}_{0,k}, \tilde{x}_k, b_j)\mathrm{d}h\right)\right\}\right].
\end{aligned}
$$

The last expression may be easily evaluated by means of the expression of the Laplace functional of $\tilde{\mu}_0$. Therefore (3.27) follows. □

Proof of Theorem 3.3.2. Let us consider d measurable functions $g_j : \mathbb{X} \to \mathbb{R}^+$, for each $j = 1,\ldots,d$, then to prove the theorem it is sufficient to evaluate the posterior Laplace functional of $(\tilde{\mu}_1,\ldots,\tilde{\mu}_d)$, i.e.,

$$
L_{(\tilde{\mu}_1,\ldots,\tilde{\mu}_d)|\boldsymbol{D},\boldsymbol{s}}(g_1,\ldots,g_d) := \mathbb{E}\left[e^{-\tilde{\mu}_1(g_1)-\ldots-\tilde{\mu}_d(g_d)} \mid \boldsymbol{D},\boldsymbol{s}\right],
$$

where $\boldsymbol{s}$ are the latent jumps introduced in (3.5).
In order to evaluate the Laplace functional, let us recall the partition $\mathscr{P}_m$ introduced in the proof of Theorem 3.3.1. Note that the posterior Laplace functional is given by

$$
\begin{aligned}
&L_{(\tilde{\mu}_1,\ldots,\tilde{\mu}_d)|\boldsymbol{D}}(g_1,\ldots,g_d) = \\
&= \lim_{m\to\infty} \frac{\mathbb{E}\left[e^{-\tilde{\mu}_1(g_1)-\ldots-\tilde{\mu}_d(g_d)}\mathbb{P}\left[\bigcap_{j=1}^d(\xi_{m,1,j},\ldots,\xi_{m,N_j,j}) \in \times_{r=1}^K A_{m,i_r}^{n_{r,j}+n_{r,j}^c} \mid \tilde{\mu}_1,\ldots,\tilde{\mu}_d\right]\right]}{\mathbb{P}\left[\bigcap_{j=1}^d(\xi_{m,1,j},\ldots,\xi_{m,N_j,j}) \in \times_{r=1}^K A_{m,i_r}^{n_{r,j}+n_{r,j}^c}\right]}.
\end{aligned}
$$

Conditionally on $\boldsymbol{s}$, one may easily determine the numerator and the denominator in the previous expression. In particular the denominator, conditionally on $\boldsymbol{s}$, can be derived from (3.5), while the numerator conditionally on $\boldsymbol{s}$ amounts to be

$$\prod_{r=1\ \Delta_r^*=1}^{K} \alpha(\mathrm{d}X_r^*) \prod_{j=1}^{d} \int_0^\infty e^{-h(\bar{n}_{r+1,j}+\tilde{n}_{r,j}^c)-hg_j(X_r^*)}(1-e^{-h})^{n_{r,j}} f_j(h \mid s_r, X_r^*, b_j)\mathrm{d}h\rho(s_r \mid X_r^*)\mathrm{d}s_r$$
$$\times \exp\left\{-\int_{\mathbb{X}}\int_0^\infty \left[1-\prod_{j=1}^{d}\int_0^\infty e^{-h(N_j(x)+g_j(x))} f_j(h \mid s, x, b_j)\mathrm{d}h\right]\rho(s \mid x)\mathrm{d}s\alpha(\mathrm{d}x)\right\}.$$

Therefore dividing the previous expression by the expression of the likelihood derived from (3.5) and taking the limit as $m \to \infty$ it is possible to determine the posterior Laplace functional conditionally on the data and the latent jumps. In particular

$$\begin{aligned} L_{(\tilde{\mu}_1,\ldots,\tilde{\mu}_d)\mid \boldsymbol{D},\boldsymbol{s}}(g_1,\ldots,g_d) &= \\ &= \exp\left\{-\int_{\mathbb{X}}\int_0^\infty \left[1-\prod_{j=1}^{d}\int_0^\infty e^{-hg_j(x)} f_j'(h \mid s, x, b_j)\mathrm{d}h\right]\nu'(\mathrm{d}s,\mathrm{d}x)\right\} \\ &\quad \times \prod_{r=1\ \Delta_r^*=1}^{K}\prod_{j=1}^{d}\int_0^\infty e^{-hg_j(X_r^*)} f_j'(h \mid s_r, X_r^*, b_j)\mathrm{d}h. \end{aligned} \tag{3.28}$$

Observe that the first term in (3.28) is the Laplace functional of the vector of CRMs as described in point *(i)* of the statement, while the last term in (3.28) is the Laplace functional referring to the vector of jumps described in point *(ii)* of the statement. Therefore the thesis follows. □

3.B Proofs of Section 3.4

Proof of Proposition 3.4.1. Note that

$$\mathbb{E}\left[e^{-\tilde{\mu}_j(0,t]}\right] = \mathbb{E}\left[\mathbb{E}\left[e^{-\tilde{\mu}_j(0,t]} \mid \tilde{\mu}_0\right]\right] = \mathbb{E}\left[\prod_{k\geq 1}\mathbb{E}\left[e^{-\tilde{h}_{j,k}\delta_{\tilde{x}_k}(0,t]} \mid \tilde{\mu}_0\right]\right] = \mathbb{E}\left[\prod_{k\geq 1:\tilde{x}_k\leq t}\mathbb{E}\left[e^{-\tilde{h}_{j,k}} \mid \tilde{\mu}_0\right]\right].$$

Note that by definition

$$1-e^{-\tilde{h}_{j,k}} \mid \tilde{\mu}_0 \sim Beta\left(c_j\left(\tilde{x}_k\right)F_j\left(\tilde{h}_{0,k}\right), c_j\left(\tilde{x}_k\right)\left(1-F_j\left(\tilde{h}_{0,k}\right)\right)\right),$$

so it follows that

$$e^{-\tilde{h}_{j,k}} \mid \tilde{\mu}_0 \sim Beta\left(c_j\left(\tilde{x}_k\right)\left(1-F_j\left(\tilde{h}_{0,k}\right)\right), c_j\left(\tilde{x}_k\right)F_j\left(\tilde{h}_{0,k}\right)\right).$$

So for each $k \geq 1$, the value $\mathbb{E}\left[e^{-\tilde{h}_{j,k}\delta_{\tilde{x}_k}} \mid \tilde{\mu}_0\right]$ is the expected value of the above Beta distribution, as a function of the locations $\tilde{x}_k$'s and of the jumps $\tilde{h}_{0,k}$'s; let us denote this expected value as $f_j\left(\tilde{x}_k, \tilde{h}_{0,k}\right)$.
So the we can write the expected value of $e^{-\tilde{\mu}_j(0,t]}$ as

$$\mathbb{E}\left[e^{-\tilde{\mu}_j(0,t]}\right] = \mathbb{E}\left[\prod_{k\geq 1:\tilde{x}_k\leq t} f_j\left(\tilde{x}_k, \tilde{h}_{0,k}\right)\right] = \mathbb{E}\left[\exp\left(\sum_{k\geq 1:\tilde{x}_k\leq t}\log\left(f_j\left(\tilde{x}_k, \tilde{h}_{0,k}\right)\right)\right)\right]$$
$$= \mathbb{E}\left[\exp\left(\int_0^t\int_{\mathbb{R}^+}\log\left(f_j\left(x,h'\right)\right)\tilde{N}\left(dx,dh\right)\right)\right],$$

where $\tilde{N} = \sum_{k\geq 1}\delta_{\left(\tilde{x}_k,\tilde{h}_{0,k}\right)}$ is the marked Poisson process corresponding to the base measure. The last quantity can be calculated, and it is equal to

$$\exp\left[-\int_0^t\int_{\mathbb{R}^+}\left(1-f_j\left(x,h\right)\right)\nu\left(dx,dh\right)\right],$$

where ν is the Lévy intensity of $\tilde{\mu}_0$.
Since

$$f_j\left(x,h\right) = \mathbb{E}\left[e^{-h\delta_x} \mid \tilde{\mu}_0\right] = \frac{c_j\left(x\right)\left(1-F_j\left(h\right)\right)}{c_j\left(x\right)\left(1-F_j\left(h\right)\right)+c_j\left(x\right)F_j\left(h\right)} = 1-F_j\left(h\right),$$

the jth expected value can be written as

$$\mathbb{E}\left[e^{-\tilde{\mu}_j(0,t]}\right] = \exp\left[-\int_0^t\int_{\mathbb{R}^+}\left(1-\left(1-F_j\left(h\right)\right)\right)\nu\left(dx,dh\right)\right] = \exp\left[-\int_0^t\int_{\mathbb{R}^+}F_j\left(h\right)\nu\left(dx,dh\right)\right].$$

Writing explicitly the Lévy intensity $\nu\left(dx,dh\right)$, the thesis of the proposition follows. □

Proof of Proposition 3.4.2. The thesis is true if and only if

$$0 = \lim_{t\to\infty}\tilde{S}_j(t) = \lim_{t\to\infty}e^{-\tilde{\mu}_j(0,t]}$$

almost surely. Since $e^{-\tilde{\mu}_j(0,t]}$ is a positive function, $\lim_{t\to\infty}e^{-\tilde{\mu}_j(0,t]} = 0$ if and only if $\mathbb{E}\left[\lim_{t\to\infty}e^{-\tilde{\mu}_j(0,t]}\right] = 0$, i.e.,

$$\lim_{t\to\infty} \mathbb{E}\left[e^{-\tilde{\mu}_j(0,t]}\right] = 0.$$

Note that

$$\mathbb{E}\left[e^{-\tilde{\mu}_j(0,t]}\right] = \mathbb{E}\left[\mathbb{E}\left[e^{-\tilde{\mu}_j(0,t]} \mid \tilde{\mu}_0\right]\right],$$

and since $\tilde{\mu}_j \mid \tilde{\mu}_0 = \sum_{k\geq 1} \tilde{h}_{j,k}\delta_{\tilde{x}_k}$, then

$$\begin{aligned}
\mathbb{E}\left[e^{-\tilde{\mu}_j(0,t]}\right] &= \mathbb{E}\left[\prod_{k\geq 1}\mathbb{E}\left[e^{-\tilde{h}_{j,k}\delta_{\tilde{x}_k}(0,t]} \mid \tilde{\mu}_0\right]\right] = \mathbb{E}\left[\prod_{k\geq 1}\mathbb{E}\left[\delta_{\tilde{x}_k}(t,\infty) + e^{-\tilde{h}_{j,k}}\delta_{\tilde{x}_k}(0,t] \mid \tilde{\mu}_0\right]\right] \\
&= \mathbb{E}\left[\prod_{k\geq 1}\left(1 - \mathbb{E}\left[1 - e^{-\tilde{h}_{j,k}} \mid \tilde{\mu}_0\right]\delta_{\tilde{x}_k}(0,t]\right)\right] \\
&= \mathbb{E}\left[\prod_{k\geq 1}\left(1 - \frac{c_j(\tilde{x}_k)\, F\left(\tilde{h}_{0,k}\right)}{c_j(\tilde{x}_k)\, F_j\left(\tilde{h}_{0,k}\right) + c_j(\tilde{x}_k)\left(1 - F_j\left(\tilde{h}_{0,k}\right)\right)}\delta_{\tilde{x}_k}(0,t]\right)\right] \\
&= \mathbb{E}\left[\prod_{k\geq 1}\left(1 - F_j\left(\tilde{h}_{0,k}\right)\delta_{\tilde{x}_k}(0,t]\right)\right] = \exp\left(-\int_{\mathbb{X}\times\mathbb{R}^+}\left(1 - 1 + F_j(s)\,\delta_x(0,t]\right)\rho(s \mid x)ds\alpha(dx)\right) \\
&= \exp\left(-\int_0^t\int_{\mathbb{R}^+} F_j(s)\,\rho(s \mid x)ds\alpha(dx)\right).
\end{aligned}$$

Therefore, the survival function is proper if and only if

$$\lim_{t\to\infty}\exp\left(-\int_0^t\int_{\mathbb{R}^+} F_j(s)\,\rho(s \mid x)ds\alpha(dx)\right) = 0.$$

Note that if $\tilde{\mu}_0$ is the CRM of a beta-stacy process with parameters (c,α), by definition

$$\rho(s \mid x)ds\alpha(dx) = \frac{e^{-sc(x)\alpha((x,\infty))}}{1 - e^{-s}}c(x)ds\alpha(dx).$$

So from (3.9) it follows that

$$\begin{aligned}
\exp\left(-\int_0^t\int_{\mathbb{R}^+} F_j(s)\,\rho(s \mid x)ds\alpha(dx)\right) &\leq \exp\left(-A_j\int_0^t\int_{\mathbb{R}^+} e^{-sc(x)\alpha((x,\infty))}c(x)ds\alpha(dx)\right) \\
= \exp\left(-A_j\int_0^t \frac{1}{c(x)\alpha((x,\infty))}c(x)\alpha(dx)\right) &= \exp\left(-A_j\int_0^t \frac{\alpha(dx)}{\alpha((x,\infty))}\right).
\end{aligned}$$

Let us conclude the proof showing that the last quantity goes to zero as $t \to \infty$.
If α is the standard exponential measure, i.e., if $\alpha((0,x]) = 1 - e^{-x}$, it follows that

$$\exp\left(-A_j \int_0^t \frac{\alpha(dx)}{\alpha((x,\infty))}\right) = \exp\left(-A_j \int_0^t \frac{e^{-x}dx}{e^{-x}}\right) = e^{-A_j t}.$$

Since the last element goes to zero as $t \to \infty$, the thesis follows.
If α is the uniform measure between 0 and τ, i.e., if $\alpha((0,x]) = \frac{x}{\tau}\mathbb{1}_{(0,\tau]}(x)$, it follows that

$$\exp\left(-A_j \int_0^t \frac{\alpha(dx)}{\alpha((x,\infty))}\right) = \exp\left(-A_j \int_0^t \frac{1}{\tau}\frac{1}{1-\frac{x}{\tau}}\mathbb{1}_{(0,\tau]}(x)dx\right) = \exp\left(-A_j \int_0^t \frac{1}{\tau - x}\mathbb{1}_{(0,\tau]}(x)dx\right)$$
$$= \exp\left(A_j \log(\tau - x)\,|_0^{\min(t,\tau)}\right) = \exp(A_j\,(\log(\tau - \min(t,\tau)) - \log(\tau)))$$
$$= \exp\left(\log\left(\left(1 - \frac{\min(t,\tau)}{\tau}\right)^{A_j}\right)\right) = \left(1 - \frac{\min(t,\tau)}{\tau}\right)^{A_j} \to 0 \quad \text{as} \quad t \to \infty.$$

□

Proof of Corollary 3.4.1. From (3.5), the joint distribution is infinitesimally equal to

$$\prod_{r=1}^{K}\prod_{\Delta_r^*=1}\left\{\alpha(\mathrm{d}X_r^*)\prod_{j=1}^{d}\int_0^\infty e^{-h(\bar{n}_{r+1,j}+\tilde{n}^c_{r,j})}(1-e^{-h})^{n_{r,j}} f_j(h \mid s_r, X_r^*, c_j, F_j)\mathrm{d}h\rho(s_r \mid X_r^*)\mathrm{d}s_r\right\}$$
$$\times \exp\left\{-\int_{\mathbb{X}}\int_0^\infty \left[1 - \prod_{j=1}^{d}\int_0^\infty e^{-hN_j(x)} f_j(h \mid s, x, c_j, F_j)\mathrm{d}h\right]\rho(s \mid x)\mathrm{d}s\alpha(\mathrm{d}x)\right\}.$$

In order to explicitly find the densities $f_j(\cdot \mid s, x, c_j, F_j)$ for each $j = 1, \dots, d$, let us recall that the laws of the jumps are defined in Section 3.4. Let $g : \mathbb{R} \to \mathbb{R}$ be a the function $g(s) = -log(1-s)$ and let us now compute the law of $g(\tilde{h}_{r,j}) \mid \tilde{\mu}_0$ for each $j = 1, \dots, d$ and $r = 1, \dots, K$ such that $\Delta_r^* = 1$.
Since if $W \sim \text{Beta}(c_j(x)F_j(s), c_j(x)(1-F_j(s)))$ then its density function is

$$f_W(w) = \frac{w^{c_j(x)F_j(s)-1}(1-w)^{c_j(x)(1-F_j(s))-1}}{Be(c_j(x)F_j(s), c_j(x)(1-F_j(s)))},$$

where $Be(\cdot)$ is the beta function. It follows that the density function of $Y = g(X)$ is

$$f_Y(y) = f_W\left(g^{-1}(y)\right)e^{-y}.$$

So the density of the jth jump is

$$f_j(h \mid s, x, c_j, F_j) = e^{-h} \frac{\left(1-e^{-h}\right)^{c_j(x)F_j(s)-1}\left(1-\left(1-e^{-h}\right)\right)^{c_j(x)(1-F_j(s))-1}}{Be(c_j(x)F_j(s), c_j(x)(1-F_j(s)))} =$$

$$= \frac{\left(1-e^{-h}\right)^{c_j(x)F_j(s)-1} e^{-hc_j(x)(1-F_j(s))}}{Be(c_j(x)F_j(s), c_j(x)(1-F_j(s)))}.$$

Moreover, the kernel density of the base measure is

$$\rho(s \mid x) = \frac{c \cdot e^{-sc\alpha((x,+\infty))}}{1-e^{-s}}$$

by definition. So finally the thesis follows. □

Proof of Corollary 3.4.2. The conditional laws of the jumps $\tilde{h}_{j,k}$'s and the jumps $J_{j,r}$'s follows directly from Theorem 8. From the same theorem it follows that the posterior base measure $\tilde{\mu}_0'$ of the hierarchical beta-stacy is a CRM having Lévy intensity given by

$$\nu'(\mathrm{d}s, \mathrm{d}x) = \prod_{j=1}^{d} \int_0^\infty e^{-hN_j(x)} f_j(h \mid s, x, c_j) \mathrm{d}h \rho(s \mid x) \mathrm{d}s \alpha(\mathrm{d}x), \tag{3.29}$$

where f_j is the prior density of the conditional jumps.
In order to solve the integral in (3.29) with respect to f_j, note that it can be written as an expected value with respect to a Beta distribution with parameters $(c_j(x)(1 - F_j(s)), c_j(x)F_j(s))$; let us call $\tilde{B}$ such a distribution. Then,

$$\int_0^\infty e^{-hN_j(x)} f_j(h \mid s, x, c_j) \mathrm{d}h = \mathbb{E}_{\tilde{B}}\left[\tilde{B}^{N_j(x)}\right],$$

and therefore (3.29) can be written as

$$\nu'(\mathrm{d}s, \mathrm{d}x) = \prod_{j=1}^{d} \left[\frac{Be(c_j(x)F_j(s), N_j(x) + c_j(x)(1-F_j(s)))}{Be(c_j(x)F_j(s), c_j(x)(1-F_j(s)))}\right] \rho(s \mid x) \mathrm{d}s \alpha(\mathrm{d}x),$$

where $Be(\cdot, \cdot)$ is the Beta function. Therefore, the Lévy intensity of the posterior base measure in (3.29) as in the thesis follows. □

Proof of Proposition 3.4.3. Let $\tilde{S}_j'$ be the jth posterior survival function from the hierarchical beta-stacy prior, whose corresponding posterior CRM is

$$\tilde{\mu}_j \mid \boldsymbol{D}, s \stackrel{d}{=} \tilde{\mu}'_j + \sum_{r=1,\Delta^*_r=1}^{K} J_{j,r}\delta_{X^*_r}.$$

Since

$$\tilde{S}'_j(t) \stackrel{d}{=} e^{-\tilde{\mu}_j((0,t]} \mid \boldsymbol{D}, s \quad \text{for each } t > 0,$$

its Bayesian estimator is

$$\hat{S}'_j(t) = \mathbb{E}\left[\tilde{S}'_j(t)\right] = \mathbb{E}\left[e^{-\tilde{\mu}_j(0,t]} \mid \boldsymbol{D}, s\right] = \mathbb{E}\left[e^{-\tilde{\mu}'_j(0,t] - \sum_r J_{j,r}\delta_{X^*_r}(0,t]}\right].$$

From the independence properties follows that the above quantity is equal to

$$\mathbb{E}\left[e^{-\tilde{\mu}'_j(0,t]}\right]\prod_r \mathbb{E}\left[e^{-J_{j,r}\delta_{X^*_r}(0,t]}\right] = \mathbb{E}\left[e^{-\tilde{\mu}'_j(0,t]}\right]\prod_r \mathbb{E}\left[e^{-J_{j,r}}\mathbb{1}_{(0,t]}(X^*_r) + 1 - \mathbb{1}_{(0,t]}(X^*_r)\right].$$

Since each observation X^*_r is fixed, the above quantity is equal to

$$\begin{aligned} &\mathbb{E}\left[e^{-\tilde{\mu}'_j(0,t]}\right]\prod_r \left(\mathbb{E}\left[e^{-J_{j,r}}\right]\mathbb{1}_{(0,t]}(X^*_r) + 1 - \mathbb{1}_{(0,t]}(X^*_r)\right) = \\ &= \mathbb{E}\left[e^{-\tilde{\mu}'_j(0,t]}\right] \times \prod_r \left(-\mathbb{E}\left[1 - e^{-J_{j,r}}\right]\mathbb{1}_{(0,t]}(X^*_r) + 1\right). \end{aligned} \tag{3.30}$$

We can separately compute each term of (3.30). The first term can be written as

$$\mathbb{E}\left[e^{-\tilde{\mu}'_j(0,t]}\right] = \mathbb{E}\left[\mathbb{E}\left[e^{-\tilde{\mu}'_j(0,t]} \mid \tilde{\mu}'_0\right]\right],$$

and since $\tilde{\mu}'_j \mid \tilde{\mu}'_0 = \sum_{k\geq 1}\tilde{h}'_{j,k}\delta_{\tilde{x}'_k}$, then

$$\mathbb{E}\left[e^{-\tilde{\mu}'_j(0,t]}\right] = \mathbb{E}\left[\prod_{k\geq 1}\mathbb{E}\left[e^{-\tilde{h}'_{j,k}\delta_{\tilde{x}'_k}(0,t]} \mid \tilde{\mu}'_0\right]\right].$$

The inner expected value in the previous expression can be explicitly computed. From Corollary 3.4.2,

$$1 - e^{-\tilde{h}'_{j,k}} \mid \tilde{\mu}'_0 \sim Beta\left(c_j\left(\tilde{x}'_k\right)F_j\left(\tilde{h}'_{0,k}\right), c_j\left(\tilde{x}'_k\right)\left(1 - F_j\left(\tilde{h}'_{0,k}\right)\right) + N_j\left(\tilde{x}'_k\right)\right),$$

so

$$e^{-\tilde{h}'_{j,k}} \mid \tilde{\mu}'_0 \sim Beta\left(c_j\left(\tilde{x}'_k\right)\left(1-F_j\left(\tilde{h}'_{0,k}\right)\right)+N_j\left(\tilde{x}'_k\right), c_j\left(\tilde{x}'_k\right)F_j\left(\tilde{h}'_{0,k}\right)\right).$$

So for each $k \geq 1$, the value $\mathbb{E}\left[e^{-\tilde{h}'_{j,k}\delta_{\tilde{x}'_k}} \mid \tilde{\mu}'_0\right]$ is the expected value of the above Beta distribution, as a function of $\tilde{x}'_k$ and $\tilde{h}'_{0,k}$; let us denote as $f_j\left(\tilde{x}'_k, \tilde{h}'_{0,k}\right)$ this expected value. Then the first element of (3.30) can be written as

$$\begin{aligned}\mathbb{E}\left[e^{-\tilde{\mu}'_j(0,t]}\right] &= \mathbb{E}\left[\prod_{k\geq 1:\tilde{x}_k\leq t} f_j\left(\tilde{x}'_k, \tilde{h}'_{0,k}\right)\right] = \mathbb{E}\left[exp\left(\sum_{k\geq 1:\tilde{x}_k\leq t} log\left(f_j\left(\tilde{x}'_k, \tilde{h}'_{0,k}\right)\right)\right)\right] \\ &= \mathbb{E}\left[exp\left(\int_0^t \int_{\mathbb{R}^+} log\left(f_j\left(x,s\right)\right)\tilde{N}\left(\mathrm{d}x, \mathrm{d}s\right)\right)\right],\end{aligned}$$

where $\tilde{N} = \sum_{k\geq 1}\delta_{\tilde{x}'_k, \tilde{h}'_{0,k}}$ is the corresponding marked Poisson process. Observe that the last quantity in the previous expression is equal to

$$exp\left[-\int_0^t \int_{\mathbb{R}^+}\left(1-f_j\left(x,s\right)\right)\nu'\left(\mathrm{d}x, \mathrm{d}s\right)\right],$$

where ν' is the Lévy intensity of $\tilde{\mu}'_0$. Since

$$\begin{aligned}f_j\left(x,s\right) = \mathbb{E}\left[e^{-s\delta_x} \mid \tilde{\mu}'_0\right] &= \frac{c_j\left(x\right)\left(1-F_j\left(s\right)\right)+N_j\left(x\right)}{c_j\left('\right)\left(1-F_j\left(s\right)\right)+N_j\left(x\right)+c_j\left(x\right)F_j\left(s\right)} \\ &= \frac{c_j\left(x\right)\left(1-F_j\left(s\right)\right)+N_j\left(x\right)}{c_j\left(x\right)+N_j\left(x\right)},\end{aligned}$$

it follows that the first term of (3.30) can be computed as

$$\begin{aligned}\mathbb{E}\left[e^{-\tilde{\mu}'_j(0,t]}\right] &= exp\left[-\int_0^t \int_{\mathbb{R}^+}\left(1-\frac{c_j\left(x\right)\left(1-F_j\left(s\right)\right)+N_j\left(x\right)}{c_j\left(x\right)+N_j\left(x\right)}\right)\nu'\left(\mathrm{d}x, \mathrm{d}s\right)\right] = \\ &= exp\left[-\int_0^t \int_{\mathbb{R}^+}\frac{c_j\left(x\right)F_j\left(s\right)}{c_j\left(x\right)+N\left(x\right)}\nu'\left(\mathrm{d}x, \mathrm{d}s\right)\right].\end{aligned} \tag{3.31}$$

We now focus on the second term if (3.30). In particular, the expected value of each $1-e^{-J_{j,r}}$ can be computed since, from Corollary 3.4.2, we have that

$$1 - e^{-J_{j,r}} \sim Beta(n_{r,j} + c_j\left(X_r^*\right)F_j(s_r), \overline{n}_{r+1,j} + \tilde{n}_{r,j}^c + c_j\left(X_r^*\right)\left(1 - F_j(s_r)\right),$$

where c_j and F_j are the parameters of the jth group. Then the expected value of $1 - e^{-J_{j,r}}$ is

$$\begin{aligned}\mathbb{E}\left[1 - e^{-J_{j,r}}\right] &= \frac{n_{r,j} + c\left(X_r^*\right)F_j(s_r)}{n_{r,j} + c\left(X_r^*\right)F_j(s_r) + \overline{n}_{r+1,j} + \tilde{n}_{r,j} + c\left(X_r^*\right)\left(1 - F_j(s_r)\right)} = \\ &= \frac{n_{r,j} + c\left(X_r^*\right)F_j(s_r)}{n_{r,j} + \overline{n}_{r+1,j} + \tilde{n}_{r,j} + c_j\left(X_r^*\right)}.\end{aligned}$$

It follows that the second term in (3.30) is

$$\begin{aligned}&\prod_r \left(-\mathbb{E}\left[1 - e^{-J_{j,r}}\right]\mathbb{1}_{(0,t]}(X_r^*) + 1\right) \\ &= \prod_{r:\Delta_r^*=1}\left(-\frac{n_r + c\left(X_r^*\right)F(s_r)}{n_r + \overline{n}_{r+1} + \tilde{n}_r + c\left(X_r^*\right)}\mathbb{1}_{(0,t]}(X_r^*) + 1\right).\end{aligned} \tag{3.32}$$

So finally the thesis follows substituting (3.31) and (3.32) into (3.30). □

3.C Full-conditional distributions

The purpose of this section is to calculate the full conditional distributions useful for the sampling algorithm described in Section 3.5.2. First of all, we rewrite the joint posterior distribution of the model as follows. From Corollary 3.4.1, the joint distribution is infinitesimally equal to

$$
\prod_{r=1}^{K}\prod_{\Delta_r^*=1}\left\{\alpha(\mathrm{d}X_r^*)\prod_{j=1}^{d}\int_0^\infty \frac{\left(1-e^{-h}\right)^{n_{r,j}+c_j(X_r^*)F_j(s_r)-1}e^{-h\left(\bar{n}_{r+1,j}+\tilde{n}_{r,j}^c+c_j(X_r^*)(1-F_j(s_r))\right)}}{Be(c_j(X_r^*)F_j(s_r),c_j(X_r^*)(1-F_j(s_r)))}\mathrm{d}h\right.
$$
$$
\left.\times\frac{e^{-s_r c(X_r^*)\alpha((X_r^*,\infty))}}{1-e^{-s_r}}c(X_r^*)\mathrm{d}s_r\right\}
$$
$$
\times\exp\left\{-\int_{\mathbb{X}}\int_0^\infty\left[1-\prod_{j=1}^{d}\int_0^\infty\frac{\left(1-e^{-h}\right)^{c_j(x)F_j(s)-1}e^{-h(N_j(x)+c_j(x)(1-F_j(s)))}}{Be(c_j(x)F_j(s),c_j(x)(1-F_j(s)))}\mathrm{d}h\right]\right.
$$
$$
\left.\times\frac{e^{-sc(x)\alpha((x,\infty))}}{1-e^{-s}}c(x)\mathrm{d}s\alpha(\mathrm{d}x)\right\}.
$$

In order to avoid the first integral, let us introduce a further level of latent variables $\boldsymbol{h}=(h_{r,j})_{r,j}$, as $j=1,\dots,d$ and $r=1,\dots,K$ such that $\Delta_r^*=1$. With the choices described in Section 3.5, $c(x)=c\in\mathbb{R}^+$, $c_j(x)=c_*\in\mathbb{R}^+$ and $F_j(s)=F(s)=1-e^{-s}$ for any $j=1,\dots,d$. So finally the joint distribution as in (3.5) is infinitesimally equal to

$$
\begin{aligned}
&\prod_{r=1}^{K}\prod_{\Delta_r^*=1}\alpha(\mathrm{d}X_r^*)\frac{e^{-s_r c\alpha((X_r^*,+\infty))}}{1-e^{-s_r}}\\
&\qquad\times c\prod_{j=1}^{d}\frac{\left(1-e^{-h_{r,j}}\right)^{n_{r,j}+c_*F(s_r)-1}e^{-h_{r,j}\left(c_*(1-F(s_r))+\bar{n}_{r+1,j}+\tilde{n}_{r,j}^c\right)}}{Be(c_*F(s_r),c_*(1-F(s_r)))}\mathrm{d}h_{r,j}\mathrm{d}s_r\\
&\quad\times\exp\left\{-\int_{\mathbb{X}}\int_0^\infty\left[1-\prod_{j=1}^{d}\int_0^\infty\frac{(1-e^{-h})^{c_*F(s)-1}\cdot e^{-h(c_*(1-F(s))+N_j(x))}}{Be(c_*F(s),c_*(1-F(s)))}\mathrm{d}h\right]\right.\\
&\qquad\left.\times\frac{e^{-sc\alpha((x,\infty))}}{1-e^{-s}}c\mathrm{d}s\alpha(\mathrm{d}x)\right\}.
\end{aligned}
\tag{3.33}
$$

C.1 Full-conditional distribution of $\boldsymbol{s}$

Starting from (3.33), we now recover the full conditional distribution of the latent vector $\boldsymbol{s}$. In particular, for each $r=1,\dots,d$ as $\Delta_r^*=1$, let us consider the rth component s_r of the latent vector $\boldsymbol{s}$. The aim of this part is to compute the full conditional distribution of the transformed variable $F(s_r)=1-e^{-s_r}$.

Denoting as y_r the variable $F(s_r)=1-e^{-s_r}$, the full conditional of the transformed variable can be obtained from (3.33) as

$$\pi(y_r \mid \boldsymbol{h}, c, c_*) = \frac{1}{1-y_r}\pi(-\log(1-y_r) \mid \boldsymbol{h}, c, c_*) =$$
$$= \frac{1}{1-y_r}\prod_{j=1}^{d} \frac{e^{h_{r,j}c_* y_r + \log(1-y_r)c\alpha((X_r^*,+\infty))} \cdot \left(1-e^{-h_{r,j}}\right)^{c_* y_r}}{y_r \cdot Be(c_* y_r, c_*(1-y_r))} =$$
$$= \frac{1}{1-y_r}\prod_{j=1}^{d} \frac{e^{h_{r,j}c_* y_r} \cdot (1-y_r)^{c\alpha((x_r^*,+\infty))} \cdot \left(1-e^{-h_{r,j}}\right)^{c_* y_r}}{y_r \cdot Be(c_* y_r, c_j(x_r^*)(1-y_r))} =$$
$$= \frac{e^{y_r \sum_{j=1}^{d} h_{r,j}c_*} \cdot (1-y_r)^{dc\alpha((x_r^*,+\infty))-1} \prod_{j=1}^{d}\left(1-e^{-h_{r,j}}\right)^{c_* y_r}}{y_r^d \cdot \prod_{j=1}^{d} Be(c_* y_r, c_*(1-y_r))}.$$

Finally, the full conditional log-distribution of the rth component of the latent vector $\boldsymbol{s}$ is

$$\begin{aligned}\log\pi(y_r \mid \boldsymbol{h}, c, c_*) =& y_r \sum_{j=1}^{d} h_{r,j}c_* + [d \cdot c \cdot \alpha((X_r^*,+\infty)) - 1] \cdot \log(1-y_r) \\ &+ y_r \sum_{j=1}^{d} c_* \log(1-e^{-h_{r,j}}) \\ &- d \cdot \log(y_r) - \sum_{j=1}^{d} \log Be(c_* y_r, c_*(1-y_r))\end{aligned} \tag{3.34}$$

C.2 Full-conditional distribution of $\boldsymbol{h}$

From (3.33) and for each $j = 1, \ldots d$ and $r = 1, \ldots, K$ such that $\Delta_r^* = 1$, the full conditional distribution of each component $h_{j,r}$ of the set of further latent variables $\boldsymbol{h}$ is

$$e^{-h_{r,j}(\bar{n}_{r+1,j}+\tilde{n}_{r,j}^c)}(1-e^{-h_{j,r}})^{n_{r,j}} e^{-h_{r,j}} \left(1-e^{-h_{r,j}}\right)^{c_* F(s_r)-1} e^{-h_{r,j}(c_*(1-F(s_r))-1)}$$
$$= e^{-h_{r,j}(\bar{n}_{r+1,j}+\tilde{n}_{r,j}^c+c_j(x_r^*)(1-F_j(s_r)))} \cdot (1-e^{-h_{j,r}})^{n_{r,j}+c_j(x_r^*)F_j(s_r)-1}.$$

Considering $x_{r,j} = e^{-h_{r,j}}$ it follows that

$$f_{X_{r,j}}(x_{r,j}) = f_{X_{r,j}}(g^{-1}(x_{r,j}) \left| \frac{d}{dx_{r,j}} g^{-1}(x_{r,j}) \right|,$$

where

$$g(h) = e^{-h} \quad \text{and so} \quad g^{-1}(x_{r,j}) = -log(x_{r,j}).$$

So finally

$$\begin{aligned} f_{X_{r,j}}(x_{r,j}) &\propto x_{r,j}^{\bar{n}_{r+1,j}+\tilde{n}^c_{r,j}+c_*(1-F(s_r))} \cdot (1-x_{r,j})^{n_{r,j}+c_*F(s_r)-1} \cdot \frac{1}{x_{r,j}} \\ &= x_{r,j}^{\bar{n}_{r+1,j}+\tilde{n}^c_{r,j}+c_*(1-F(s_r))-1} \cdot (1-x_{r,j})^{n_{r,j}+c_*F(s_r)-1}. \end{aligned}$$

So for each j, r it follows that the component $h_{r,j}$ can be sampled as

$$e^{-h_{r,j}} \sim \mathrm{Beta}\left(\mathrm{\bar{n}_{r+1,j}} + \mathrm{\tilde{n}^c_{r,j}} + \mathrm{c_*}\left(1 - \mathrm{F(s_r)}\right), \mathrm{n_{r,j}} + \mathrm{c_* F(s_r)}\right).$$

It follows that

$$1 - e^{-h_{r,j}} \sim \mathrm{Beta}\left(\mathrm{n_{r,j}} + \mathrm{c_* F(s_r)}, \mathrm{\bar{n}_{r+1,j}} + \mathrm{\tilde{n}^c_{r,j}} + \mathrm{c_*}\left(1 - \mathrm{F(s_r)}\right)\right).$$

Note that the previous equation represents the same law of the discrete jumps $J_{r,j}$ provided in (3.12), so that we can sample each $J_{r,j}$ and set $h_{r,j} = J_{r,j}$ for each r, j.

C.3 Full-conditional distribution of c

Let us assume a Gamma(k, θ) prior on c. So from (3.33) the full conditional of c becomes

$$\begin{aligned} &\prod_{r=1}^{K}\prod_{\Delta^*_r=1} \alpha(\mathrm{d}x^*_r) \frac{e^{-s_r c\alpha((X^*_r,+\infty))}}{1-e^{-s_r}} c \prod_{j=1}^{d} \frac{\left(1-e^{-h_{r,j}}\right)^{n_{r,j}+c_*F(s_r)-1} e^{-h_{r,j}\left(c_*(1-F(s_r))+\bar{n}_{r+1,j}+\tilde{n}^c_{r,j}\right)}}{Be(c_*F(s_r), c_*(1-F(s_r)))} \mathrm{d}h_{j,r}\mathrm{d}s_r \\ &\times \exp\left\{-\int_{\mathbb{X}}\int_0^\infty \left[1 - \prod_{j=1}^{d}\int_0^\infty \frac{(1-e^{-h})^{c_*F(s)-1} \cdot e^{-h(c_*(1-F(s))+N_j(x))}}{Be(c_*F(s), c_*(1-F(s)))}\mathrm{d}h\right] \frac{e^{-sc\alpha((x,\infty))}}{1-e^{-s}} c\mathrm{d}s\alpha(\mathrm{d}x)\right\} \\ &\times c^{k-1} e^{-\frac{c}{\theta}}. \end{aligned}$$

Note that

$$\begin{aligned} &\int_0^\infty \frac{(1-e^{-h})^{c_*F(s)-1} \cdot e^{-h(c_*(1-F(s))+N_j(x))}}{Be(c_*F(s), c_*(1-F(s)))}\mathrm{d}h = \\ &= \frac{\int_0^1 (1-x)^{c_*F(s)-1} \cdot x^{c_*(1-F(s))+N_j(x)-1} dx}{Be(c_*F(s), c_*(1-F(s)))} = \frac{Be(c_*(1-F(s))+N_j(x), c_*F(s))}{Be(c_*F(s), c_*(1-F(s)))}. \end{aligned}$$

So the full-conditional for c becomes

$$\prod_{r=1}^{K}\prod_{\Delta_r^*=1}\alpha(\mathrm{d}x_r^*)\frac{e^{-s_r c\alpha((X_r^*,+\infty))}}{1-e^{-s_r}}c\prod_{j=1}^{d}\frac{\left(1-e^{-h_{r,j}}\right)^{n_{r,j}+c_*F(s_r)-1}e^{-h_{r,j}\left(c_*(1-F(s_r))+\bar{n}_{r+1,j}+\tilde{n}^c_{r,j}\right)}}{Be(c_*F(s_r),c_*(1-F(s_r)))}\mathrm{d}h_{j,r}\mathrm{d}s_r$$
$$\times\exp\left\{-\int_{\mathbb{X}}\int_0^\infty\left[1-\prod_{j=1}^{d}\frac{Be(c_*(1-F(s))+N_j(x),c_*F(s))}{Be(c_*F(s),c_*(1-F(s)))}\right]\frac{e^{-sc\alpha((x,\infty))}}{1-e^{-s}}c\mathrm{d}s\alpha(\mathrm{d}x)\right\}$$
$$\times c^{k-1}e^{-\frac{c}{\theta}}.$$

Since α is distributed as a uniform distribution between 0 and τ, $\alpha((x,\infty))=\left(1-\frac{x}{\tau}\right)\mathbb{1}_{(0,\tau)}(x)$ and $\alpha(dx)=\frac{1}{\tau}\mathbb{1}_{(0,\tau)}(dx)$. Switching the integrals, the double integral can be rewritten as follows:

$$\int_{\mathbb{X}}\int_0^\infty\left[1-\prod_{j=1}^{d}\frac{Be(c_*(1-F(s))+N_j(x),c_*F(s))}{Be(c_*F(s),c_*(1-F(s)))}\right]\frac{e^{-scB_0((x,\infty))}}{1-e^{-s}}c\mathrm{d}s\alpha(\mathrm{d}x)$$
$$=\int_0^\infty\left[\int_0^\tau\left[1-\prod_{j=1}^{d}\frac{Be(c_*(1-F(s))+N_j(x),c_*F(s))}{Be(c_*F(s),c_*(1-F(s)))}\right]\frac{e^{-sc\left(1-\frac{x}{\tau}\right)}}{1-e^{-s}}\cdot c\cdot\frac{1}{\tau}\mathrm{d}x\right]\mathrm{d}s$$
$$=\int_0^\infty\frac{1}{1-e^{-s}}\left[\int_0^\tau\left[1-\prod_{j=1}^{d}\frac{Be(c_*(1-F(s))+N_j(x),c_*F(s))}{Be(c_*F(s),c_*(1-F(s)))}\right]\frac{e^{-sc\left(1-\frac{x}{\tau}\right)}}{\tau}\cdot c\mathrm{d}x\right]\mathrm{d}s.$$

Recalling that K is the number of distinct observations from all the d groups,

$$X_1^*,\ldots,X_K^*,$$

let us define the number of observations from group j equal to X_r^* as $n_{r,j}^*\geq 0$, for each $j=1,\ldots,d$ and $r=1,\ldots,K$. Let us assume that $X_0^*=0$, and note that $X_K^*=\tau$.
Note that for each time $x>0$ the general at-risk process for group j can be written as follows:

$$N_j(x)=\sum_{i=1}^{N_j}\mathbb{1}_{[x,\infty)}(X_{i,j})=\sum_{r=1}^{K}\left(\sum_{g=r}^{K}n_{g,j}^*\right)\mathbb{1}_{(X_{r-1}^*,X_r^*]}(x).$$

By definition, it follows that in each interval $(X^*_{r-1}, X^*_r]$ the function $N_j(x)$ is a constant, and it is equal to

$$N^*_{r,j} := \sum_{g=r}^{K} n^*_{g,j}.$$

Note that $N_j(x) = 0$ on (X^*_K, ∞). We can therefore write the inner integral as follows:

$$\begin{aligned}
&\int_0^\tau \left[1 - \prod_{j=1}^{d} \frac{Be(c_*(1-F(s)) + N_*(x), c_*F(s))}{Be(c_*F(s), c_*(1-F(s)))}\right] \frac{e^{-sc\left(1-\frac{x}{\tau}\right)}}{\tau} \cdot c\mathrm{d}x \\
&= \sum_{r=1}^{K} \int_{X^*_{r-1}}^{X^*_r} \left[1 - \prod_{j=1}^{d} \frac{Be(c_*(1-F(s)) + N_j(x), c_*F(s))}{Be(c_*F(s), c_*(1-F(s)))}\right] \frac{e^{-sc\left(1-\frac{x}{\tau}\right)}}{\tau} \cdot c\mathrm{d}x \\
&= \sum_{r=1}^{K} \int_{x^*_{r-1}}^{x^*_r} \left[1 - \prod_{j=1}^{d} \frac{Be(c_*(1-F(s)) + N^*_{r,j}, c_*F(s))}{Be(c_*F(s), c_*(1-F(s)))}\right] \frac{e^{-sc\left(1-\frac{x}{\tau}\right)}}{\tau} \cdot c\mathrm{d}x \\
&= \sum_{r=1}^{K} \left[1 - \prod_{j=1}^{d} \frac{Be(c_*(1-F(s)) + N^*_{r,j}, c_*F(s))}{Be(c_*F(s), c_*(1-F(s)))}\right] \int_{x^*_{r-1}}^{x^*_r} \frac{c}{\tau} \cdot e^{-sc\left(1-\frac{x}{\tau}\right)}\mathrm{d}x \\
&= \sum_{r=1}^{K} \left[1 - \prod_{j=1}^{d} \frac{Be(c_*(1-F(s)) + N^*_{r,j}, c_*F(s))}{Be(c_*F(s), c_*(1-F(s)))}\right] \cdot \frac{e^{-sc\left(1-\frac{X^*_r}{\tau}\right)} - e^{-sc\left(1-\frac{X^*_{r-1}}{\tau}\right)}}{s}.
\end{aligned}$$

Since $N^*_{r,j}$ can be computed for any $r = 1, \ldots, K$, the double integral can be rewritten as follows:

$$\begin{aligned}
&\int_{\mathbb{X}} \int_0^\infty \left[1 - \prod_{j=1}^{d} \frac{Be(c_*(1-F(s)) + N_j(x), c_*F(s))}{Be(c_*F(s), c_*(1-F(s)))}\right] \frac{e^{-sc\alpha((x,\infty))}}{1-e^{-s}} c\mathrm{d}s\alpha(\mathrm{d}x) \\
&= \int_0^\infty \frac{1}{1-e^{-s}} \left[\int_0^\tau \left[1 - \prod_{j=1}^{d} \frac{Be(c_*(1-F(s)) + N_j(x), c_*F(s))}{Be(c_*F(s), c_*(1-F(s)))}\right] \frac{e^{-sc\left(1-\frac{x}{\tau}\right)}}{\tau} \cdot c\mathrm{d}x\right] \mathrm{d}s \\
&= \int_0^\infty \frac{1}{1-e^{-s}} \sum_{r=1}^{K} \left[1 - \prod_{j=1}^{d} \frac{Be(c_*(1-F(s)) + N^*_{r,j}, c_*F(s))}{Be(c_*F(s), c_*(1-F(s)))}\right] \cdot \frac{e^{-sc\left(1-\frac{X^*_r}{\tau}\right)} - e^{-sc\left(1-\frac{X^*_{r-1}}{\tau}\right)}}{s} \mathrm{d}s \\
&= \sum_{r=1}^{K} \int_0^\infty \left[1 - \prod_{j=1}^{d} \frac{Be(c_*(1-F(s)) + N^*_{r,j}, c_*F(s))}{Be(c_*F(s), c_*(1-F(s)))}\right] \cdot \frac{e^{-sc\left(1-\frac{X^*_r}{\tau}\right)} - e^{-sc\left(1-\frac{X^*_{r-1}}{\tau}\right)}}{s \cdot (1-e^{-s})} \mathrm{d}s.
\end{aligned}$$

It follows that the full conditional for c is proportional to

$$\prod_{r=1}^{K} \exp\left\{-\int_0^\infty \left[1-\prod_{j=1}^{d}\frac{Be(c_*(1-F(s))+N_{r,j}^*, c_*F(s))}{Be(c_*F(s), c_*(1-F(s)))}\right]\cdot\frac{e^{-s\cdot c\cdot\alpha(X_r^*)}-e^{-s\cdot c\cdot\alpha\left(X_{r-1}^*\right)}}{s\cdot(1-e^{-s})}\mathrm{d}s\right\}$$
$$\times\left[\prod_{r=1\ \Delta_r^*=1}^{K} e^{-s_r c\alpha((X_r^*,+\infty))}\right]\times c^{k-1}e^{-\frac{c}{\theta}}.$$

Let us now perform the change of variable $t := e^{-s}$ in the last integral left. So finally the full conditional distribution of c is

$$\pi\left(c \mid \boldsymbol{s},\boldsymbol{h},c_*\right)=\prod_{r=1}^{K}\exp\left\{-\int_0^1\left[1-\frac{\prod_{j=1}^{d} Be(c_*\cdot t+N_{r,j}^*, c_*\cdot(1-t))}{\left(Be(c_*\cdot(1-t), c_*\cdot t)\right)^d}\right]\cdot\frac{t^{c\cdot\alpha(X_r^*)}-t^{c\cdot\alpha\left(X_{r-1}^*\right)}}{-\log(t)\cdot t\cdot(1-t)}\mathrm{d}t\right\}$$
$$\times\left[\prod_{r=1\ \Delta_r^*=1}^{K} e^{-s_r c\alpha((X_r^*,+\infty))}\right]\times c^{k-1}e^{-\frac{c}{\theta}},$$

while the log-full conditional of c is

$$\log\pi\left(c\mid \boldsymbol{s},\boldsymbol{h},c_*\right)=$$
$$-\sum_{r=1}^{K}\int_0^1\left[1-\frac{\prod_{j=1}^{d} Be(c_*\cdot t+N_{r,j}^*, c_*\cdot(1-t))}{\left(Be(c_*\cdot(1-t), c_*\cdot t)\right)^d}\right]\cdot\frac{t^{c\cdot\alpha(X_r^*)}-t^{c\cdot\alpha\left(X_{r-1}^*\right)}}{-\log(t)\cdot t\cdot(1-t)}\mathrm{d}t$$
$$-c\sum_{r=1\ \Delta_r^*=1}^{K} s_r\alpha((X_r^*,+\infty))+(k-1)\log c-\frac{c}{\theta}.$$

C.4 Full-conditional distribution of c_*

Let us assume a $\Gamma(k_*,\theta_*)$ prior on c_*. So from (3.33) the full conditional of c_* becomes

$$\prod_{r=1}^{K}\prod_{\Delta_r^*=1}\alpha(\mathrm{d}x_r^*)\frac{e^{-s_r c\alpha((X_r^*,+\infty))}}{1-e^{-s_r}}c\prod_{j=1}^{d}\frac{\left(1-e^{-h_{r,j}}\right)^{n_{r,j}+c_*F_*(s_r)-1}e^{-h_{r,j}\left(c_*(1-F_*(s_r))+\bar{n}_{r+1,j}+\tilde{n}_{r,j}^c\right)}}{Be(c_*F_*(s_r),c_*(1-F_*(s_r)))}\mathrm{d}h_{j,r}\mathrm{d}s_r\times$$
$$\times\exp\left\{-\int_{\mathbb{X}}\int_0^\infty\left[1-\prod_{j=1}^{d}\frac{Be(c_*(1-F_*(s))+N_j(x),c_*F_*(s))}{Be(c_*F_*(s),c_*(1-F_*(s)))}\right]\frac{e^{-sc\alpha((x,\infty))}}{1-e^{-s}}c\mathrm{d}s\alpha(\mathrm{d}x)\right\}\times$$
$$\times c_*^{k_*-1}e^{-\frac{c_*}{\theta_*}}\propto$$
$$\propto\left[\prod_{r=1}^{K}\prod_{\Delta_r^*=1}\prod_{j=1}^{d}\frac{\left(1-e^{-h_{r,j}}\right)^{c_*F_*(s_r)}e^{-h_{r,j}c_*(1-F_*(s_r))}}{Be(c_*F_*(s_r),c_*(1-F_*(s_r)))}\right]\times c_*^{k_*-1}e^{-\frac{c_*}{\theta_*}}\times$$
$$\times\prod_{r=1}^{K}\exp\left\{-\int_0^\infty\left[1-\prod_{j=1}^{d}\frac{Be(c_*(1-F_*(s))+N_{r,j}^*,c_*F_*(s))}{Be(c_*F_*(s),c_*(1-F_*(s)))}\right]\cdot\frac{e^{-s\cdot c\cdot\alpha(X_r^*)}-e^{-s\cdot c\cdot\alpha\left(X_{r-1}^*\right)}}{s\cdot(1-e^{-s})}\mathrm{d}s\right\}.$$

Applying again the change of variable $t:=e^{-s}$ in the last integral left, the full conditional distribution of c_* becomes

$$\pi\left(c_*\mid\boldsymbol{s},\boldsymbol{h},c\right)=c_*^{k_*-1}e^{-\frac{c_*}{\theta_*}}\times\left[\prod_{r=1}^{K}\prod_{\Delta_r^*=1}\frac{\prod_{j=1}^{d}\left(1-e^{-h_{r,j}}\right)^{c_*F_*(s_r)}e^{-h_{r,j}c_*(1-F_*(s_r))}}{\left(Be(c_*F_*(s_r),c_*(1-F_*(s_r)))\right)^d}\right]\times$$
$$\times\prod_{r=1}^{K}\exp\left\{-\int_0^1\left[1-\frac{\prod_{j=1}^{d}Be(c_*\cdot t+N_{r,j}^*,c_*\cdot(1-t))}{\left(Be(c_*\cdot(1-t),c_*\cdot t)\right)^d}\right]\cdot\frac{t^{c\cdot\alpha(X_r^*)}-t^{c\cdot\alpha\left(X_{r-1}^*\right)}}{-\log(t)\cdot t\cdot(1-t)}\mathrm{d}t\right\},$$

while the log-full conditional of c_* is

$$\log\pi\left(c_*\mid\boldsymbol{s},\boldsymbol{h},c\right)=-\sum_{r=1}^{K}\int_0^1\left[1-\frac{\prod_{j=1}^{d}Be(c_*\cdot t+N_{r,j}^*,c_*\cdot(1-t))}{\left(Be(c_*\cdot(1-t),c_*\cdot t)\right)^d}\right]\cdot\frac{t^{c\cdot\alpha(X_r^*)}-t^{c\cdot\alpha\left(X_{r-1}^*\right)}}{-\log(t)\cdot t\cdot(1-t)}\mathrm{d}t$$
$$+\sum_{r=1}^{K}\sum_{\Delta_r^*=1}\left[c_*\sum_{j=1}^{d}\left(F_*(s_r)\log\left(1-e^{-h_{r,j}}\right)-h_{r,j}(1-F_*(s_r))\right)-d\cdot\log\left(Be(c_*F_*(s_r),c_*(1-F_*(s_r)))\right)\right]$$
$$+(k_*-1)\log(c_*)-\frac{c_*}{\theta_*}.$$

3.D Conditional algorithm

In this section, we provide a more detailed description of the conditional algorithm discussed in Section 3.5. Let us call R the number of iterations of the MCMC algorithm, R_{burnin} the burn-in parameter and R_{thin} the thinning parameter. We denote the initial values of the vector $\boldsymbol{s}$ and the parameters c and c_* as $\boldsymbol{s}_0$, c_0 and $c_{*,0}$ respectively. Note that in the posterior representation provided in Corollary 3.4.2, the latent vector $\boldsymbol{s}$ is only used to define the law described in (3.12); in this formula, each element s_r of $\boldsymbol{s}$ always appears as $F(s_r) = 1 - e^{s_r}$. We therefore define the transformed latent vector

$$\boldsymbol{y} = (y_r)_{r=1;\Delta_r^*=1}^K = (F(s_r))_{r=1;\Delta_r^*=1}^K .$$

The sampling of the latent variables and the two parameters c and c_* is done from the respective full-conditional distributions, which are described in Section 3.C. Each step of the sampling algorithm is divided into 4 parts: first of all, the sampling of the jumps of the discrete component of the posterior distribution; then, the sampling of the latent variables and the parameters c and c_*; finally, the approximation of the absolutely continuous part of the posterior distribution via the algorithm described in Wolpert and Ickstadt (1998) and the evaluation of the sampled survival functions on an evaluation grid T^{eval}. The number of total steps is $R_{burnin} + R_{thin} \times R$. For each step of the MCMC algorithm the following points are implemented.

(1.) Sample the discrete jumps $J_{j,r}$'s as in (3.12), for each $j = 1, \ldots, d$ and $r = 1, \ldots, K$ such that $\Delta_r^* = 1$.

(2.) Sample the latent variable $\boldsymbol{y} = (y_r)_{r=1;\Delta_r^*=1}^K$ and the parameters c and c_* according to their full conditional distributions (see Section 3.C).

(3.) Exploit the algorithm described in Wolpert and Ickstadt (1998) to sample a set of locations $(\hat{x}'_k)_{k=1}^M$ and jumps $(\hat{h}'_{0,k})_{k=1}^M$ for a M-dimensional finite approximation of the posterior base measure $\tilde{\mu}'_0$ as in (3.10), i.e.,

$$\tilde{\mu}_0'^M := \sum_{k=1}^M \hat{h}'_{0,k} \delta_{\hat{x}'_k}.$$

(4.) Sample the dependent continuous jump $\hat{h}'_{j,k} \mid \tilde{\mu}_0'^M$ as in (3.11), for each $j = 1, \ldots, d$ and $k = 1, \ldots, M$.

(5.) Each survival function is estimated on the grid T^{eval}, after a number of steps equal to R_{burnin}, and from that point onward, every R_{thin} steps. Therefore, for each $j = 1, \dots d$ and for each evaluation time $t \in T^{eval}$, the posterior measure $\tilde{\mu}_j(t) \mid \boldsymbol{D}, \boldsymbol{s}$ is computed as in (3.6), i.e.,

$$\tilde{\mu}_j(t) \mid \boldsymbol{D}, \boldsymbol{s} = \sum_{k=1}^{M} \hat{h}'_{j,k} \delta_{\hat{x}'_k}(0, t] + \sum_{r=1;\Delta^*_r=1}^{K} J_{j,r} \delta_{X^*_r}(0, t].$$

Finally, the evaluation of the jth survival function is

$$S_j(t) \mid \boldsymbol{D}, \boldsymbol{s} = e^{-\tilde{\mu}_j(t) \mid \boldsymbol{D}, \boldsymbol{s}}.$$

Aknowledgement

The work carried out over these years and the writing of this thesis were only made possible thanks to the understanding, the moral and technical support, and the constant availability to help and encourage me which have always been provided by my supervisors. For this reason, my acknowledgments gladly begin with Prof. **Federico Camerlenghi** and Prof. **Tommaso Rigon**, as they are literally the reason I have the opportunity to write them. A thank you and a thought also go to all my colleagues and fellow travelers, especially **Luca**, **Francesco**, **Jiefeng**, **Ludovica**, and **Claudia**. It was intense, it was strange, but without a doubt, it was fun.

Secondly, I would like to thank my managers **Andrea**, **Mauro** and **Maddalena** for giving me the opportunity to embark on this unusual multi-year experience and for granting me the necessary degrees of freedom to see it through. My thanks must also explicitly go to **Giovanni**: in these years, he certainly hasn't found himself without things to do, nor has he ever been at risk of getting bored due to a lack of tasks to carry out, but despite this, he dedicated time from the very beginning to help me organize my double life, acting as a filter and supporter, receiving friendship and socks in return.

I also thank my friends, each for a different reason. **Eric**, for sharing the beginning of this journey with me in a complicated situation, and for sharing the most iconic Ph.D. stories I have ever heard. **Giulio**, for being a reassuring presence and for making it clear to me that while it's always possible to do better, it's also true that it can always be worse. **Simone**, for showing me that although a Ph.D. hits hard, Greek tragedy hits harder. **Andrea**, for demonstrating to me every day that there are no crazy choices, only choices. **Daniele** and **Simone**, for being references and examples, perhaps unknowingly. And to all the other people with whom I shared these years of this experience, either in part or in its entirety.

The expressions of esteem and gratitude to my family certainly go beyond just this thesis, but I don't often have the opportunity to write down my thanks, so I'll take this chance: I thank **Donatella**, **Gilberto**, **Elena**, **Luca**, **Francesco**, **Teresa**, **Verza**, **Lavanda**, and **Mirtilla** for always being my home.

To conclude, it is customary that the last acknowledgment is the most heartfelt, and for this reason, it doesn't need to be the longest. I will therefore be brief: since every person needs a goal, the motivation to achieve it, and help along the way, I thank **Yuliya** for being all three.

Riccardo