Reinforced urns and the subdistribution beta-Stacy process prior for competing risks analysis

Andrea Arfè

Department of Decision Sciences, Bocconi University, 20136 Milan, Italy.

E-mail: andrea.arfe@phd.unibocconi.it

Stefano Peluso

Department of Statistical Sciences, Università Cattolica del Sacro Cuore, 20124 Milan, Italy.

E-mail: stefano.peluso@unicatt.it

Pietro Muliere

Department of Decision Sciences, Bocconi Institute of Data Science and Analytics, Bocconi University, 20136 Milan, Italy.

E-mail: pietro.muliere@unibocconi.it

Summary. In this paper we introduce the subdistribution beta-Stacy process, a novel Bayesian nonparametric process prior for subdistribution functions useful for the analysis of competing risks data. In particular, we i) characterize this process from a predictive perspective by means of an urn model with reinforcement, ii) show that it is conjugate with respect to right-censored data, and iii) highlight its relations with other prior processes for competing risks data. Additionally, we consider the subdistribution beta-Stacy process prior in a nonparametric regression model for discrete-time competing risks data, with application to the prognosis of HIV-infected men from the Amsterdam Cohort Studies.

Keywords: Bayesian nonparametrics; competing risks; prediction; reinforcement; subdistribution beta-Stacy; urn process.

1. Introduction

In the setting of clinical prognostic research with a time-to-event outcome, often the occurrence of one of several competing risks may preclude the occurrence of another event of interest (Kalbfleisch and Prentice, 2002, Chapter 8). In such cases it is typically of interest to assess i) the probability that one of the considered competing risks occurs within some time interval and ii) how this probability changes in association with predictors of interest (Wolbers et al., 2009; Fine, 1999; Putter et al., 2007). For example, using data from the Amsterdam Cohort Studies (Geskus et al., 2003), Putter et al. (2007) assessed the prognosis of HIV-infected men as a function of whether they presented or not a potentially protective deletion in the C–C chemokine receptor 5 gene (CCR5). Specifically, they considered the problem of predicting the time elapsed from seroconversion to the onset of two clinically relevant events: the emergence of the so-called Syncytium Inducing (SI) phenotype or the development of AIDS with a non-SI

phenotype (non-SI AIDS). Since the onset of either event is associated with a different subsequent prognosis, the two events are considered as competing risks.

Competing risks data has received widespread attention in the frequentist literature. It suffices to recall the comprehensive textbooks of Kalbfleisch and Prentice (2002), Aalen et al. (2008), Andersen et al. (2012), Lawless (2011), Crowder (2012), and Pintilie (2006). Putter et al. (2007), Wolbers et al. (2009), and Andersen et al. (2002) provide an introductory overview of standard approaches for competing risks data. Classical approaches to prediction in presence of competing risks focus on the *subdistribution function*, also known as the *cumulative incidence function*, which represents the probability that a specific event occurs within a given time period. Kalbfleisch and Prentice (2002, Chapter 8) describe a frequentist nonparametric estimator for the subdistribution function, while Fine and Gray, in their pivotal 1999 paper, introduced a semiparametric proportional hazards model for the subdistribution function. Fine (1999) and Scheike et al. (2008) considered alternative semiparametric estimators, whilst Larson and Dinse (1985), Jeong and Fine (2007), and Hinchliffe and Lambert (2013) considered parametric regression models for the subdistribution function.

In contrast with the frequentist literature, the Bayesian literature on competing risks is still sparse, although several relevant contributions can be identified. Ge and Chen (2012) introduced a semiparametric model for competing risks by separately modelling the subdistribution function of the primary event of interest and the conditional time-to-event distributions of the other competing risks. They modelled the baseline subdistribution hazards and the cause-specific hazards by means of a gamma process prior (see Nieto-Barajas and Walker, 2002 and Kalbfleisch and Prentice, 2002, Section 11.8). De Blasi and Hjort (2007) suggested a semiparametric proportional hazards regression model with logistic relative risk function for cause-specific hazards. For inference, they assign the common baseline cumulative hazard a beta process prior (Hjort, 1990). With the same approach, Hjort's extension of the beta process for nonhomogeneous Markov Chains (Hjort, 1990, Section 5) may be considered as a prior distribution on the set of cause-specific baseline hazards in a more general multiplicative hazards model (see Andersen et al., 2012, Chapter III and Lawless, 2011, Chapter 9). In the beta process for nonhomogeneous Markov Chains the individual transition hazards are necessarily independent (Hjort, 1990, Section 5). The beta-Dirichlet process, a generalization of the beta process introduced by Kim et al. (2012), relaxes this assumption by allowing for correlated hazards. Kim et al. (2012) use the beta-Dirichlet process to define a semiparametric semi-proportional transition hazards regression model for nonhomogeneous Markov Chains which, in the competing risks setting, could be used to model the cause-specific hazards. With the same purpose, Chae et al. (2013) proposed a nonparametric regression model based on a mixture of beta-Dirichlet process priors.

In this paper we introduce a novel stochastic process, a generalization of Walker and Muliere's beta-Stacy process (Walker and Muliere, 1997), which represents a prior distribution for the Bayesian nonparametric analysis of discrete-time competing risks data. This new process, which we call the *subdistribution beta-Stacy* process, is conjugate with respect to right censored observations, greatly simplifying the task of performing probabilistic predictions. We will also use the subdistribution beta-Stacy process to specify a Bayesian nonparametric competing risks regression model, thus generalizing the

survival regression approach of Rigat and Muliere (2012). Our perspective is Bayesian because the related interpretation of probability is intimately linked with the task of making predictions in presence of uncertainty (De Finetti, 1964; Singpurwalla, 1988) and nonparametric because we aim to avoid potentially restrictive and/or arbitrary parametric model components (Müller and Mitra, 2013; Hjort et al., 2010).

To characterize the subdistribution beta-Stacy process we adhere to the predictive approach for the construction of nonparametric process priors, i.e. of prior probability distributions on infinite-dimensional spaces of distribution functions (Ferguson, 1973; Hjort et al., 2010; Müller and Mitra, 2013). In this framework, a nonparametric prior distribution is implicitly characterized by specifying the predictive distribution of the observable quantities and then by appealing to results related to the celebrated de Finetti Representation Theorem (Walker and Muliere, 1999; Muliere et al., 2000; Epifani et al., 2002; Muliere et al., 2003; Bulla and Muliere, 2007; Fortini and Petrone, 2012). In the context of this paper, the predictive distribution represents a specific rule which prescribes how probabilistic predictions for a new patient should be performed after observing the experience of other similar (exchangeable) patients. Not only this approach is tailored to our predictive purposes, but it also avoids some conceptual difficulties that arise when specifying prior distributions for unobservable quantities, such as cause-specific hazards (De Finetti, 1964; Singpurwalla, 1988; Wechsler, 1993; Cifarelli and Regazzini, 1996; Bernardo and Smith, 2000, Chapter 4).

The predictive rule underlying the subdistribution beta-Stacy process will be described in terms of the laws determining the evolution of a *reinforced urn process* (Muliere et al., 2000). Other well-known prior processes were predictively characterized by Muliere et al. (2000) using reinforced urn processes, including Pólya trees (Mauldin et al., 1992), the beta-Stacy process (Walker and Muliere, 1997), and, more generally, neutral-to-the-right processes (Doksum, 1974). In detail, a reinforced urn process is a stochastic process with countable state-space S. Each point $x \in S$ is associated with an urn containing coloured balls. The possible colors of the ball are represented by the elements of the finite set E. Each urn $x \in S$ initially contains $n_x(c) \geq 0$ balls of color $c \in E$. The quantities $n_x(c)$ need not be integers, although thinking them as such simplifies the description of the process. For a fixed initial state x_0, recursively define the process as follows: i) if the current state is $x \in S$, then a ball is sampled from the corresponding urn and replaced together with a fixed amount $m > 0$ of additional balls of the same color; hence, the extracted color is "reinforced", i.e. made more likely to be extracted in future draws from the same urn (Coppersmith and Diaconis, 1987; Pemantle, 1988, 2007); ii) if $c \in E$ is the color of the sampled ball, then the next state of the process is $q(x, c)$, where $q : S \times E \to S$ is a known function, called the *law of motion* of the process, such that for every $x, y \in S$ there exists a unique $c(x, y) \in E$ satisfying $q(x, c(x, y)) = y$. For our purposes, the sequence of colors extracted from the urns will represent the history of a series of sequentially observed patients. The "reinforcement" of colors will then correspond to the notion of "learning from the past" that allows predictions to be performed and which is central in the Bayesian paradigm (Muliere et al., 2000, 2003; Bulla and Muliere, 2007; Peluso et al., 2015).

2. The subdistribution beta-Stacy process

Suppose that the random variable $T \in \{0, 1, 2, \ldots\}$ represents the time until an at-risk individual experiences some event of interest (e.g. time from seroconversion to AIDS onset for a HIV-infected man). If the distribution of T is unknown, then, in the Bayesian framework, it may be assigned a nonparametric prior to perform inference. In other words, it may be assumed that, conditionally on some random distribution function G defined on $\{0, 1, 2, \ldots\}$, T is distributed according to G itself: $P(T \leq t|G) = G(t)$ for all $t \geq 0$. Thus the random distribution function G assumes the role of a infinite-dimensional parameter, while its distribution corresponds to the nonparametric prior distribution. The *beta-Stacy process* of Walker and Muliere (1997) is one of such nonparametric priors which has received frequent use. Specifically, a random distribution function G on $\{0, 1, 2, \ldots\}$ is a discrete-time beta-Stacy process with parameters $\{(\beta_t, \gamma_t) : t \geq 1\}$, where

$$\lim_{t \to +\infty} \prod_{u=1}^{t} \frac{\gamma_k}{\beta_k + \gamma_k} = 0, \tag{1}$$

if: i) $G(0) = 0$ with probability 1 and ii) $\Delta G(t) = U_t \prod_{u=1}^{t-1}(1 - U_u)$ for all $t \geq 2$, where $\{U_t : t \geq 1\}$ is a sequence of independent random variables such that $U_t \sim Beta(\beta_k, \gamma_k)$ (Walker and Muliere, 1997). It can be shown that condition (1) is both necessary and sufficient for a random function $G(t)$ satisfying points i) and ii) to be a cumulative distribution function with probability one. The beta-Stacy process prior is conjugate with respect to right-censored data, a property that makes it especially suitable in survival analysis applications (Walker and Muliere, 1997).

To generalize this approach to competing risks, we introduce the following definitions:

Definition 2.1. *A function $F : \{0, 1, 2, \ldots\} \times \{1, \ldots, k\} \to [0, 1]$, $k \geq 1$, is called a (discrete-time)* subdistribution function *if it is the joint distribution function of some random vector $(T, \delta) \in \{0, 1, 2, \ldots\} \times \{1, \ldots, k\}$: $F(t, c) = P(T \leq t, \delta = c)$ for all $t \geq 0$ and $c \in \{1, \ldots, k\}$. A* random subdistribution function *is defined as a stochastic process indexed by $\{0, 1, 2, \ldots\} \times \{1, \ldots, k\}$ whose sample paths form a subdistribution function almost surely.*

Suppose now that T represents the time until one of k specific competing events occurs and that $\delta = 1, \ldots, k$ indicates the type of the occurring event. For instance, for $k = 2$, T may represent time from seroconversion to onset of either the SI phenotype or of non-SI AIDS, while δ may represent the specific event that occurred: $\delta = 1$ if the event is onset of the SI phenotype, $\delta = 2$ if the event is onset of non-SI AIDS. As before, if the distribution of (T, δ) is unknown, then in the Bayesian nonparametric framework it is assumed that, conditionally on some random subdistribution function F, (T, δ) is distributed according to F itself: $P(T \leq t, \delta = c|F) = F(t, c)$ for all $t \geq 0$ and $c = 1, \ldots, k$.

Remark 2.1. *Conditionally on F, $\Delta F(t, c) = F(t, c) - F(t-1, c)$ is the probability of experiencing an event of type c at time t: $\Delta F(t, c) = P(T = t, \delta = c|F)$. Additionally, if $G(t) = \sum_{d=1}^{k} F(t, k)$, $\Delta G(t) = G(t) - G(t-1)$, and $V_{t,k} = \Delta F(t, c)/\Delta G(t)$, then: $G(t) = P(T \leq t|F)$ is the cumulative probability of experiencing an event by time t, $\Delta G(t) =$*

$P(T = t|F)$ is the probability of experiencing an event at time t, and $V_{t,c} = P(\delta = c|T = t, F)$ is the probability of experiencing an event of type c at time t given that some event occurs at time t. Moreover, it can be shown that $F(t,c) = \sum_{u=1}^{t} S(t-1)\Delta A_c(t)$, where $S(t) = 1 - G(t)$ and $A_c(t) = \Delta F(t,c)/(1 - \sum_{d=1}^{k} F(t-1,d))$ is the cumulative hazard *of experiencing an event of type c by time t (Kalbfleisch and Prentice, 2002, Chapter 8).*

To specify a suitable prior on the random subdistribution function F, we now introduce the subdistribution beta-Stacy process:

DEFINITION 2.2. *Let $\{(\alpha_{t,0}, \ldots, \alpha_{t,k}) : t \geq 1\}$ be a collection of $(k+1)$-dimensional vectors of positive real numbers satisfying the following condition:*

$$\lim_{\tau \to +\infty} \prod_{t=1}^{\tau} \frac{\alpha_{t,0}}{\sum_{d=0}^{k} \alpha_{t,d}} = 0. \tag{2}$$

A random subdistribution function F is said to be a discrete-time subdistribution beta-Stacy *process with parameters $\{(\alpha_{t,0}, \ldots, \alpha_{t,k}) : t \geq 1\}$ if:*

(a) $F(0,c) = 0$ with probability 1 for all $c = 1, \ldots, k$;
(b) for all $c = 1, \ldots, k$ and all $t \geq 1$,

$$\Delta F(t,c) = W_{t,c} \prod_{u=1}^{t-1} \left(1 - \sum_{d=1}^{k} W_{u,d}\right),$$

with the usual convention that empty products are equal to 1, and where $\{W_t = (W_{t,0}, \ldots, W_{t,k}) : t \geq 1\}$ is a sequence of independent random vectors such that for all $t \geq 1$, $W_t \sim Dirichlet_{k+1}(\alpha_{t,0}, \ldots, \alpha_{t,k})$.

REMARK 2.2. *In Section 3, Remark 3.1, it will be shown that condition (2) is both necessary and sufficient for a random function $F(t,c)$ satisfying points 1 and 2 of Definition 2.2 to be a subdistribution function with probability 1. This justifies regarding the subdistribution beta-Stacy process as a potential prior distribution on the space of all subdistribution functions.*

LEMMA 2.1. *Let F be a subdistribution beta-Stacy process with parameters $\{(\alpha_{t,0}, \ldots, \alpha_{t,k}) : t \geq 1\}$. Then*

$$E[\Delta F(t,c)] = K_{t,c} \prod_{d=0}^{c-1} (1 - K_{t,d}) = \frac{\alpha_{t,c}}{\sum_{d=0}^{k} \alpha_{t,d}} \prod_{u=1}^{t-1} \frac{\alpha_{u,0}}{\sum_{d=0}^{k} \alpha_{u,d}}, \tag{3}$$

$$\mathrm{Var}\left(\Delta F(t,c)\right) = R_{t,c} \prod_{d=0}^{c-1} (1 - 2K_{t,d} + R_{t,d}) - E[\Delta F(t,c)]^2 \tag{4}$$

for all $t \geq 1$ *and* $c = 1, \ldots, k$, *where*

$$
\begin{aligned}
K_{t,0} &= 1 - \left(1 - \frac{\alpha_{t,0}}{\sum_{d=0}^{k} \alpha_{t,d}}\right) \prod_{u=1}^{t-1} \frac{\alpha_{u,0}}{\sum_{d=0}^{k} \alpha_{u,d}}, \\
K_{t,c} &= \frac{\alpha_{t,c}}{\sum_{d=c}^{k} \alpha_{t,d}}, \\
R_{t,0} &= (1 - K_{t,0}) \left(\frac{1 + \sum_{d=1}^{k} \alpha_{t,d}}{1 + \sum_{d=0}^{k} \alpha_{t,d}} \prod_{u=1}^{t-1} \frac{1 + \alpha_{u,0}}{1 + \sum_{d=0}^{k} \alpha_{u,d}} - 2 \right) + 1, \\
R_{t,c} &= K_{t,c} \frac{1 + \alpha_{t,c}}{1 + \sum_{d=c}^{k} \alpha_{t,d}}.
\end{aligned}
$$

PROOF. Using Theorem 2.5 of Ng et al. (2011) it is possible to show that the vector of random probabilities $(1 - \sum_{d=1}^{k} \Delta F(t,d), \Delta F(t,1), \ldots, \Delta F(t,k))$ is *completely neutral* in the sense of Connor and Mosimann (1969). Equations (3) and (4) can therefore be directly computed from Definition 2.2 using formulas (4) and (9) of Connor and Mosimann (1969).

DEFINITION 2.3. *Let* F_0 *be a fixed subdistribution function and let* $\omega_t > 0$ *for all* $t \geq 1$. *Then a random subdistribution function* F *has a* $s\mathcal{BS}(\omega, F_0)$ *distribution if it is a subdistribution beta-Stacy process with parameters*

$$
\alpha_{t,c} = \omega_t \Delta F_0(t,c), \quad \alpha_{t,0} = \omega_t \left(1 - \sum_{d=1}^{k} F_0(t,d)\right),
$$

where $t \geq 1$ *and* $c = 1, \ldots, k$.

REMARK 2.3. *If* $F \sim s\mathcal{BS}(\omega, F_0)$, *then condition (2) is automatically satisfied since* $\sum_{d=0}^{k} \alpha_{t,d} = \omega_t(1 - \sum_{d=1}^{k} F_0(t-1,d))$ *and so* $\prod_{t=1}^{+\infty} [\alpha_{t,0} / \sum_{d=0}^{k} \alpha_{t,d}] = \lim_{\tau \to +\infty}(1 - \sum_{d=1}^{k} F_0(\tau, d)) = 0$, *as* $\lim_{\tau \to +\infty} \sum_{d=1}^{k} F_0(\tau, d) = 1$ *(provided occurrence of at least one of the* k *events is inevitable). Moreover, it can be shown from Equation (3) in Lemma 2.1 that* $E[\Delta F(t,c)] = \Delta F_0(t,c)$ *for all* $t \geq 1$ *and* $c = 1, \ldots, k$, *implying that* F *is centered on* F_0. *From Equation (4) it can be further shown that* $\mathrm{Var}\,(\Delta F(t,c))$ *is a decreasing function of* ω_t *and* $\mathrm{Var}\,(\Delta F(t,c)) \to 0$ *as* $\omega_t \to +\infty$. *The parameters* ω_t *can thus be used to control the prior precision of the* $s\mathcal{BS}(\omega, F_0)$ *process.*

3. Predictive characterisation of the subdistribution beta-Stacy process

Muliere et al. (2000) described a predictive construction of the discrete-time beta-Stacy process by means of a reinforced urn process $\{Y_n : n \geq 0\}$ with state space $\{0, 1, 2, \ldots\}$. The urns of this process contain balls of only two colors, black and white (say), and reinforcement is performed by the addition of a single ball ($m = 1$). To intuitively describe this process, suppose that each patient in a series is observed from an initial time point until the onset of an event of interest. The process $\{Y_n : n \geq 0\}$ starts from $Y_0 = 0$, signifying the start of the observation for the first patient, and then

evolves as follows: if $Y_n = t$ and a black ball is extracted, then the current patient does not experience the event at time t and $Y_{n+1} = t + 1$; if instead a white ball is extracted, then the current patient experiences the event at time t and $Y_{n+1} = 0$, so the process is restarted to signify the start of the observation of a new patient. With this interpretation, the number T_n of states visited by $\{Y_n : n \geq 0\}$ between the $(n-1)$-th and n-th visits to the initial state 0 correspond to the time of event onset for the n-th patient. If the process $\{Y_n : n \geq 0\}$ is recurrent (so the times T_n are almost surely finite), a representation theorem for reinforced urn processes implies that the process $\{Y_n : n \geq 0\}$ is a mixture of Markov Chains. The corresponding mixing measure is such that the rows of the transition matrix are independent Dirichlet processes (Muliere et al., 2000, Theorem 2.16; see Ferguson, 1973 for the definition of a Dirichlet process). Using this representation, Muliere et al. (2000) showed that the sequence $\{T_n : n \geq 1\}$ is exchangeable and that there exists a random distribution function G such that i) conditionally on G, the times $T_1, T_2, \ldots$ are i.i.d. with common distribution function G, and ii) G is a beta-Stacy process (Muliere et al., 2000, Section 3).

In this section, we will generalize the predictive construction of Muliere et al. (2000) to yield a similar characterization of the subdistribution beta-Stacy process. To do so, consider a reinforced urn process $\{X_n : n \geq 0\}$ with state space $S = \{0, 1, 2, \ldots\} \times E$, set of colors $E = \{0, 1, \ldots, k\}$ $(k \geq 1)$, starting point $X_0 = (0, 0)$, and law of motion defined by $q((t, 0), c) = (t + 1, c)$ and $q((t, d), c) = (0, 0)$ for all for all integers $t \geq 0$ and $c, d = 0, 1, \ldots, k$, $d \neq 0$. Further suppose that reinforcement is performed by the addition of a single ball $(m = 1)$, as before, and that the initial composition of the urns is given as follows: i) $n_{(t,0)}(c) = \alpha_{t+1,c}$ for all integers $t \geq 0$ and $c = 0, 1, \ldots, k$; ii) $n_{(t,d)}(0) = 1$, $n_{(t,d)}(c) = 0$ for all integers $t \geq 0$ and $c, d = 1, \ldots, k$, $d \neq 0$. Now, define $\tau_0 = 0$ and $\tau_{n+1} = \inf\{t > \tau_n : X_t = (0, 0)\}$ for all integers $n \geq 0$. The process $\{X_n : n \geq 0\}$ is said to be *recurrent* if $P(\cap_{n=1}^{+\infty}\{\tau_n < +\infty\}) = 1$. Additionally, let $T((t, c)) = t$ and $D((t, c)) = c$ for all $(t, c) \in S$. For all $n \geq 1$, set $T_n = T(X_{\tau_n - 1})$, the length of the sequence of states between the $(n-1)$-th and the n-th visits to the initial state $(0, 0)$, and $D_n = D(X_{\tau_n - 1})$, the color of the last ball extracted before the n-th visit to $(0, 0)$.

The process $\{X_n : n \geq 0\}$ can be interpreted as follows: a patient initially at risk of experiencing any of k possible outcomes is followed in time starting from time $t = 0$; at each time point t, the color of the extracted ball represents the status of the patient at the next time point $t + 1$; if a ball of color 0 is extracted, the patient remains at risk at the next time point; if instead a ball of color $c \in \{1, \ldots, k\}$ is extracted, then the patient will experience an outcome of type c at the next time point. The process returns to the initial state after such an occurrence to signify the arrival of a new patient. With this interpretation, the variable T_n represents the time at which the n-th patient experiences one of the k events under study, while D_n encodes the type of the realized outcome. These concepts are illustrated in Figure 3. Moreover, note that, although slightly different, the reinforced urn process used to construct the beta-Stacy process by Muliere et al. (2000) is essentially equivalent to the process $\{X_n : n \geq 0\}$ in the particular case where $k = 1$, with color 0 being black and color 1 being white in the above description.

Continuing, in accordance with Diaconis and Freedman (1980) we say that the process $\{X_n : n \geq 0\}$ is *Markov exchangeable* if $P(X_0 = x_0, \ldots, X_n = x_n) = P(X_0 =$

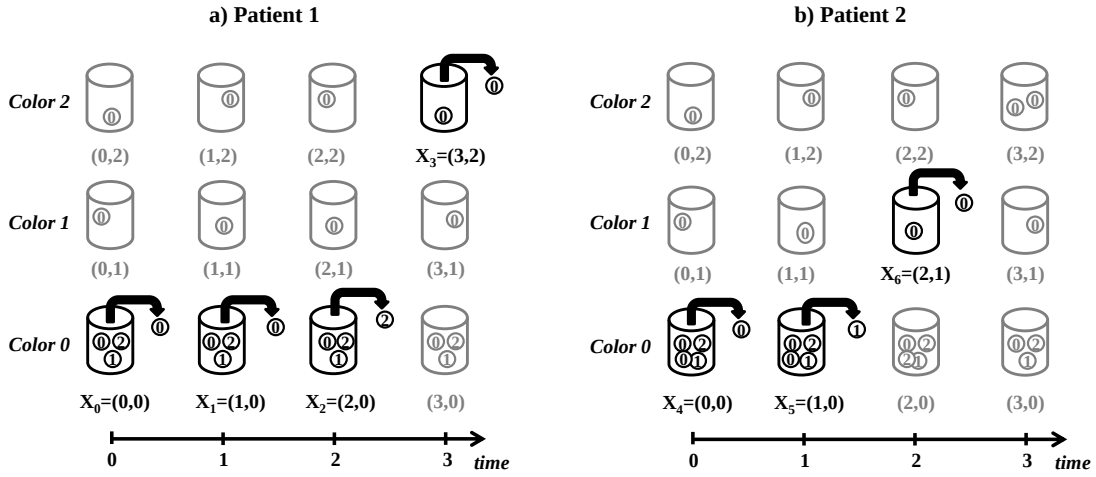

Fig. 1. Illustration of the reinforced urn process characterizing the subdistribution beta-Stacy process assuming $k = 2$. In both panels, the horizontal axis measures the time since the last visit to the urn representing the state $(0, 0)$. The process starts from the $(0, 0)$ urn in Panel a, in which all urns are represented at their initial composition. In this example, balls of colors 0, 0, and 2 are successively extracted from the urns visited by the process, respectively at times 0, 1, and 2. At time 3 the process visits the $(3, 2)$ urn, from which only balls of color 0 can be extracted. The process then returns to the $(0, 0)$ urn and continues as shown in Panel b, where the composition of the urns has been updated by reinforcement. Suppose now that each visit to $(0, 0)$ represents the arrival of a new HIV-infected man at the moment of seroconversion. If color 1 represents onset of the SI phenotype and color 2 represents onset of non-SI AIDS, then the sequence of urns visited in Panel a corresponds to the history of a man (Patient 1) who develops non-SI AIDS after 3 time instants ($T_1 = 3$, $D_1 = 2$), while Panel b represents the history of a subsequently observed man (Patient 2) who experiences onset of the SI phenotype after 2 time instants ($T_2 = 2$, $D_2 = 1$).

$y_0, \ldots, X_n = y_n)$ for all finite sequences $(x_0, \ldots, x_n)$ and $(y_0, \ldots, y_n)$ of elements of S such that i) $x_0 = y_0$ and ii) for any $s_1, s_2 \in S$, the number of transitions from s_1 to s_2 is the same in both sequences.

LEMMA 3.1. *The process $\{X_n : n \geq 0\}$ is Markov exchangeable. Consequently, if $\{X_n : n \geq 0\}$ is recurrent, then it is also a mixture of Markov Chains with state space S. In other words, there exists a probability measure μ on the space $\mathcal{M}$ of all transition matrices on $S \times S$ and a $\mathcal{M}$-valued random element $\Pi \sim \mu$ such that for all $n \geq 1$ and all sequences $x_0, \ldots, x_n \in S$ with $x_0 = (0, 0)$,*

$$P(X_0 = x_0, \ldots, X_n = x_n | \Pi) = \prod_{i=0}^{n-1} \Pi(x_i, x_{i+1}),$$

where $\Pi(x, y)$ is the element on the x-row and y-th column of Π. Additionally, for each $x = (t, c) \in S$, let $\mathcal{N}_x(\cdot)$ be the measure on S (together with the Borel σ-algebra generated

by the discrete topology) which gives mass $n_{(t,c)}(d)$ *to* $q((t,c),d)$ *for all* $d = 0, 1, \ldots, k$, *and null mass to all other points in* S. *Then, the random probability measure* $\Pi(x, \cdot)$ *on* S *is a Dirichlet process with parameter measure* $\mathcal{N}_x(\cdot)$.

PROOF. The thesis follows immediately from Theorem 2.3 and 2.16 of Muliere et al. (2000) and Theorem 7 of Diaconis and Freedman (1980).

LEMMA 3.2. *The process* $\{X_n : n \geq 0\}$ *is recurrent if and only if* $\{(\alpha_{t,0}, \ldots, \alpha_{t,k}) : t \geq 1\}$ *satisfies condition (2).*

PROOF. First observe that

$$\begin{aligned} P(\tau_1 = +\infty) &= \lim_{n \to +\infty} P(\tau_1 > n) \\ &= \lim_{n \to +\infty} P(X_0 = (0,0), X_1 = (1,0), \ldots, X_{n-1} = (n-1, 0)) \\ &= \lim_{n \to +\infty} \prod_{t=0}^{n-1} \frac{n_{(t,0)}(0)}{\sum_{d=1}^{k} n_{(t,0)}(d)} \\ &= \lim_{n \to +\infty} \prod_{t=1}^{n} \frac{\alpha_{t,0}}{\sum_{d=0}^{k} \alpha_{t,d}}. \end{aligned}$$

Consequently, if $\{X_n : n \geq 0\}$ is recurrent, then $P(\tau_1 = \infty) = 0$ and so condition (2) must hold. Conversely, suppose that condition (2) is satisfied. Then $P(\tau_1 < +\infty) = 1$. By induction on $n \geq 1$, suppose that $P(\cap_{i=1}^{n} \{\tau_i < +\infty\}) = 1$. Then

$$P(\tau_{n+1} = +\infty) = \int_{\cap_{i=1}^{n} \{\tau_i < +\infty\}} P(\tau_{n+1} = +\infty | T_1, \ldots, T_n) dP.$$

Given $T_1, \ldots, T_n$, if $\tau_{n+1} = +\infty$ then the process must visit all states $(t, 0)$ with $t \geq 0$ starting from time τ_n. Since the states $(t, 0)$ for $t > L := \max(T_1, \ldots, T_n) + 1$ correspond to previously unvisited urns, the probability of this event is bounded above by

$$\lim_{n \to +\infty} \prod_{i=L}^{n} \frac{n_{(i,0)}(0)}{\sum_{d=1}^{k} n_{(i,0)}(d)} = \lim_{n \to +\infty} \prod_{i=L+1}^{n} \frac{\alpha_{i,0}}{\sum_{d=1}^{k} \alpha_{i,d}}.$$

Hence

$$P(\tau_{n+1} = +\infty) \leq \int_{\cap_{i=1}^{n} \{\tau_i < +\infty\}} \lim_{n \to +\infty} \prod_{i=L+1}^{n} \frac{\alpha_{i,0}}{\sum_{d=1}^{k} \alpha_{i,d}} dP = 0,$$

where the last equality follows from condition (2). Consequently, $P(\cap_{i=1}^{n+1} \{\tau_i < +\infty\}) = 1$. This argument shows that $P(\cap_{i=1}^{+\infty} \{\tau_i < +\infty\}) = 1$ and so the process must be recurrent, as needed.

THEOREM 3.1. *Suppose that the process* $\{X_n : n \geq 0\}$ *is recurrent. Then there exists a random subdistribution function* F, *such that, given* F, *then the* (T_n, D_n) *are i.i.d. distributed according to* F. *Moreover, i)* F *is determined as a function of the random transition matrix* Π *from Lemma 3.1, and ii)* F *is a subdistribution beta-Stacy process with parameters* $\{(\alpha_{t,0}, \ldots, \alpha_{t,k}) : t \geq 1\}$.

PROOF. Let Π be the random transition matrix on $S \times S$ provided by Lemma 3.1 and define $F(t,c) = P(T_1 \leq t, D_1 = c|\Pi)$, which is clearly a random subdistribution function. Moreover, for all $c = 1, \ldots, k$,

$$F(0,c) = P(T_1 = 0, D_1 = c|\Pi) \leq P(T(X_{\tau_1 - 1}) = 1|\Pi) = P(\tau_1 = 1|\Pi) = 0.$$

Instead, for all $c = 1, \ldots, k$ and all $t \geq 1$,

$$\begin{aligned}\Delta F(t,c) &= P(T_1 = t, D_1 = c|\Pi) \\ &= P(X_0 = (0,0), \ldots, X_{t-1} = (t-1,0), X_t = (t,c)|\Pi) \\ &= \Pi((t-1,0),(t,c)) \prod_{u=0}^{t-2} \Pi((u,0),(u+1,0)).\end{aligned}$$

Now, for all $t \geq 1$ and $d = 0, 1, \ldots, k$, $\mathcal{N}_{(t-1,0)}(\{(t,d)\}) = \mathcal{N}_{(t-1,0)}(\{q((t-1,0),d)\}) = n_{(t-1,0)}(d) = \alpha_{t,d}$. Then, from Lemma 3.1 again and from the properties of the Dirichlet process (Ferguson, 1973), it follows that, for all $t \geq 1$, $\big(\Pi((t-1,0),(t,0)), \ldots, \Pi((t-1,0),(t,k))\big) \sim \mathit{Dirichlet}_{k+1}(\alpha_{t,0}, \ldots, \alpha_{t,k})$. Hence, Lemma 3.2 implies that F is a subdistribution beta-Stacy process with parameters $\{(\alpha_{t,0}, \ldots, \alpha_{t,k}) : t \geq 1\}$.

To show that, given F, the (T_n, D_n) are i.i.d. distributed according to F, it suffices to note that for all $(t_1,d_1), \ldots, (t_n,d_n) \in S$ such that $t_i \geq 1$ for all $i = 1, \ldots, n$, it holds that

$$\begin{aligned}&P\left((T_1,D_1) = (t_1,d_1), \ldots, (T_n,D_n) = (t_n,d_n)|\Pi\right) \\ &= \prod_{i=1}^{n} \left\{ \Pi((t_i-1,0),(t_i,d_i)) \prod_{t=0}^{t_i-1} \Pi((t,0),(t+1,0)) \right\} \\ &= \prod_{i=1}^{n} \Delta F(t_i,d_i).\end{aligned}$$

Since F is a function of Π, this concludes the proof.

REMARK 3.1. *Suppose that F is a random function satisfying points 1 and 2 of Definition 2.2. The proof of Theorem 3.1 also shows that, if condition (2) is satisfied, then F is a random subdistribution function. This is because condition (2) coincides with the recurrency condition in Lemma 3.2. Suppose instead that F is a subdistribution function with probability 1. Then $\tilde{F}(t,c) = E[F(t,c)]$ is an subdistribution function and*

$$P(T_1 \leq t, D_1 = c) = \tilde{F}(t,c) = \frac{\alpha_{t,d}}{\sum_{c=0}^{k} \alpha_{t,c}} \prod_{u=1}^{t-1} \frac{\alpha_{u,0}}{\sum_{c=0}^{k} \alpha_{u,c}}$$

for all $t \geq 0$ and $c = 1, \ldots, k$. Hence it must be

$$0 = P(T_1 = +\infty) = \lim_{t \to +\infty} P(X_0 = (0,0), \ldots, X_t = (t,0)) = \lim_{t \to +\infty} \prod_{u=1}^{t} \frac{\alpha_{u,0}}{\sum_{c=0}^{k} \alpha_{u,c}}.$$

Thus condition (2) must hold. Therefore, condition (2) is both necessary and sufficient for F to be a random subdistribution function, justifiying the claim anticipated in Remark 2.2 of Section 2.

Another immediate consequence of Theorem 3.1 is the following corollary:

COROLLARY 3.1. *The sequence of random variables* $\{(T_n, D_n) : n \geq 1\}$ *induced by the reinforced urn process* $\{X_n : n \geq 0\}$ *is exchangeable.*

This fact could also have been proved directly through an argument similar to that at the end of Section 2 of Muliere et al. (2000). To elaborate, suppose that $\{Y_n : n \geq 0\}$ is a recurrent stochastic process with countable state space S and such that $X_0 = x_0 \in S$ with probability one. Then a x_0-*block* is defined as any finite sequence of states visited by process which begins from x_0 and ends at the state immediately preceding the successive visit to x_0. Diaconis and Freedman (1980) showed that if $\{Y_n : n \geq 0\}$ is also Markov exchangeable, then the sequence $\{B_n : n \geq 1\}$ of its x_0-blocks is exchangeable. Now, consider the reinforced urn process $\{Y_n : n \geq 0\}$ used by Muliere et al. (2000) for constructing the beta-Stacy process and described at the beginning of this section. This process is Markov exchangeable and so, under a recurrency condition, its sequence of 0-blocks $\{B_n : n \geq 1\}$ is exchangeable. Consequently, so must be the corresponding sequence of total survival times $\{T_n = f(B_n) : n \geq 1\}$, where $f(B)$ is the length of the 0-block B after excluding its initial element. Each 0-block B_n must have the form $(0, 1, \ldots, t)$ for some $t \geq 1$ and $f((0, 1, \ldots, t)) = t$ for all $t \geq 1$.

In our setting, it can easily be seen that the $(0, 0)$-blocks of the reinforced urn process $\{X_n : n \geq 0\}$ introduced in this section are finite sequences of states of the form $((0, 0), (1, 0), \ldots, (t - 1, 0), (t, c))$ for some $t \geq 1$ and $c = 1, \ldots, k$. Any such $(0, 0)$-block represents the entire observed history of an individual at risk of developing any one of the k considered competing risks. For example, the history of Patient 1 in Figure 3(a) is represented by the $(0, 0)$-block $B_1 = ((0, 0), (1, 0), (2, 0), (3, 2))$, while that of Patient 2 in Figure 3(b) is represented by the $(0, 0)$-block $B_2 = ((0, 0), (1, 0), (2, 1))$. If $\{X_n : n \geq 0\}$ is recurrent, by Lemma 3.1 its sequence of $(0, 0)$-blocks $\{B_n : n \geq 1\}$ is exchangeable. Hence, so must be the sequence $\{(T_n, D_n) = f(B_n) : n \geq 1\}$, as claimed, where $f(B)$ is the last state in the $(0, 0)$-block B. For the example in Figure 3, $f(B_1) = (T_1, D_1) = (3, 2)$ and $f(B_2) = (T_2, D_2) = (2, 1)$.

4. Posterior distributions and censoring

Suppose that (T_i, D_i) is distributed according to some subdistribution function F and $T_i > 0$ with probability 1 for all $i = 1, \ldots, n$. If the value (T_i, D_i) can be potentially right-censored at the known time $c_i \in \{0, 1, 2, \ldots\} \cup \{+\infty\}$, then instead of observing the actual value (T_i, D_i) one is only able to observed (T_i^*, D_i^*), where $(T_i^*, D_i^*) = (T_i, D_i)$ if $T_i \leq c_i$ and $(T_i^*, D_i^*) = (c_i, 0)$ if $T_i > c_i$ (if $c_i = +\infty$, then (T_i, D_i) is not affected by censoring). The following theorem shows that the subdistribution beta-Stacy process has a useful conjugacy property even in presence of such right-censoring mechanism.

THEOREM 4.1. *Suppose that* (T_1, D_1), ..., (T_n, D_n) *is an i.i.d. sample from a subdistribution function* F *distributed as a subdistribution beta-Stacy process with parameters* $\{(\alpha_{t,0}, \ldots, \alpha_{t,k}) : t \geq 1\}$. *If* $(T_1, D_1), \ldots, (T_n, D_n)$ *are potentially right-censored at the known times* $c_1, \ldots, c_n$, *respectively, then the posterior distribution of* F *given* (T_1^*, D_1^*),..., (T_n^*, D_n^*) *is a subdistribution beta-Stacy with parameters* $\{(\alpha_{t,0}^*, \ldots, \alpha_{t,k}^*) :$

$t \geq 1\}$, *where* $\alpha^*_{t,0} = \alpha_{t,0} + l_t + m_{t,0}$, $\alpha^*_{t,d} = \alpha_{t,d} + m_{t,d}$ *for all integers* $t \geq 1$ *and for* $d = 1, \ldots, k$, *where* $l_t = \sum_{i=1}^n I\{T_i^* > t\}$ *and* $m_{t,d} = \sum_{i=1}^n I\{T_i^* = t, D_i = d\}$ *for all* $t \geq 1$ *and* $d = 0, 1, \ldots, k$.

PROOF. To prove the thesis, it suffices it is true for $n = 1$, as the general case will then follow from an immediate induction argument. To do so, first note that, with reference to the renforced urn process $\{X_n : n \geq 0\}$ of Section 3, condition (2) implies that F can be seen as a function of some random transition matrix Π as in the proof of Theorem 3.1. Assume now that $(T_1^*, D_1^*) = (t, d)$ for some $t \geq 1$ and $d = 0, 1, \ldots, k$. Since observing (T_1^*, D_1^*) is equivalent to observing $X_0 = (0,0), \ldots, X_{t-1} = (t-1, 0), X_t = (t, d)$, Corollary 2.21 of Muliere et al. (2000) implies that the rows of Π are independent and, for all $x \in S$, the parameter measure of the x-th row of Π assigns mass $n_{(0,0)}(0)+1, \ldots, n_{(t-2,0)}(0)+1, n_{(t-1,0)}(d)+1$ to the states $(1,0), \ldots, (t-1,0), (t,d)$, respectively, and mass $n_{(t',d')}(c)$ to all other states $q((t',d'),c) \neq (1,0), \ldots, (t-1,0), (t,d)$ in S. Since $\alpha_{t,d} = n_{(t-1,0)}(d)$ for all $t \geq 1$ and $d = 0, 1, \ldots, k$, it can now be seen that, conditionally on (T_1^*, D_1^*), F must be subdistribution beta-Stacy with parameters $\{(\alpha^*_{t,0}, \ldots, \alpha^*_{t,k}) : t \geq 1\}$ defined by $\alpha^*_{t,0} = \alpha_{t,0} + I\{T_1^* > t\} + I\{T_1^* = t, D_1^* = 0\}$, $\alpha^*_{t,d} = \alpha_{t,d} + I\{T_1^* = t, D_1^* = d\}$ for all integers $t \geq 1$ for $d = 1, \ldots, k$. This concludes the proof.

The following corollary is now a direct consequence of Equation (3) in Lemma 2.1.

COROLLARY 4.1. *The predictive distribution of a new (non-censored) observation* (T_{n+1}, D_{n+1}) *from* F *having previously observed* $(T_1^*, D_1^*), \ldots, (T_n^*, D_n^*)$ *is*

$$\begin{aligned} P((T_{n+1}, D_{n+1}) = (t,d) | (T_1^*, D_1^*), \ldots, (T_n^*, D_n^*)) &= \\ &= E\left[\Delta F(t,d) | (T_1^*, D_1^*), \ldots, (T_n^*, D_n^*)\right] \\ &= \frac{\alpha^*_{t,d}}{\sum_{c=0}^k \alpha^*_{t,c}} \prod_{u=1}^{t-1} \frac{\alpha^*_{u,0}}{\sum_{c=0}^k \alpha^*_{u,c}}. \end{aligned}$$

for all $t \geq 1$ *and* $d = 1, \ldots, k$.

The following result instead follows from Corollary 4.1 and Remark 2.3.

COROLLARY 4.2. *Assume that* $F \sim s\mathcal{BS}(\omega, F_0)$ *a priori. Then, the posterior distribution of* F *given the observed values of* $(T_1^*, D_1^*), \ldots, (T_n^*, D_n^*)$ *is* $s\mathcal{BS}(\omega^*, F^*)$, *where*

$$\begin{aligned} F^*(t,c) &= \sum_{u=1}^t S^*(u-1) \Delta A_c^*(u), \\ A_c^*(t) &= \sum_{u=1}^t \frac{\omega_u \Delta F_0(u,c) + m_{u,c}}{\omega_u (1 - \sum_{d=1}^k F_0(u-1,d)) + l_u + \sum_{d=0}^k m_{u,d}}, \\ S^*(t) &= \prod_{u=1}^t \left(1 - \frac{\omega_u \sum_{d=1}^k \Delta F_0(u,d) + \sum_{d=1}^k m_{u,d}}{\omega_u (1 - \sum_{d=1}^k F_0(u-1,d)) + l_u + \sum_{d=0}^k m_{u,d}}\right), \end{aligned}$$

and

$$\omega_t^* = \frac{\omega_t \left[1 - \sum_{d=1}^k F_0(t,d)\right] + l_t + m_{t,0}}{1 - \sum_{d=1}^k F^*(t,d)}.$$

REMARK 4.1. *As* $\max_{u=1,\dots,t}(\omega_u) \to 0$, $S^*(t)$ *converges to the discrete-time Kaplan-Meier estimate* $\widehat{S}(t) = \prod_{u=1}^{t}(1 - [\sum_{d=1}^{k} m_{u,d}]/[l_u + \sum_{d=0}^{k} m_{u,d}])$, *while* $A_c^*(t)$ *converges to the Nelson-Aalen estimate* $\widehat{A}_c(t) = \sum_{u=1}^{t} m_{u,c}/(l_u + \sum_{d=0}^{k} m_{u,d})$. *All in all,* $F^*(t,c)$, *which coincides with the optimal Bayesian estimate of* F *under a squared-error loss, converges to* $\widehat{F}(t,c) = \sum_{u=1}^{t} \widehat{S}(u-1)\Delta\widehat{A}_c(u)$, *the classical non-parametric estimate of* $F(t,c)$ *of Kalbfleisch and Prentice (2002, Chapter 8). Conversely, if* $\min_{u=1,\dots,t}(\omega_u) \to +\infty$, *then* $S^*(t)$ *converges to* $1 - \sum_{d=1}^{k} F_0(t,d)$, $A_c(t)$ *converges to the corresponding cumulative hazard of* F_0, *and therefore* $F^*(t,c)$ *converges to the prior mean* $F_0(t,c)$.

REMARK 4.2. *(Censored data likelihood) Given a sample* (t_1^*, d_1^*),..., (t_n^*, d_n^*) *of censored observations from a subdistribution function* $F(t,c)$, *define* $z_i = I\{d_i^* \neq 0\}$ *for all* $i = 1, \dots, n$. *It can then be shown that the likelihood function for* F *is*

$$\begin{aligned} L(F) &= P\left((T_1^*, D_1^*) = (t_1^*, d_1^*), \dots, (T_n^*, D_n^*) = (t_n^*, d_n^*)|F\right) \\ &= \prod_{i=1}^{n} \Delta F(t_i^*, d_i^*)^{z_i} \left[1 - \sum_{d=1}^{k} F(t_i^*, d)\right]^{1-z_i}. \end{aligned} \tag{5}$$

So far the censoring times $c_1, \dots, c_n$ have been considered fixed and known. Theorem 4.1 however continues to hold also in the following more general setting in which censoring times are random: let the censored data be defined as $T_i^* = \min(T_i, C_i)$ and $D_i^* = I\{T_i \leq C_i\}$ for all $i = 1, \dots, n$, where i) $C_1, \dots, C_n$ are independent random variable with common distribution function $H(t)$, ii) conditional on F and H, $(T_1, D_1), \dots, (T_n, D_n)$ and $C_1, \dots, C_n$ are independent, and iii) F and H are apriori independent. Adapting the terminology of Heitjan and Rubin (1991; 1993), in this case the random censoring mechanism is said to be *ignorable*.

THEOREM 4.2. *If censoring is random and ignorable and* F *is a priori a subdistribution beta-Stacy process, then the marginal likelihood for* F *is proportional to the likelihood* $L(F)$ *defined in Equation (5). Consequently, the posterior distribution of* F *given* (T_1^*, D_1^*), ..., (T_n^*, D_n^*) *is the same as that described in Theorem 4.1.*

PROOF. The likelihood function for F and H given a sample (t_1^*, d_1^*),..., (t_n^*, d_n^*) of observations affected from ignorable random censoring is

$$\begin{aligned} L^*(F,H) &= P\left((T_1^*, D_1^*) = (t_1^*, d_1^*), \dots, (T_n^*, D_n^*) = (t_n^*, d_n^*)|F, H\right) \\ &= L(F) \prod_{i=1}^{n} \Delta H(t_i^*)^{1-z_i} \left[1 - H(t_i^*)\right]^{z_i} \\ &= L(F) L^*(H), \end{aligned}$$

where L and the z_i are defined as in Equation 5. Therefore, the marginal likelihood for F is $L^{\text{marginal}}(F) = L(F) E_H[L^*(H)] \propto L(F)$, where the constant of proportionality only depends on the data and $E_H[\cdot]$ represents expectation with respect to the prior distribution of H. As a consequence, the posterior distribution of F can be computed ignoring the randomness in the censoring times $C_1, \dots, C_n$ by considering their observed values as fixed and their unobserved values as fixed to $+\infty$. Hence, if F is a priori

a subdistribution beta-Stacy process, then its posterior distribution is the same as in Theorem 4.1.

REMARK 4.3. *The update-rule of Theorem 4.1 could be shown to hold under even more general censoring mechanisms. In fact, the marginal likelihood for F remains proportional to $L(F)$ as long as i) the distribution H of censoring times is independent of F and ii) censoring only depends on the past and outside variation (Kalbfleisch and Prentice, 2002).*

5. Relation with other prior processes

5.1. Relation with the beta-Stacy process

By construction, the subdistribution beta-Stacy process can be regarded as a direct generalization of the beta-Stacy process. In fact, the two processes are linked with each other, as highlighted by the following theorem:

THEOREM 5.1. *A random subdistribution function F is a discrete-time subdistribution beta-Stacy process with parameters $\{(\alpha_{t,0}, \ldots, \alpha_{t,k}) : t \geq 1\}$ if and only if i) $G(t) = \sum_{d=1}^{k} F(t, d)$ is a discrete-time beta-Stacy process with parameters $\{(\sum_{d=1}^{k} \alpha_{t,d}, \alpha_{t,0}) : t \geq 1\}$ and ii) $\Delta F(t, c) = V_{t,c} \Delta G(t)$ for all $t \geq 1$ and $c = 1, \ldots, k$, where $\{V_t = (V_{t,1}, \ldots, V_{t,k}) : t \geq 1\}$ is a sequence of independent random vectors independent of G and such that $V_t \sim Dirichlet_k(\alpha_{t,1}, \ldots, \alpha_{t,k})$ for all $t \geq 1$ (where, if $k = 1$, we let the distribution $Dirichlet_1(\alpha_{t,1})$ be the point mass at 1).*

PROOF. The proof is provided in the on-line supplementary material.

5.2. Relation with the beta process

Suppose $A(t) = (A_1(t), \ldots, A_k(t))$ collects the cumulative hazards of the subdistribution function $F(t, c)$ and let $\Delta A(t) = (\Delta A_1(t), \ldots, \Delta A_k(t))$, $A_0(t) = \sum_{d=1}^{k} A_d(t)$. Then, following Hjort (Hjort, 1990, Section 2), a discrete time beta-process prior for non-homogeneous Markov Chains with parameters $\{(\alpha_{t,0}, \ldots, \alpha_{t,k}) : t \geq 1\}$ could be specified for $A(t)$ by independently letting $(1 - \Delta A_0(t), \Delta A_1(t), \ldots, \Delta A_k(t))$ have a $Dirichlet(\alpha_{t,0}, \ldots, \alpha_{t,k})$ distribution for all $t \geq 1$. In such case, from Definition 2.2 it would follow that F is subdistribution beta-Stacy with the same set of parameters. The converse is also true, since if F is subdistribution beta-Stacy then it can be easily seen from Definition 2.2 that $(1 - \Delta A_0(t), \Delta A_1(t), \ldots, \Delta A_k(t)) = (W_{t,0}, W_{t,1}, \ldots, W_{t,k})$. Therefore, if interest is in the subdistribution function $F(t, c)$ itself, one should consider the subdistribution beta-Stacy process, whereas if interest is in the cumulative hazards $A(t)$, one should consider the beta process for non-homogeneous Markov Chains. This equivalence parallels an analogous relation between the usual beta-stacy and beta processes (Walker and Muliere, 1997).

5.3. Relation with the beta-Dirichlet process

The subdistribution beta-Stacy process is also related to the discrete-time version of the *beta-Dirichlet* process, a generalization of Hjort's beta process prior (Hjort, 1990)

introduced by Kim et al. (2012). The cumulative hazards $\{A(t) : t \geq 1\}$ are said to be a beta-Dirichlet process with parameters $\{(\beta_{t,1}, \beta_{t,2}, \gamma_{t,1}, \ldots, \gamma_{t,k}) : t \geq 1\}$ if i) the $\Delta A(t)$ are independent, ii) $\Delta A_0(t) \sim Beta(\beta_{t,1}, \beta_{t,2})$ for all $t \geq 1$, and iii) $\Delta A(t)/\Delta A_0(t) \sim Dirichlet_k(\gamma_{t,1}, \ldots, \gamma_{t,k})$ independently of $\Delta A_0(t)$ for all $t \geq 1$. From Definition 2.2 it is clear that if $F(t, c)$ is subdistribution beta-Stacy with parameters $\{(\alpha_{t,0}, \ldots, \alpha_{t,k}) : t \geq 1\}$, then from $(1 - \Delta A_0(t), \Delta A_1(t), \ldots, \Delta A_k(t)) = (W_{t,0}, W_{t,1}, \ldots, W_{t,k})$ and Theorem 2.5 of Ng et al. (2011), then the corresponding cumulative hazards $A(t)$ must be beta-Dirichlet with parameters $\beta_{t,1} = \sum_{d=1}^{k} \alpha_{t,d}$, $\beta_{t,2} = \alpha_{t,0}$, and $\gamma_{t,d} = \alpha_{t,d}$ for all $d = 1, \ldots, k$ and $t \geq 1$. The converse is not true unless $\beta_{t,1} = \sum_{d=1}^{k} \gamma_{t,d}$ for all $t \geq 1$.

6. Nonparametric cumulative incidence regression

In this section, we will illustrate a hierarchical subdistribution beta-Stacy regression approach for competing risks. We assume that data are a sample of possibly-right censored survival times and cause-of-failure indicators (t_1^*, d_1^*), ..., (t_n^*, d_n^*). Each observation (t_i^*, d_i^*) is associated with a known vector w_i of predictors. In this context, we assume events occur in continuous time, but their times of occurrence have been discretized according to some fixed partition $0 = \tau_0 < \tau_1 < \tau_2 < \cdots$ of the time axis. Hence, $(t_i^*, d_i^*) = (t, d)$ for some $d = 1, \ldots, k$ if an event of type d has been observed in the time interval $(\tau_{t-1}, \tau_t]$. Instead, $(t_i^*, d_i^*) = (t, 0)$ if no event has been observed during $(\tau_{t-1}, \tau_t]$ and censoring took place in the same interval.

We specify the model hierarchically as follows. The observations $(t_1^*, d_1^*), \ldots, (t_n^*, d_n^*)$ are independent and each generated by some corresponding subdistribution function $F(t, c; w_i)$ under some censoring mechanism, as described in Section 4. If $w_{(1)}, \ldots, w_{(m)}$ are the distinct values of $w_1, \ldots, w_n$, then the subdistribution functions $F(\cdot; w_{(i)})$ are also assumed to be independent and $F(\cdot; w_{(i)}) \sim s\mathcal{BS}(\omega, F_0(\cdot|\theta, w_{(i)}))$ for all $i = 1, \ldots, m$, where $F_0(t, c|\theta, w_i)$ is some fixed parametric subdistribution function. Lastly, the parameter vector θ is also assigned a prior distribution.

In more detail, we specify the centering parametric subdistribution functions

$$F_0(t, c|\theta, w_i) = F_0^{(1)}(c|\theta, w_i) F_0^{(2)}(t|\theta, c, w_i)$$

using the same strategy of Larson and Dinse (1985), i.e. by separately modeling the probability $F_0^{(1)}(c|\theta, w_i)$ of observing a failure of type c and the conditional time-to-event distribution $F_0^{(2)}(t|\theta, c, w_i)$ given the specific failure type c. For the first, we specify a multinomial logistic regression model

$$F_0^{(1)}(c|\theta, w_i) = \frac{\exp(w_i' b_c)}{1 + \sum_{d=1}^{k-1} \exp(w_i' b_d)} \quad (c = 1, \ldots, k-1),$$
$$F_0^{(1)}(k|\theta, w_i) = \frac{1}{1 + \sum_{d=1}^{k-1} \exp(w_i' b_d)},$$

while for the second we let $F_0^{(2)}(t|\theta, c, w_i)$ be defined by

$$\Delta F_0^{(2)}(t|\theta, c, w_i) = G_0(\tau_t|\theta, c, w_i) - G_0(\tau_{t-1}|\theta, c, w_i),$$

where

$$G_0(t|\theta, c, w_i) = 1 - \exp(-t^{u_c} \exp(w_i' v_c))$$

is the cumulative distribution function of a Weibull random variable with scale parameter $\exp(-w_i' v_c / u_c)$ and shape parameter u_c. This choice corresponds to assuming a parametric Weibull regression model (Aalen et al., 2008, Chapter 5) for the conditional times to event (conditional with respect to the type of the occurring event). The parameter vector θ is thus so formed: $\theta = (b_1, \ldots, b_{k-1}, v_1, \ldots, v_k, u_1, \ldots, u_k)$.

For the precision parameters ω_t we let

$$\omega_t = \frac{\tau_t - \tau_{t-1}}{\sum_{d=1}^k F_0(\tau_t, d|\theta, w_i) - \sum_{d=1}^k F_0(\tau_{t-1}, d|\theta, w_i)}.$$

Similarly to Rigat and Muliere (2012), this choice allows the model to rely on its parametric component over the times where observations are not available, whereas it should allow for more flexibility over the times where most data is observed.

Finally, we specify a prior distribution for θ as follows. First, we assume all of $b_1, \ldots, b_{k-1}, u_1, v_1, \ldots, u_k, v_k$ to be independent a priori. For the regression parameters b_c and v_c we specify multivariate normal distributions: $b_c \sim N(b_{c,0}, \Sigma_{b_c})$ $(c = 1, \ldots, k-1)$ and $v_c \sim N(v_{c,0}, \Sigma_{v_c})$ $(c = 1, \ldots, k)$ for fixed vectors of Gaussian means $b_{c,0}$, $v_{c,0}$ and variance-covariance matrices Σ_{b_c}, Σ_{v_c} (each of the appropriate dimension). For the Weibull shape parameters u_c we instead specify $u_c \sim Gamma(p_c, q_c)$ $(c = 1, \ldots, k)$ for fixed positive constants p_c (shape parameter) and q_c (rate parameter).

6.1. Sampling from the posterior distribution

To fix notations, let $t^* = (t_1^*, \ldots, t_n^*)$, $d^* = (d_1^*, \ldots, d_n^*)$, $w = (w_1, \ldots, w_n)$, and $\mathcal{F} = (F(\cdot; w_{(1)}), \ldots, F(\cdot; w_{(m)}))$. With these notations, from Lemma 2.1 and the arguments in Section 4, under ignorable right censoring we can assume that the marginal likelihood of θ is

$$P(t^*, d^*|\theta, w) = \prod_{i=1}^n \left\{ \Delta F_0(t_i^*, d_i^*|\theta, w_i)^{z_i} \left[1 - \sum_{d=1}^k F_0(t_i^*, d|\theta, w_i) \right]^{1-z_i} \right\}.$$

Using this fact, the joint posterior distribution $P(\mathcal{F}, \theta|t^*, d^*, z)$ of $\mathcal{F}$ and θ is

$$P(\mathcal{F}, \theta|t^*, d^*, w) \propto P(\theta) P(t^*, d^*|\theta, w) \prod_{j=1}^m P_j(F(\cdot; w_{(j)})|\theta, w), \tag{6}$$

where $P(\theta)$ represents the prior distribution of θ (which is independent of w) and the term $P_j(F(\cdot; w_{(j)})|\theta, w)$ represents the posterior distribution of the subdistribution $F(\cdot; w_{(j)}) \sim s\mathcal{BS}(\omega, F_0(\cdot|\theta, w_{(j)}))$ obtained (for fixed θ) from the data $\mathcal{D}_j = \{(t_i^*, d_i^*) : w_i = w_{(j)}, i = 1, \ldots, n\}$ using the update rule described in Theorem 4.1.

Now, although the posterior distribution for θ is not available for exact sampling, Equation (6) suggests the use of a Markov Chain Monte Carlo strategy such as the following to perform approximate posterior inferences. First, a sample $\{\theta_i\}_{i=1}^S$ from the

marginal posterior distribution of θ is obtained, after discarding an appropriate number of burn-in iterations, via a Random Walk Metropolis-Hastings algorithm (Robert and Casella, 2004, Section 7.5). A multivariate Gaussian distribution can be considered after the reparametrization induced by a logarithmic transformation of each shape parameter u_c (to account for their positive support). Second, having obtained a sample $\{\theta_i\}_{i=1}^S$ as just described, the conditional posterior distribution of $F(\cdot; w_{(j)})$ given θ_i and the data $\mathcal{D}_j$ is obtained by direct simulation for all $i = 1, \ldots, M$ and $j = 1, \ldots, m$. Specifically, the parameters of the conditional posterior distribution $P_j(F(\cdot; w_{(j)})|\theta, w)$ of $F(\cdot; w_{(j)})$ given θ_i and $\mathcal{D}_j$ are obtained using Theorem 4.1. Then a sample $F_i(\cdot; w_{(j)})$ from $P_j(F(\cdot; w_{(j)})|\theta, w)$ is obtained using Definition 2.2 by sampling from the relevant Dirichlet distributions. The sample $\{(\theta_i, F_i(\cdot; w_{(1)}), \ldots, F_i(\cdot; w_{(m)}))\}_{i=1}^S$ so obtained then represents a sample from the joint posterior distribution of Equation (6).

6.2. Estimating the predictive distributions

Let T_{n+1} and D_{n+1} be the unknown uncensored survival time and type of realized outcome, respectively, for a new individual with covariate profile w_{n+1}. The objective if to estimate the predictive distribution of (T_{n+1}, D_{n+1}) given the data (t_1^*, d_1^*), …, (t_n^*, d_n^*). We distinguish two cases: i) $w_{n+1} = w_{(j)}$ for some $j = 1, \ldots, m$, and ii) $w_{n+1} \neq w_{(1)}, \ldots, w_{(m)}$. In the first case, simply obtain a sample $\{F_i(\cdot; w_{n+1}) = F_i(\cdot; w_{(j)})\}_{i=1}^S$ from the posterior distribution of $F(\cdot; w_{(j)})$ using the output of the procedure described above. The predictive distribution of (T_{n+1}, D_{n+1}) is then estimated as $S^{-1} \sum_{i=1}^S F_i(\cdot; w_{n+1})$. In the second case it is still possible to estimate the predictive distribution of (T_{n+1}, D_{n+1}) by recycling the sample $\{\theta_i\}_{i=1}^S$. Specifically, for each θ_i, $F_i(\cdot; w_{n+1})$ is simulated directly from the $s\mathcal{BS}(\omega, F_0(\cdot|\theta_i, w_{n+1}))$ distribution. The predictive distribution of (T_{n+1}, D_{n+1}) is then estimated as the average of the sampled subdistribution functions, as before.

7. Application example: analysis of Amsterdam Cohort Studies data

7.1. Analysis objectives and data description

As an illustration, we consider the problem of assessing the long-term prognosis of HIV infected men with respect to the risk of AIDS onset (event of type 1) or SI onset (event of type 2), considered as competing events, as a function of two predictors: i) CCR5 genotype, with possible values *WW* (wild type allele on both chromosomes) or *WM* (mutant allele on one chromosome), or ii) age at HIV infection. With this aim, we analyse data on 324 men who participated in the Amsterdam Cohort Studies on HIV infection and AIDS. These data are freely available in the *aidssi2* dataset of the *mstate* R package (de Wreede et al., 2011) and are described in more detail by Geskus et al. (2003) and Putter et al. (2007). For our purposes, each patient was followed-up from the date of HIV infection to the earliest among the dates of AIDSs onset, SI phenotype onset, or right censoring, i.e. death, study drop-out, or end of the study period (July 2006, when highly-active antiretroviral therapy became widely available). A total of 65 (20.1%) patients had a WM genotype, while mean age at HIV infection was about 34.6 years (s.d., 7.2 years). Overall, the 324 patients accumulated 2262.2 person-years of

follow up (minimum - maximum follow-up: 0.1 years - 13.9 years), generating 117 cases of AIDS onset and 107 cases of SI onset.

7.2. Model specification and prior distributions

We consider two separate regression models, both specified as explained in Section 6. In the first model, only the indicator of WM CCR5 genotype is considered as predictor and are inserted in the model (together with an intercept term). In the second model, the indicator for WM CCR5 genotype, the age at HIV infection (centered about its mean and considered as a simple linear term), as well as an interaction term between CCR5 genotype and age at infection inserted in the model (again together with an intercept term). For all hyperparameters b_c, v_c, and u_c of these two models, we consider the following prior specifications. First, for all vectors of regression parameters b_c and v_c we specified independent $N(0, I)$ multivariate normal prior distributions, where I is the identity matrix of the appropriate dimensions. Second, for all the Weibull shape parameters u_c we instead consider independent $Gamma(2, 2)$ prior distributions. Numerical simulations reported in the on-line supplementary material suggest that these choices yield fairly diffuse prior distributions for the subdistribution function of both models.

7.3. Posterior analysis

For both models, posterior inference was performed by a Random Walk Metropolis-Hastings algorithm with a multivariate Gaussian proposal distribution as suggested in Section 6, by means of the *MCMCpack* R package Martin et al. (2011). The proposal distribution was centered at the current sampled value, with a proposal covariance matrix equal to the negative inverse Hessian matrix of the log-posterior distribution, evaluated at the posterior mode and scaled by $(2.4)^2/d$, where d is the dimension of θ, as suggested by Gelman and Meng (1998, Section 12.2). To improve mixing, all predictors were standardized before running the algorithm. All parameters were initialized at 0, except for each u_c, which were initialized at 1. In all cases, the Metropolis-Hastings algorithm was run for a total of 26000 iterations: the first 1000 were discarded as burn-in, while the remaining 25000 were thinned by retaining only one generated sample every 25 iterations. Convergence was assessed by means of Geweke's test (Geweke, 1992) and visual inspection of the trace plots.

The Markov Chain Monte Carlo algorithm converged fairly quickly with an acceptance rate of about 20% for both models. The estimated autocorrelation functions, trace plots, and posterior density estimates for both models, as well as posterior summaries and Geweke's diagnostic p-values for the second model are reported in the on-line supplementary material. The results were found to be fairly robust when different prior distributions were considered (data not shown). Additionally, the obtained posterior distributions were found to be much more concentrated than the considered prior distributions (see the on-line supplementary material). The R code used to perform the analyses is available on request from the authors.

The posterior summaries for the hyperparameters of the model including only the WM CCR5 genotype indicator as predictor are reported in the on-line supplementary material. These suggest that individuals with the WM genotype are at a lower risk of de-

veloping AIDS instead of SI at any time compared with individuals with WW genotype (estimated log-odds ratio and 95% credible interval: -0.51 $(-1.55, 0.48)$). Additionally, both among individuals who develop AIDS or SI, the WM genotype seems to be associated with a decreased instantaneous hazard of experiencing the onset of the disease, although for SI this finding is affected by a higher uncertainty (estimated log-hazard rates and 95% credible intervals: -0.52 $(-0.94, -0.08)$ for AIDS, -0.35 $(-0.85, 0.22)$ for SI). The estimated shape parameters indicate that for both AIDS and SI, the conditional time to disease onset distribution is characterized by an increasing instantaneous hazard (estimated shape parameters and 95% credible intervals: 2.03 $(1.70, 2.44)$ for AIDS, 1.43 $(1.17, 1.78)$ for SI).

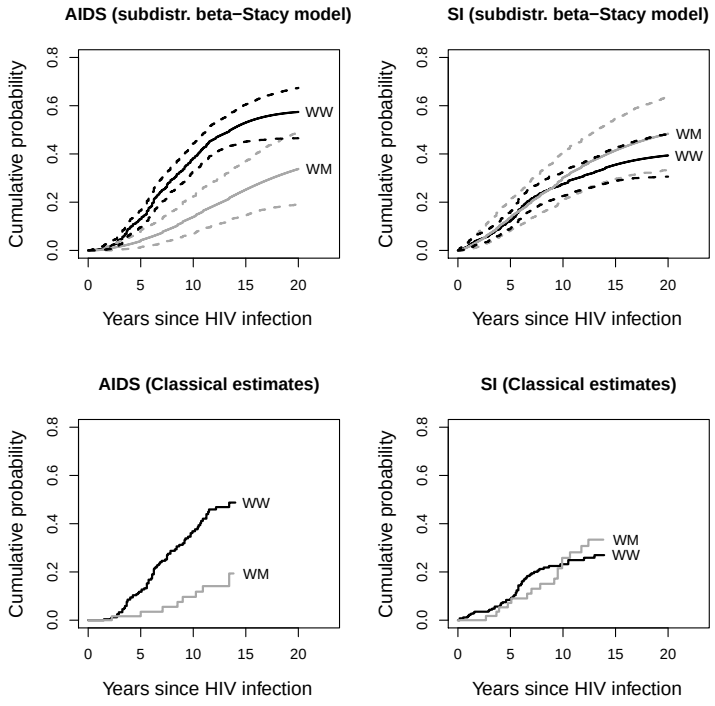

Fig. 2. Posterior predictive distributions (i.e. posterior expectations of the subdistribution functions) obtained from the subdistribution beta-Stacy regression model including the CCR5 genotype as predictor, with 95% pointwise credible bands for the subdistribution function (top), and classical nonparametric estimatates of the subdistribution functions (bottom).

Figure 2 shows the posterior predictive distributions, i.e. the posterior expectations of the subdistribution functions, obtained from the subdistribution beta-Stacy regression model including the CCR5 genotype as predictor, as well as the estimated subdistribution functions obtained from the classical nonparametric estimators. Both sets of estimates are in good agreement for all times up to about 14 years since HIV infection, i.e. the maximum observed time. This was to be expected, since in the range of the observed data the specified subdistribution model is mostly driven by the data through the classical nonparametric estimators. Note, however, that our estimates avoid the unreal-

istic shape characterizing the classical estimators, which are defined as constant over the time periods between successive uncensored observations. Furthermore, after the maximum observed time in the data, the classical nonparametric estimators are undefined, restricting their usefulness for evaluating the long-term prognosis. Instead, by relying more on its parametric component, the subdistribution beta-Stacy model provides an extrapolated prognostic risk estimate beyond the range of observed event times.

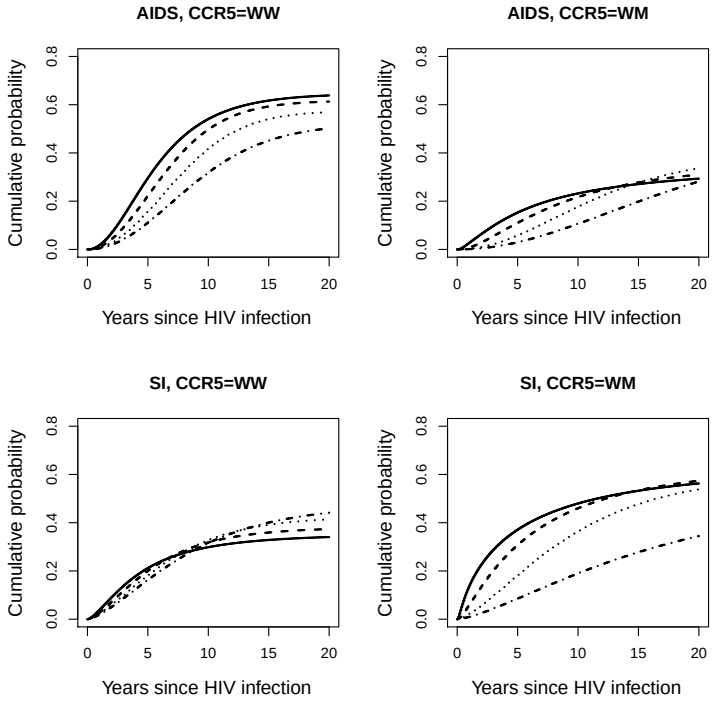

Fig. 3. Posterior predictive distributions obtained from the subdistribution beta-Stacy regression model including both the CCR5 genotype and age at HIV infection as predictors for four hypothetical patients characterized by the following ages at HIV infection: 20 years (dot-dashed curve), 40 years (dotted curve), 60 years (dashed curve), and 80 years (continuous curve).

Figure 3 shows the posterior predictive distributions obtained from the subdistribution beta-Stacy model including both the CCR5 genotype and age at HIV infection as predictors for four hypothetical patients respectively characterized by an age at HIV infection of 20, 40, 60, and 80 years. This analysis suggests that age at HIV infection may be an important prognostic factor for both individuals with WW or WM CCR5 genotype. In fact, the posterior predictive distributions suggest that time to AIDS onset or SI onset is delayed in younger HIV infected men compared with older infected men. Additionally, these results also highlight how prognosis may be widely heterogeneous depending on both CCR5 genotype and age at HIV infection.

8. Concluding remarks

The approach developed in this paper is amenable to several potential generalizations. First, other reinforced urn schemes could be contemplated. For example, each extracted ball may be reinforced by a general amount $m > 0$ of new similar balls, instead of just one as considered in this paper. More generally, the value m may also be allowed to be a random variable depending on the color of the extracted balls, as in Muliere et al. (2006). In general, urn-based schemes represent a useful tool to characterize different process priors on the space of subdistribution functions from a predictive point of view. From this perspective, we are currently investigating the predictive characterization of the beta-Dirichlet process prior described in Section 5. Second, a continuous-time generalization of the subdistribution beta-Stacy process could be considered. We are currently developing a characterization of such process from a predictive perspective by means of the continuous-time urn models of Muliere et al. (2003) and Bulla and Muliere (2007) in a similar way as done in this paper. Third and last, a generalization of the reinforced urn process considered in Section 3 could be attempted to characterize a process prior on the space of transition kernels of a Markovian multistate process. Such process could be useful for the predictive Bayesian nonparametric analysis of event-history data (Aalen et al., 2008) and it will be the subject of our future work.

References

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

Andersen, P. K., Abildstrom, S. Z. and Rosthøj, S. (2002) Competing risks as a multi-state model. *Statistical Methods in Medical Research*, **11**, 203–215.

Andersen, P. K., Borgan, O., Gill, R. D. and Keiding, N. (2012) *Statistical models based on counting processes*. New York: Springer Science & Business Media.

Bernardo, J. M. and Smith, A. F. (2000) *Bayesian theory*. Chichester, England: John Wiley & Sons.

Bulla, P. and Muliere, P. (2007) Bayesian nonparametric estimation for reinforced markov renewal processes. *Statistical Inference for Stochastic Processes*, **10**, 283–303.

Chae, M., Weißbach, R., Cho, K. H. and Kim, Y. (2013) A mixture of beta–dirichlet processes prior for bayesian analysis of event history data. *Journal of the Korean Statistical Society*, **42**, 313–321.

Cifarelli, D. M. and Regazzini, E. (1996) De Finetti's contribution to probability and statistics. *Statistical Science*, **11**, 253–282.

Connor, R. J. and Mosimann, J. E. (1969) Concepts of independence for proportions with a generalization of the dirichlet distribution. *Journal of the American Statistical Association*, **64**, 194–206.

Coppersmith, D. and Diaconis, P. (1987) Random walk with reinforcement. unpublished manuscript.

Crowder, M. J. (2012) *Multivariate survival analysis and competing risks*. Boca Raton, Florida: CRC Press.

De Blasi, P. and Hjort, N. L. (2007) Bayesian survival analysis in proportional hazard models with logistic relative risk. *Scandinavian Journal of Statistics*, **34**, 229–257.

De Finetti, B. (1964) Foresight: its logical laws, its subjective sources. (English translation of 1937 article in French, "La prévision: ses lois logiques, ses sources subjectives", Annales de l'institut Henri Poincaré, 7:1-68) in H. E. Kyburg and H. E. Smokler (editors), Studies in Subjective Probability. New York: Wiley.

Diaconis, P. and Freedman, D. (1980) De Finetti's theorem for Markov chains. *The Annals of Probability*, **8**, 115–130.

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

Epifani, I., Fortini, S. and Ladelli, L. (2002) A characterization for mixtures of semi-Markov processes. *Statistics & Probability Letters*, **60**, 445–457.

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

Fine, J. P. (1999) Analysing competing risks data with transformation models. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, **61**, 817–830.

Fine, J. P. and Gray, R. J. (1999) A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, **94**, 496–509.

Fortini, S. and Petrone, S. (2012) Predictive construction of priors in bayesian nonparametrics. *Brazilian Journal of Probability and Statistics*, **26**, 423–449.

Ge, M. and Chen, M.-H. (2012) Bayesian inference of the fully specified subdistribution model for survival data with competing risks. *Lifetime Data Analysis*, **18**, 339–363.

Gelman, A. and Meng, X.-L. (1998) Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. *Statistical Science*, **13**, 163–185.

Geskus, R. B., Miedema, F. A., Goudsmit, J., Reiss, P., Schuitemaker, H. and Coutinho, R. A. (2003) Prediction of residual time to aids and death based on markers and cofactors. *Journal of Acquired Immune Deficiency Syndromes*, **32**, 514–521.

Geweke, J. (1992) Evaluating the accuracy of sampling-based approaches to calcualting posterior moments. In *Bayesian Statistics 4* (eds. J. Bernardo, J. Berger, A. Dawid and S. A.F.M.), 169–193. Oxford, UK: Clarendon Press.

Heitjan, D. F. (1993) Ignorability and coarse data: Some biomedical examples. *Biometrics*, **49**, 1099–1109.

Heitjan, D. F. and Rubin, D. B. (1991) Ignorability and coarse data. *The Annals of Statistics*, **19**, 2244–2253.

Hinchliffe, S. R. and Lambert, P. C. (2013) Flexible parametric modelling of cause-specific hazards to estimate cumulative incidence functions. *BMC Medical Research Methodology*, **13**, 13.

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

Hjort, N. L., Holmes, C., Müller, P. and Walker, S. G. (2010) *Bayesian nonparametrics*. Cambridge, UK: Cambridge University Press.

Jeong, J.-H. and Fine, J. P. (2007) Parametric regression on cumulative incidence function. *Biostatistics*, **8**, 184–196.

Kalbfleisch, J. D. and Prentice, R. L. (2002) *The statistical analysis of failure time data*. Hoboken, New Jersey: John Wiley & Sons, 2nd edition edn.

Kim, Y., James, L. and Weissbach, R. (2012) Bayesian analysis of multistate event history data: beta-dirichlet process prior. *Biometrika*, **99**, 127–140.

Larson, M. G. and Dinse, G. E. (1985) A mixture model for the regression analysis of competing risks data. *Applied Statistics*, **34**, 201–211.

Lawless, J. F. (2011) *Statistical models and methods for lifetime data*. Hoboken, New Jersey: John Wiley & Sons.

Martin, A. D., Quinn, K. M. and Park, J. H. (2011) MCMCpack: Markov chain Monte Carlo in R. *Journal of Statistical Software*, **42**, 22. URL: `http://www.jstatsoft.org/v42/i09/`.

Mauldin, R. D., Sudderth, W. D. and Williams, S. (1992) Polya trees and random distributions. *The Annals of Statistics*, **20**, 1203–1221.

Muliere, P., Paganoni, A. M. and Secchi, P. (2006) A randomly reinforced urn. *Journal of Statistical Planning and Inference*, **136**, 1853–1874.

Muliere, P., Secchi, P. and Walker, S. (2000) Urn schemes and reinforced random walks. *Stochastic Processes and their Applications*, **88**, 59–78.

Muliere, P., Secchi, P. and Walker, S. G. (2003) Reinforced random processes in continuous time. *Stochastic Processes and their Applications*, **104**, 117–130.

Müller, P. and Mitra, R. (2013) Bayesian nonparametric inference–why and how. *Bayesian Analysis*, **8**, 269–302.

Ng, K. W., Tian, G.-L. and Tang, M.-L. (2011) *Dirichlet and related distributions: Theory, methods and applications*. Chichester, England: John Wiley & Sons.

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

Peluso, S., Mira, A. and Muliere, P. (2015) Reinforced urn processes for credit risk models. *Journal of Econometrics*, **184**, 1–12.

Pemantle, R. (1988) *Random processes with reinforcement.* Ph.D. thesis, Massachussets Institute of Technology.

— (2007) A survey of random processes with reinforcement. *Probabability Surveys*, **4**, 1–79.

Pintilie, M. (2006) *Competing risks: a practical perspective.* Chichester, England: John Wiley & Sons.

Putter, H., Fiocco, M. and Geskus, R. (2007) Tutorial in biostatistics: competing risks and multi-state models. *Statistics in Medicine*, **26**, 2389–2430.

Rigat, F. and Muliere, P. (2012) Nonparametric survival regression using the beta-stacy process. *Journal of Statistical Planning and Inference*, **142**, 2688–2700.

Robert, C. and Casella, G. (2004) *Monte Carlo Statistical Methods.* Springer Texts in Statistics. New York: Springer, 2nd edn.

Scheike, T. H., Zhang, M.-J. and Gerds, T. A. (2008) Predicting cumulative incidence probability by direct binomial regression. *Biometrika*, **95**, 205–220.

Singpurwalla, N. D. (1988) Foundational issues in reliability and risk analysis. *SIAM Review*, **30**, 264–282.

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

— (1999) A characterization of a neutral to the right prior via an extension of Johnson's sufficientness postulate. *The Annals of Statistics*, **27**, 589–599.

Wechsler, S. (1993) Exchangeability and predictivism. *Erkenntnis*, **38**, 343–350.

Wolbers, M., Koller, M. T., Witteman, J. C. and Steyerberg, E. W. (2009) Prognostic models with competing risks: methods and application to coronary risk prediction. *Epidemiology*, **20**, 555–561.

de Wreede, L. C., Fiocco, M., Putter, H. et al. (2011) mstate: an r package for the analysis of competing risks and multi-state models. *Journal of Statistical Software*, **38**, 1–30.