Statistica Sinica **17**(2007), 427-444

BAYESIAN NONPARAMETRIC ESTIMATION OF A BIVARIATE SURVIVAL FUNCTION

Paolo Bulla[1], Pietro Muliere[1] and Stephen Walker[2]

[1] *Università Bocconi and* [2] *University of Kent*

Abstract: We use a bivariate reinforced process derived from a Generalized Pólya Urn scheme to provide an estimator for a bivariate survival function. The estimator may be considered as a Bayesian nonparametric predictive estimator and it can be obtained easily via an implementation of a Gibbs sampler. Consistency of the estimator is studied.

Key words and phrases: Bayesian nonparametric, beta-Stacy process, bivariate survival function, Pólya urn scheme.

1. Introduction

A difficult problem in survival and reliability analysis is the estimation of a bivariate survival function for lifetimes subjected to censoring.

Such problems arise in studies where the experimental unit consists of couples of components (for example, twins, eyes, kidneys) or pairs of lifetimes are observed for the same individual (for example, response times for successive courses of a medical treatment).

Compared to the univariate problem, relatively less work has been done in the nonparametric and semiparametric areas; nevertheless, recently, some contributions have been made. Campbell (1981) and Hanley and Parnes (1983) studied self-consistent maximum likelihood estimators for discrete data. Campbell and Földes (1982) and Tsai, Leurgans and Crowley (1986) proposed estimators based on a decomposition of the bivariate survival function. Dabrowska (1988) and Pruitt (1991) extended the Kaplan-Meier estimator on the plane by means of a product integral representation for a bivariate cumulative hazard. Oakes (1989) exploited a frailty model, while Prentice and Cai (1992) used covariance between counting process martingales to characterize the dependence of the survival times. van der Laan (1996) and Wang and Wells (1997) adopted a non-parametric approach.

Most of the proposed estimators have drawbacks; some are not proper while others have no explicit form or are quite complicated to obtain in practice. Furthermore, efficiency results often hold only under restrictive conditions.

In this paper, using a reinforcement scheme, a model for coupled survival times is proposed. We work exclusively with discrete observations, and so with observation space $\mathbb{N}_0^2 = \{0, 1, \ldots\} \times \{0, 1, \ldots\}$, but in many applications this does not seem to represent a severe limitation.

Our approach can be considered as Bayesian predictive in that the main concern is to provide a sensible procedure to predict the future behaviour of a bivariate lifetimes when data having the same form were collected in the past. Moreover in the model, observations are exchangeable and, hence, define via de Finetti's Representation Theorem, a prior on the space of bivariate distributions. Although the prior has a structure which makes the posterior intractable, it is possible to study its support and obtain some consistency results. Inference will be done using sampling based methods.

Here we describe the layout of the paper. Section 2 recalls the definition and some features of Generalized Pólya Urn scheme. In Section 3 the construction of a bivariate reinforced process is carried out, a prior on the space of distribution functions on $\mathbb{N}_0^2$ is derived, and some properties are pointed out, with particular attention to its support and consistency. Section 4 shows how to use this prior to make inference about about a bivariate survival function with data eventually subjected to censorship. An example is given in Section 5.

2. Preliminaries

The basic building block of the model is the Generalized Pólya Urn (GPU) scheme. Here we recall the definition of Walker and Muliere (1997).

Definition 1. Consider a sequence of random variables $\{T_n;\, n \geq 1\}$ with values on the non-negative integers $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$, and let $\{\alpha_j, \beta_j,\, j \in \mathbb{N}_0\}$ satisfy

1. $\alpha_j, \beta_j \geq 0\, \forall\, j$;
2. $\alpha_j + \beta_j > 0\, \forall\, j$;
3. $\lim_{n\to\infty} \prod_{j=0}^{n} \beta_j/(\alpha_j + \beta_j) = 0$.

The sequence $\{T_n;\, n \geq 1\}$ is said to come from a GPU scheme if

$$P\left[T_1 = t\right] = \frac{\alpha_t}{\alpha_t + \beta_t} \prod_{j=0}^{t-1} \frac{\beta_j}{\alpha_j + \beta_j}$$

and

$$P\left[T_{n+1} = t \mid \mathbf{T}_n = \mathbf{t}_n\right] = \frac{\alpha_t + m_t\left(\mathbf{t}_n\right)}{\alpha_t + \beta_t + s_t\left(\mathbf{t}_n\right)} \prod_{j=0}^{t-1} \frac{\beta_j + r_j\left(\mathbf{t}_n\right)}{\alpha_j + \beta_j + s_j\left(\mathbf{t}_n\right)}, \quad (2.1)$$

where $m_j\left(\mathbf{t}_n\right) = \sum_{k=1}^{n} \mathbb{1}_{[t_k = j]}$, $r_j\left(\mathbf{t}_n\right) = \sum_{k=1}^{n} \mathbb{1}_{[t_k > j]}$ and $s_j\left(\mathbf{t}_n\right) = m_j\left(\mathbf{t}_n\right) + r_j\left(\mathbf{t}_n\right)$.

Henceforward bold letters indicate vectors, for example $\mathbf{T}_n = (T_1, \ldots, T_n)$; relations and operations between them are done componentwise.

These predictive distributions arise from a beta process model of Hjort (1990). Briefly, the model puts independent beta prior distributions on hazards at $t = 1, 2, \ldots$. That is, a random hazard at time t from the prior is given by $V_t \prod_{s<t}(1-V_s)$, with $V_s \sim$ beta (α_s, β_s). A reason for this name can be found in the following scheme to generate the sequence $\{T_n\}$.

Consider a Pólya urn U_j that has α_j white balls and β_j black balls, $j = 0, 1, \ldots$. A ball is drawn starting from U_0; if white put $T_1 = 0$, otherwise a ball is drawn from U_1 and so on, until a white ball is obtained. If the white ball is drawn from U_t, say, then put $T_1 = t$. All the urns are updated according to the traditional Pólya reinforcement rule - the drawn ball is reintroduced in the corresponding urn along with another of the same color. So, $T_2, T_3, \ldots$ are generated from continually updated urns.

The GPU scheme translates the idea of Bayesian learning, specifying through (2.1) how to update the knowledge about survival times when data come from follow-up studies on different subjects.

Walker and Muliere (1997) and Muliere, Secchi and Walker (2000) show that the sequence $\{T_n\}$ is exchangeable and, by de Finetti's Representation Theorem, there exists a random distribution function F, such that, given F, the random variables T_n are independent and identically distributed with distribution F. Moreover, F is distributed according to a beta-Stacy process on the integers with parameters $\{\alpha_j, \beta_j,\ j \in \mathbb{N}_0\}$. The beta-Stacy process is *neutral to the right* and, so, conjugate to right censored observations.

Let $T_1, \ldots, T_n$ be independent and identically distributed and subject to right censoring. Thus, what is observed is $(T_1^*, \delta_1), \ldots, (T_n^*, \delta_n)$ where

$$\begin{aligned} T_i^* &= t, \delta_i = 0 \quad \Leftrightarrow \text{a censoring took place: } T_i > t \\ T_i^* &= t, \delta_i = 1 \quad \Leftrightarrow \text{a death happened: } T_i = t. \end{aligned}$$

With a quadratic loss function, the predictive distribution of T_{n+1} given $(\mathbf{T}_n, \boldsymbol{\delta}_n)$ is the Bayes estimator for the random distribution function.

So, analogously for the survival function and under a beta-Stacy prior, we have

$$\begin{aligned} \hat{S}(t) &= P\left[T_{n+1} > t \,|\, \mathbf{T}_n^* = \mathbf{t}_n, \boldsymbol{\delta}_n = \mathbf{d}_n\right] \\ &= \prod_{j=0}^{t} \left[1 - \frac{\alpha_j + m_j^*(\mathbf{t}_n, \mathbf{d}_n)}{\alpha_j + \beta_j + s_j(\mathbf{t}_n)}\right], \end{aligned} \tag{2.2}$$

where $\mathbf{d}_n \in \{0,1\}^n$ and $m_j^*(\mathbf{t}_n, \mathbf{d}_n) = \sum_{k=1}^n \mathbb{1}_{[t_k=j, d_k=1]}$.

Finally, note the following:

1. without censoring (2.2) reduces to

$$\hat{S}(t) = P\left[T_{n+1} > t \,|\, \mathbf{T}_n = \mathbf{t}_n\right] = \prod_{j=0}^{t} \frac{\beta_j + r_j\left(\mathbf{t}_n\right)}{\alpha_j + \beta_j + s_j\left(\mathbf{t}_n\right)};$$

2. the limit of (2.2) for $\alpha_j, \beta_j \to 0$, $\forall\, j$, reduces to classical Kaplan-Meier estimator end extend it when it was not defined.

3. A Bivariate Reinforced Random Process

3.1. Definition

Our aim is to construct a bivariate random process $\{(X_n, Y_n)\,,\ n \geq 1\}$ providing a model for coupled lifetimes and incorporating the basic principle of reinforcement similar to the GPU. The idea underlying the GPU scheme is that of reinforcing the path from the origin to the point $T_n = t$ before generating T_{n+1}, sequentially for $n \geq 1$. This is relatively easy because the random variables T_n can be represented as points on the non-negative line. On the other hand, (X_n, Y_n) is a point in the non-negative orthant and, unfortunately, there is no unique path from $(0,0)$ to (x, y). So, the reinforcement procedure must be introduced in an alternative way.

Let $\{A_n, n \geq 1\}$, $\{B_n, n \geq 1\}$ and $\{C_n, n \geq 1\}$ be independent sequences from GPUs with parameters (α_j^A, β_j^B), (α_j^B, β_j^B) and (α_j^C, β_j^C), $j \in \mathbb{N}_0$, respectively. Now, define a bivariate random process $\{(X_n, Y_n)\,,\, n \geq 1\}$ by

$$\begin{aligned} X_n &= A_n + B_n \\ Y_n &= A_n + C_n, \quad n \geq 1. \end{aligned} \tag{3.1}$$

The relations above postulate a particular and very easy form of dependence between X_n and Y_n: for a given couple, each element is supposed to have a common component (A_n) and a individual specific one (B_n and C_n). This is a natural way to construct a dependence without forming a parametric model. And, mathematically, it is that the dependence between X_n and Y_n is determined by the variance of A_n.

By this construction it turns out, conditionally on A_n, that X_n and Y_n are independent. Moreover, $\sigma\left(\mathbf{A}_n, \mathbf{B}_n, \mathbf{C}_n\right) = \sigma\left(\mathbf{A}_n, \mathbf{X}_n, \mathbf{Y}_n\right)$.

The structure of the dependence is described by

$$\begin{aligned} &\text{Cov}\ (X_1, Y_1) = \text{Var}\ (A_1) \geq 0, \\ &\text{Cov}\ \left(X_{n+1}, Y_{n+1} |\, \mathbf{A}_n, \mathbf{B}_n, \mathbf{C}_n\right) = \text{Var}\ \left(A_{n+1} |\, \mathbf{A}_n\right), \quad n \geq 1. \end{aligned}$$

Hence, we have introduced, correspondingly to each lifetimes couple, an auxiliary variable whose variance describes the covariance between X_n and Y_n. Note that even though $\mathbf{A}_n$ cannot be directly observed, the predictive distribution of this bivariate process can be easily computed in terms of those of the three GPUs, as is shown in the Appendix A.

The most interesting feature of this bivariate process given in the following proposition.

Proposition 1. *The sequence of couples* $\{(X_n, Y_n), n \geq 1\}$ *is exchangeable.*

Proof. Note that (X_n, Y_n) is a measurable function of (A_n, B_n, C_n). This is an exchangeable sequence and hence so is $\{(X_n, Y_n),\, n \geq 1\}$.

The de Finetti Representation Theorem assures the existence of a bivariate random distribution function F_{XY}, conditionally on which the couples (X_n, Y_n) are independent and identically distributed with distribution F_{XY}. The random distribution functions corresponding to GUPs $\{A_n\}$, $\{B_n\}$ and $\{C_n\}$ are F_A, F_B and F_C; moreover, they are distributed according to beta-Stacy processes with parameters $(\alpha_j^A, \beta_j^A), (\alpha_j^B, \beta_j^B)$ and (α_j^C, β_j^C), $j \in \mathbb{N}_0$, respectively. On the other hand, the de Finetti's measure of $\{(X_n, Y_n)\}$, that is the distribution of the bivariate distribution function F_{XY}, is not explicitly known. Nevertheless, it is possible to elicit some of its properties.

First of all, if F_X and F_Y are the marginal distributions of $\{X_n\}$ and $\{Y_n\}$, given F_A, F_B and F_C, (3.1) implies

$$\begin{aligned} F_X &= F_A * F_B \\ F_Y &= F_A * F_C. \end{aligned} \tag{3.2}$$

Therefore, each of the two marginals is a convolution of two beta-Stacy processes.

In terms of probability functions, we can write

$$P_{XY}(x, y) = \sum_{a=0}^{x \wedge y} P_A(a) P_B(x-a) P_C(y-a) \quad \forall\,(x, y) \in \mathbb{N}_0^2, \tag{3.3}$$

where P is the probability function corresponding to the distribution F. For the marginals we get the usual expression for the convolution.

Moreover, given F_{XY}, F_A, F_B and F_C, if $\sigma_A^2 = Var_{F_A}(A)$, the dependence between X and Y is described by

$$\text{Cov}_{\,F_{XY}}(X, Y) = \sigma_A^2. \tag{3.4}$$

Thus, assuming a model for coupled lifetimes data, represented by the bivariate process built on the basis of GPU and equations (3.1) is equivalent, in a Bayesian point of view, to defining a probability measure on the space of the bivariate distribution functions on $\mathbb{N}_0^2$. Let Π_2 be such a measure.

3.2. The support of Π_2

Before exploring the properties of the support of Π_2, it is worth recalling the topology on spaces of probability measures. Given $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ a measurable space, where $\mathcal{X}$ is a Polish space and $\mathcal{B}(\mathcal{X})$ its Borel σ-algebra, let $\mathcal{M}$ denote the space of all probability measures on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ and $\mathcal{B}(\mathcal{M})$ a suitable Borel σ-algebra.

If, for a random element P of $(\mathcal{M}, \mathcal{B}(\mathcal{M}))$, Π is the prior and $\Pi(\cdot\,|\mathbf{X}_n)$ the posterior, given a vector of observations $\mathbf{X}_n = (X_1, \ldots, X_n) \in \mathcal{X}^n$ independent and identically distributed according to P, the definition of a suitable topology on $(\mathcal{M}, \mathcal{B}(\mathcal{M}))$ allows us to deal with some interesting problems. First of all, we can point out which probability measures belong to the support of Π and, secondly, we can describe the asymptotic behaviour of the posterior when the number of observations grows.

Definition 2. A subset U of $\mathcal{M}$ is said to be a weak neighborhood of P_0 if it contains a set of the form $\{P : |P(A_i) - P_0(A_i)| < \varepsilon_i,\, i = 1, \ldots, k\}$ where A_i are P_0-continuous sets and $\varepsilon_i > 0$. The sequence $\{\Pi\left(\cdot\,|\mathbf{X}_n\right), n \geq 1\}$ is said to be weak consistent at P_0, if for every weak neighborhood U of P_0, $\Pi\left(U|\mathbf{X}_n\right) \to 1,\ n \to \infty \quad \text{a.s. } P_0$.

Definition 3. Let $L(\mu)$ the set of all densities with respect to a finite σ-measure μ. For $f, f_0 \in L(\mu)$ the Kullback-Leibler divergence of f from f_0 is

$$K\left(f, f_0\right) = \int f_0 \log \frac{f_0}{f} d\mu. \tag{3.5}$$

For $\varepsilon > 0$, $\{f : K(f_0, f) < \varepsilon\}$ is the Kullback-Leibler neighborhood of f_0 with radius ε.

It is possible to give a definition of Kullback-Leibler consistency, but we do not deal with it. More important for our purpose is a corollary of a classical and well-known result of Schwartz (1965), providing a sufficient condition for weak consistency.

Proposition 2.(Schwartz (1965)) If $\forall\, \varepsilon > 0$, $\Pi\left(f : K(f_0, f) < \varepsilon\right) > 0$, then the posterior is weakly consistent at f_0.

Exploiting the properties of the beta-Stacy process, we can obtain some knowledge about the width of the support of Π_2 with respect to weak topology. We let $P_{XY}(i,j) = p_{ij}, \forall\, (i,j) \in \mathbb{N}_0^2$. Moreover, for a distribution (or more generally for a measure) μ, let $\mathcal{S}_\mu$ indicate its support.

Let $\mathcal{P}_{c2}$ be the set of the probability distributions on $\mathbb{N}_0^2$ such that there exist three probability distributions on $\mathbb{N}_0$ which (3.3) is satisfied.

The above proposition transfers a well-known result for the Dirichlet process, easily proved also for discrete beta-Stacy, to $\mathcal{P}_{c2}$. Proofs of following are in the Appendix B.

Proposition 3. *If Π_2 is the prior on the space of the bivariate distributions on $\mathbb{N}_0^2$ determined by the parameters $\alpha_j^A, \beta_j^A, \alpha_j^B, \beta_j^B, \alpha_j^C, \beta_j^C > 0$, $\forall\, j \in \mathbb{N}_0$, then*

$$\mathcal{S}_{\Pi_2} \supseteq \mathcal{P}_{c2}. \tag{3.6}$$

Remark 1. As will be pointed out in the next section, it is possible to center Π_2 with respect to three distributions on $\mathbb{N}_0$, Q_A, Q_B and Q_C; some of the parameters could be null and, instead of (3.6), we have

$$\begin{aligned} \mathcal{S}_{\Pi_2} \supseteq \{ P_{XY} \text{ s.t. } \exists P_A, P_B, P_c \text{ satisfying (3.3) and} \\ \mathcal{S}_{P_A} \subseteq \mathcal{S}_{Q_A}, \mathcal{S}_{P_B} \subseteq \mathcal{S}_{Q_B}, \mathcal{S}_{P_C} \subseteq \mathcal{S}_{Q_C} \}, \end{aligned}$$

where the set on the right-hand side is just a subset of $\mathcal{P}_{c2}$.

A more detailed knowledge of the support of Π_2 is provided by analysing the behaviour of the prior in the Kullback-Leibler neighborhoods. As in the previous case, the best strategy is to extend to $\mathcal{P}_{c2}$ the properties of the beta-Stacy process on $\mathbb{N}_0$.

Proposition 4. *Let Π the measure induced by a beta-Stacy process with parameters α_j, β_j, $j \in \mathbb{N}_0$, on the space of distribution functions on $\mathbb{N}_0$ with $p^o = \{p_j^o,\, j \in \mathbb{N}_0\}$ being one of them. If α_j, β_j, $j \in \mathbb{N}_0$ are such that*

$$\sum_{j=0}^{\infty} p_j^o \left[\frac{\beta_j}{\alpha_j(\alpha_j + \beta_j)} + \beta_j\, T(\alpha_j, \beta_j) \right] < +\infty, \tag{3.7}$$

$$\sum_{j=0}^{\infty} \overline{p_j^o} \left[\frac{\alpha_j}{\beta_j(\alpha_j + \beta_j)} + \alpha_j\, T(\beta_j, \alpha_j) \right] < +\infty, \tag{3.8}$$

where $\overline{p_j^o} = \sum_{k>j} p_k^o$, and

$$T(\alpha, \beta) = \sum_{n=0}^{\infty} \frac{1}{\alpha + n + 1} \frac{1}{\alpha + \beta + n + 1} < +\infty \quad \alpha, \beta > 0,$$

then $\forall\, \varepsilon > 0$,

$$\Pi\left[p : \sum_{j=0}^{\infty} p_j^o \log \frac{p_j^o}{p_j} < \varepsilon \right] > 0.$$

The previous proposition gives sufficient conditions for the parameters to produce positive probability in each Kullback-Leibler neighborhood of p^o; straightforwardly, from Schwartz' result, we get the weak consistency at p^o.

An easy condition for consistency when p^o is unknown is, for instance,

$$0 < \underline{\alpha} \leq \alpha_j \leq \overline{\alpha},\ \forall\, j \in \mathbb{N}_0, \tag{3.9}$$

$$0 < \underline{\beta} \leq \beta_j \leq \overline{\beta},\ \forall\, j \in \mathbb{N}_0. \tag{3.10}$$

For the bivariate prior, the following holds.

Proposition 5. *Let* Π_2 *indicate the prior on the space of bivariate distribution function on* $\mathbb{N}_0^2$ *with parameters* $\alpha_j^A, \beta_j^A, \alpha_j^B, \beta_j^B, \alpha_j^C, \beta_j^C > 0$, $j \in \mathbb{N}_0$, *and* $p^o \in \mathcal{P}_{c2}$ *with marginals* $p_X^o = \{p_{i\cdot}^o,\, i \in \mathbb{N}_0\}$ *and* $p_Y^o = \{p_{\cdot j}^o,\, j \in \mathbb{N}_0\}$, *where* $p_{i\cdot}^o = \sum_{j=0}^{\infty} p_{ij}^o$ *and* $p_{\cdot j}^o = \sum_{i=0}^{\infty} p_{ij}^o$.

If α_j^B, β_j^B *for* $j \in \mathbb{N}_0$ *satisfy*

$$\sum_{j=0}^{\infty} p_{X\,j}^o \left[\frac{\beta_j}{\alpha_j(\alpha_j + \beta_j)} + \beta_j\, T(\alpha_j, \beta_j) \right] < +\infty,$$

$$\sum_{j=0}^{\infty} \overline{p_{X\,j}^o} \left[\frac{\alpha_j}{\beta_j(\alpha_j + \beta_j)} + \alpha_j\, T(\beta_j, \alpha_j) \right] < +\infty,$$

and analogous conditions hold for α_j^C, β_j^C *subtituting* $p_{X\,j}^o$ *with* $p_{Y\,j}^o$, *then* $\forall\, \varepsilon > 0$

$$\Pi_2 \left[p : \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} p_{ij}^o \log \frac{p_{ij}^o}{\sum_{a=0}^{i \wedge j} p_a^A p_{i-a}^B p_{j-a}^C} < \varepsilon \right] > 0. \tag{3.11}$$

Therefore, provided (3.11) holds, the crucial conditions must be satisfied just for α_j^B, β_j^B and α_j^C, β_j^C with respect to the marginals. Taking these parameters as suggested, for instance in (3.9) and (3.10), we obtain that $\Pi_2\,(\cdot\,|\mathbf{X}_n, \mathbf{Y}_n)$ is consistent at every $p^o \in \mathcal{P}_{c2}$.

The next section shows how to make inference about the bivariate survival function when Π_2 is the prior.

4. Estimation of a Bivariate Survival Function

Though the lack of a well-determined form prevents the direct computation of the posterior, it is possible to obtain an estimate of the survival function.

More precisely, let

$$S(x, y) = P\,[X > x, Y > y] \tag{4.1}$$

be a bivariate survival function on $\mathbb{N}_0^2$ and $(\mathbf{X}_n, \mathbf{Y}_n)$ be an independent and identically distributed sample from S, where each of the two components is subject to independent right censoring, so that the data take the form $(\mathbf{X}_n^*, \boldsymbol{\delta}_n, \mathbf{Y}_n^*, \boldsymbol{\xi}_n)$.

With this Π_2 as prior, we are interested in the predictive distribution

$$\hat{S}(x,y) = P\left[X_{n+1} > x, Y_{n+1} > y \,|\, \mathbf{X}_n^*, \boldsymbol{\delta}_n, \mathbf{Y}_n^*, \boldsymbol{\xi}_n\right]. \tag{4.2}$$

Remark 2. Note that, if the data are not subjected to censoring, weak consistence of the posterior implies weak consistence of the predictive distribution. Hence the conditions on the parameters of the previous section are sufficient for a consistent predictive distribution as will.

Before any inference, an interpretation of the parameters of the model must be provided. This coincides, in the Bayesian perspective, with specifying the initial distribution centered in accordance with some prior choice.

Recall that if F is distributed according to a discrete beta-Stacy process, it is possible to center on a given discrete distribution G (Walker and Muliere (1997)) so that $E\left[F\left(\{j\}\right)\right] = G\left(\{j\}\right)$, putting, $\forall\, j, c_j > 0$ and

$$\alpha_j = c_j G\left(\{j\}\right) \qquad \beta_j = c_j\left(1 - \sum_{i=0}^{j} G\left(\{i\}\right)\right).$$

Moreover, if $c_j = c \;\forall\, j$, then $\alpha_j + \beta_j = \beta_{j-1}$ and the beta-Stacy reduces to a discrete Dirichlet process. While G is the distribution corresponding to the initial guess, the parameters c_j play the role of *strength of belief*, so that higher values mean higher weight is given to this guess in the posterior distribution.

Hence, to determine the parameters of our model we do not need some complicated idea about the bivariate distribution of the lifetimes, but rather *a priori* guesses on the covariance between X and Y, $\mathrm{Cov}\,(X,Y)$, and on their marginal distributions.

In fact, (3.2) and (3.4) suggest proceeding as follow:

1. choose an initial distribution for A, F_A^0, having variance $\mathrm{Cov}\,(X,Y)$, and determine α_j^A, β_j^A;
2. given the prior guess F_X^0, F_Y^0 and F_A^0, solve (3.2) to obtain F_B^0 and F_C^0 and then compute α_j^B, β_j^B, α_j^C, β_j^C.

In the absence of further information about F_A, the choice of the distribution for A is free, except for the variance; hence, it could be convenient to choose a distribution making the procedure simpler at point 2.

For particular values of the parameters it is possible to abandon the hypothesis of dependence between X and Y. Indeed, taking $\alpha_0^A > 0$, $\alpha_j^A = 0 \;\forall\, j \geq 1$, $\beta_j^A = 0 \;\forall\, j \geq 0$, we obtain $A_n = 0$, $X_n = B_n$, $Y_n = C_n \;\forall\, n \; a.s.$ and

$$P\left[X_{n+1} > x, Y_{n+1} > y \right| \mathbf{X}_n^* = \mathbf{x}_n, \boldsymbol{\delta}_n = \mathbf{d}_n, \mathbf{Y}_n^* = \mathbf{y}_n, \boldsymbol{\xi}_n = \mathbf{e}_n]$$

$$= P\left[B_{n+1} > x \,|\, \mathbf{B}_n^* = \mathbf{x}_n, \boldsymbol{\delta}_n = \mathbf{d}_n\right] P\left[C_{n+1} > y \,|\, \mathbf{C}_n^* = \mathbf{y}_n, \boldsymbol{\xi}_n = \mathbf{e}_n\right]$$
$$= P\left[X_{n+1} > x \,|\, \mathbf{X}_n^* = \mathbf{x}_n, \boldsymbol{\delta}_n = \mathbf{d}_n\right] P\left[Y_{n+1} > y \,|\, \mathbf{Y}_n^* = \mathbf{y}_n, \boldsymbol{\xi}_n = \mathbf{e}_n\right].$$

This corresponds to assuming as prior for F_A a Dirichlet process centered in a distribution degenerate at 0. If, in addition, $\alpha_j^B, \beta_j^B, \alpha_j^C, \beta_j^C \to 0 \;\forall j$, the result is the product of Kaplan-Meier estimators for X and Y.

When the parameters are fixed, the right hand side of (4.2) can be directly expressed in terms of the predictives of A_n, B_n and C_n, as in the previous section. The practice computation may yet be cumbersome and laborious.

A Markov Chain Monte Carlo estimation procedure can be achieved without difficulty through the following steps.

1. Given the observations $(\mathbf{X}_n^*, \boldsymbol{\delta}_n, \mathbf{Y}_n^*, \boldsymbol{\xi}_n)$, $\mathbf{A}_n$ is generated via a Gibbs sampler, so the full conditional of A_n, $P_{A_n|\mathbf{A}_{n-1},\mathbf{X}_n^*,\boldsymbol{\delta}_n,\mathbf{Y}_n^*,\boldsymbol{\xi}_n}$, is

$$\begin{aligned} &P\left[A_n = a_n \,|\mathbf{A}_{n-1} = \mathbf{a}_{n-1}, \mathbf{X}_n^* = \mathbf{x}_n, \boldsymbol{\delta}_n = \mathbf{d}_n, \mathbf{Y}_n^* = \mathbf{y}_n, \boldsymbol{\xi}_n = \mathbf{e}_n\right] \\ &\quad \propto P\left[B_n^* = x_n - a_n, \delta_n = d_n \left|\mathbf{B}_{n-1}^* = \mathbf{x}_{n-1} - \mathbf{a}_{n-1}, \boldsymbol{\delta}_{n-1} = \mathbf{d}_{n-1}\right.\right] \\ &\qquad P\left[C_n^* = y_n - a_n, \xi_n = e_n \left|\mathbf{C}_{n-1}^* = \mathbf{y}_{n-1} - \mathbf{a}_{n-1}, \boldsymbol{\xi}_{n-1} = \mathbf{e}_{n-1}\right.\right] \\ &\qquad P\left[A_n = a_n \,|\mathbf{A}_{n-1} = \mathbf{a}_{n-1}\right], \end{aligned}$$

where

$$\begin{aligned} &P\left[B_n^* = b, \delta_n = d \left|\mathbf{B}_{n-1}^* = \mathbf{b}_{n-1}, \boldsymbol{\delta}_{n-1} = \mathbf{d}_{n-1}\right.\right] \\ &\quad = \begin{cases} P\left[B_n \geq b \left|\mathbf{B}_{n-1}^* = \mathbf{b}_{n-1}, \boldsymbol{\delta}_{n-1} = \mathbf{d}_{n-1}\right.\right], \; d = 0 \\ \\ P\left[B_n = b \left|\mathbf{B}_{n-1}^* = \mathbf{b}_{n-1}, \boldsymbol{\delta}_{n-1} = \mathbf{d}_{n-1}\right.\right], \; d = 1 \end{cases} \end{aligned}$$

and similarly for $P\left[C_n^* = y_n - a_n, \xi_n = d_n \left|\mathbf{C}_{n-1}^* = \mathbf{c}_{n-1}, \boldsymbol{\xi}_{n-1} = \mathbf{e}_{n-1}\right.\right]$. For exchangeability of $\{A_n\}$, the other full conditionals $P_{A_i|\mathbf{A}_{-i},\mathbf{X}_n^*,\boldsymbol{\delta}_n,\mathbf{Y}_n^*,\boldsymbol{\xi}_n}$, $i = 1, \ldots, n-1$ have an analogous form. (Let $\mathbf{A}_{-i} = (A_1, \ldots, A_{i-1}, A_{i+1}, \ldots, A_n)$). Then $\mathbf{B}_n^* = \mathbf{X}_n^* - \mathbf{A}_n$ and $\mathbf{C}_n^* = \mathbf{Y}_n^* - \mathbf{A}_n$ are computed.
2. A_{n+1}, B_{n+1} and C_{n+1} are sampled according to the predictive distributions $P_{A_{n+1}|\mathbf{A}_n}$, $P_{B_{n+1}|\mathbf{B}_n^*,\boldsymbol{\delta}_n}$ and $P_{C_{n+1}|\mathbf{C}_n^*,\boldsymbol{\xi}_n}$.
3. Take $X_{n+1} = A_{n+1} + B_{n+1}$ and $Y_{n+1} = A_{n+1} + C_{n+1}$.

This is straightforward to implement in practice.

5. An Example

We give an example involving a dataset. The following table reports the data (Woolson and Lachenbruch (1980) and Lin and Ying (1993)) consisting of survival times of two kinds of skin grafts on the same burn patient; more precisely, X is referred to as closely matched grafts, Y as poorly matched ones in relation to the HL-A antigen system.

Table 1. Days of survival of skin grafts on burn patients.

Patient	1	2	3	4	5	6	7	8	9	10	11
Close match (X)	37	19	57^+	93	16	22	20	18	63	29	60^+
Poor match (Y)	29	13	15	26	11	17	26	21	43	15	40

First, we try to reproduce the Kaplan-Meier ("empirical") estimator. Using well-understood ideas connecting Bayes nonparametric estimators, and empirical estimators, we allow F_A to be the distribution degenerate at 0. We take $\alpha_0^A = 1,000$, $\alpha_j^A = 0 \;\forall\; j \geq 1$, $\beta_j^A = 0 \;\forall\; j \geq 0$, while $\alpha_j^B, \beta_j^B, \alpha_j^C, \beta_j^C$ are all taken close to 0. The marginal estimators are represented by black circles in Figures 1 and 2, and compare very well with the Kaplan-Meier estimates, provided with their standard associated confidence bounds.

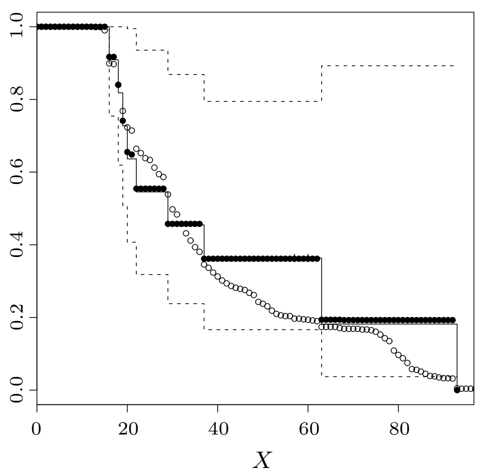

Figure 1. Marginal estimate for X.

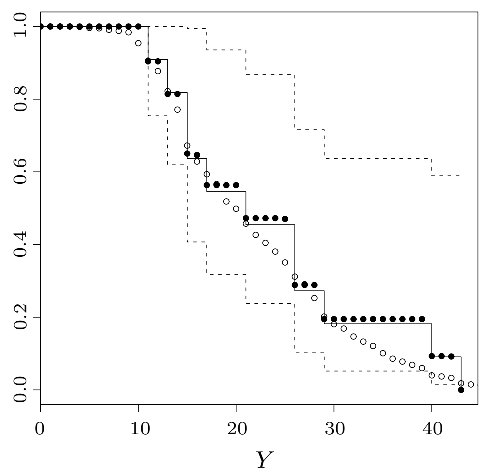

Figure 2. Marginal estimate for Y.

Figure 3 shows the joint estimator. As we know, according to theoretical issues about the meaning of the parameters, in this case the estimators are constructed as if the X and Y data are different (independent) data sets; that is we have lost any dependence between the pairs in order to obtain the empirical marginal estimates.

Here we also give a "smooth" estimate when the prior choice embodies some knowledge about the dependence between X and Y. This will carry through, and be updated, into the posterior. We choose to center the priors on Poisson (Po) distributions to exploit some of their closure properties under convolution. For illustrative purposes, we center F_A on the Po(10) distribution, which corresponds to $\mathrm{Cov}\,(X, Y) = 10$, and center F_X and F_Y on the Po(40) and Po(25) distributions, respectively. Solving the equations in (3.2), we obtain a Po(40) distribution for F_B and a Po(15) distribution for F_C. Finally, we put all the parameters related to the degree of belief c_j^A, c_j^B and c_j^C equal to $1, \forall\, j \in \mathbb{N}_0$.

The estimates are represented by white circles in Figures 1 and 2. As can be seen, the marginal estimators are smoothed version of Kaplan-Meier estimators. Similarly, the joint survival function of Figure 4 is a smoothed version of Figure 3.

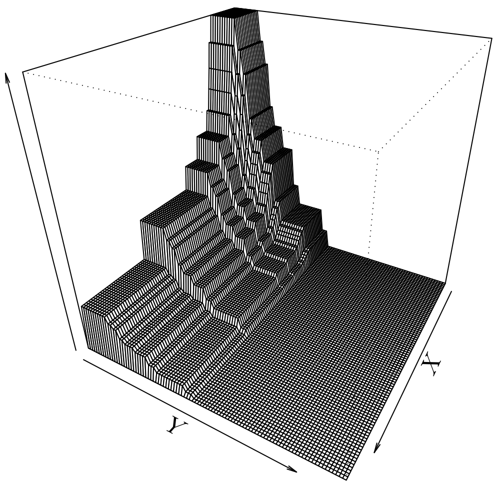

Figure 3. Empirical independent case, estimate of the bivariate survival function.

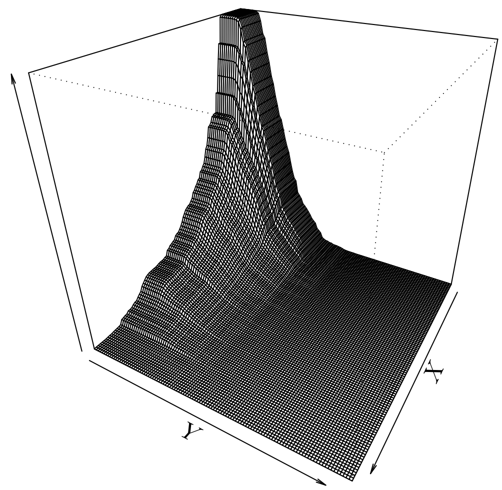

Figure 4. Smooth case, estimate of the bivariate survival function.

Acknowledgement

The third author is financed by an EPSRC Advanced Research Fellowship.

Appendix A. Predictive Distribution

The predictive distribution $P\left[X_{n+1}>x, Y_{n+1}>y \middle| \mathbf{X}_n=\mathbf{x}_n, \mathbf{Y}_n=\mathbf{y}_n\right]$ is nothing but

$$P\left[X_{n+1}>x, Y_{n+1}>y \middle| \mathbf{X}_n=\mathbf{x}_n, \mathbf{Y}_n=\mathbf{y}_n)\right] = \frac{P\left[X_{n+1}>x, Y_{n+1}>y, \mathbf{X}_n=\mathbf{x}_n, \mathbf{Y}_n=\mathbf{y}_n\right]}{P\left[\mathbf{X}_n=\mathbf{x}_n, \mathbf{Y}_n=\mathbf{y}_n\right]},$$

where

$$P\left[X_{n+1}>x, Y_{n+1}>y, \mathbf{X}_n=\mathbf{x}_n, \mathbf{Y}_n=\mathbf{y}_n\right] = \sum_{a_{n+1}=0}^{x\wedge y} \sum_{a_n=0}^{x_n\wedge y_n} \cdots \sum_{a_1=0}^{x_1\wedge y_1} \left[\frac{\alpha^A_{a_{n+1}} + m_{a_{n+1}}\left(\mathbf{a}_n\right)}{\alpha^A_{a_{n+1}} + \beta^A_{a_{n+1}} + s_{a_{n+1}}\left(\mathbf{a}_n\right)} \prod_{j=0}^{a_{n+1}-1} \frac{\beta^A_j + r_j\left(\mathbf{a}_n\right)}{\alpha^A_j + \beta^A_j + s_j\left(\mathbf{a}_n\right)} \right.$$

$$\prod_{j=0}^{x-a_{n+1}} \frac{\beta_j^B + r_j\left(\mathbf{x}_n - \mathbf{a}_n\right)}{\alpha_j^B + \beta_j^B + s_j\left(\mathbf{x}_n - \mathbf{a}_n\right)} \prod_{j=0}^{y-a_{n+1}} \frac{\beta_j^C + r_j\left(\mathbf{y}_n - \mathbf{a}_n\right)}{\alpha_j^C + \beta_j^C + s_j\left(\mathbf{y}_n - \mathbf{a}_n\right)}$$

$$\prod_{i=1}^{n-1}\left(\frac{\alpha_{a_{i+1}}^A + m_{a_{i+1}}\left(\mathbf{a}_i\right)}{\alpha_{a_{i+1}}^A + \beta_{a_{i+1}}^A + s_{a_{i+1}}\left(\mathbf{a}_i\right)} \prod_{j=0}^{a_{i+1}-1} \frac{\beta_j^A + r_j\left(\mathbf{a}_i\right)}{\alpha_j^A + \beta_j^A + s_j\left(\mathbf{a}_i\right)}\right.$$

$$\frac{\alpha_{x_{i+1}-a_{i+1}}^B + m_{x_{i+1}-a_{i+1}}\left(\mathbf{x}_i-\mathbf{a}_i\right)}{\alpha_{x_{i+1}-a_{i+1}}^B + \beta_{x_{i+1}-a_{i+1}}^B + s_{x_{i+1}-a_{i+1}}\left(\mathbf{x}_i-\mathbf{a}_i\right)} \prod_{j=0}^{x_{i+1}-a_{i+1}-1} \frac{\beta_j^B + r_j\left(\mathbf{x}_i-\mathbf{a}_i\right)}{\alpha_j^B + \beta_j^B + s_j\left(\mathbf{x}_i-\mathbf{a}_i\right)}$$

$$\left.\frac{\alpha_{y_{i+1}-a_{i+1}}^C + m_{y_{i+1}-a_{i+1}}\left(\mathbf{y}_i-\mathbf{a}_i\right)}{\alpha_{y_{i+1}-a_{i+1}}^C + \beta_{y_{i+1}-a_{i+1}}^C + s_{y_{i+1}-a_{i+1}}\left(\mathbf{y}_i-\mathbf{a}_i\right)} \prod_{j=0}^{y_{i+1}-a_{i+1}-1} \frac{\beta_j^C + r_j\left(\mathbf{y}_i-\mathbf{a}_i\right)}{\alpha_j^C + \beta_j^C + s_j\left(\mathbf{y}_i-\mathbf{a}_i\right)}\right)$$

$$\frac{\alpha_{a_1}^A}{\alpha_{a_1}^A + \beta_{a_1}^A} \prod_{j=0}^{a_1-1} \frac{\beta_j^A}{\alpha_j^A + \beta_j^A} \frac{\alpha_{x_1-a_1}^B}{\alpha_{x_1-a_1}^B + \beta_{x_1-a_1}^B} \prod_{j=0}^{x_1-a_1-1} \frac{\beta_j^B}{\alpha_j^B + \beta_j^B}$$

$$\left.\frac{\alpha_{y_1-a_1}^C}{\alpha_{y_1-a_1}^C + \beta_{y_1-a_1}^C} \prod_{j=0}^{y_1-a_1-1} \frac{\beta_j^C}{\alpha_j^C + \beta_j^C}\right],$$

$$P\left[\mathbf{X}_n = \mathbf{x}_n, \mathbf{Y}_n = \mathbf{y}_n\right]$$

$$= \sum_{a_n=0}^{x_n \wedge y_n} \cdots \sum_{a_1=0}^{x_1 \wedge y_1} \left[\prod_{i=1}^{n-1}\left(\frac{\alpha_{a_{i+1}}^A + m_{a_{i+1}}\left(\mathbf{a}_i\right)}{\alpha_{a_{i+1}}^A + \beta_{a_{i+1}}^A + s_{a_{i+1}}\left(\mathbf{a}_i\right)} \prod_{j=0}^{a_{i+1}-1} \frac{\beta_j^A + r_j\left(\mathbf{a}_i\right)}{\alpha_j^A + \beta_j^A + s_j\left(\mathbf{a}_i\right)}\right.\right.$$

$$\frac{\alpha_{x_{i+1}-a_{i+1}}^B + m_{x_{i+1}-a_{i+1}}\left(\mathbf{x}_i-\mathbf{a}_i\right)}{\alpha_{x_{i+1}-a_{i+1}}^B + \beta_{x_{i+1}-a_{i+1}}^B + s_{x_{i+1}-a_{i+1}}\left(\mathbf{x}_i-\mathbf{a}_i\right)} \prod_{j=0}^{x_{i+1}-a_{i+1}-1} \frac{\beta_j^B + r_j\left(\mathbf{x}_i-\mathbf{a}_i\right)}{\alpha_j^B + \beta_j^B + s_j\left(\mathbf{x}_i-\mathbf{a}_i\right)}$$

$$\left.\frac{\alpha_{y_{i+1}-a_{i+1}}^C + m_{y_{i+1}-a_{i+1}}\left(\mathbf{y}_i-\mathbf{a}_i\right)}{\alpha_{y_{i+1}-a_{i+1}}^C + \beta_{y_{i+1}-a_{i+1}}^C + s_{y_{i+1}-a_{i+1}}\left(\mathbf{y}_i-\mathbf{a}_i\right)} \prod_{j=0}^{y_{i+1}-a_{i+1}-1} \frac{\beta_j^C + r_j\left(\mathbf{y}_i-\mathbf{a}_i\right)}{\alpha_j^C + \beta_j^C + s_j\left(\mathbf{y}_i-\mathbf{a}_i\right)}\right)$$

$$\frac{\alpha_{a_1}^A}{\alpha_{a_1}^A + \beta_{a_1}^A} \prod_{j=0}^{a_1-1} \frac{\beta_j^A}{\alpha_j^A + \beta_j^A} \frac{\alpha_{x_1-a_1}^B}{\alpha_{x_1-a_1}^B + \beta_{x_1-a_1}^B} \prod_{j=0}^{x_1-a_1-1} \frac{\beta_j^B}{\alpha_j^B + \beta_j^B}$$

$$\left.\frac{\alpha_{y_1-a_1}^C}{\alpha_{y_1-a_1}^C + \beta_{y_1-a_1}^C} \prod_{j=0}^{y_1-a_1-1} \frac{\beta_j^C}{\alpha_j^C + \beta_j^C}\right].$$

Appendix B: Proofs of Propositions and Lemmas

Proof of Proposition 3. As $\mathbb{N}_0^2$ is discrete we need only show $\Pi_2[|p_{x_iy_i} - p^o_{x_iy_i}| < \varepsilon,\ i = 1,\ldots,k] > 0$ for each $p^o \in \mathcal{P}_{c2}$, $\forall\ \varepsilon > 0$, $k \geq 1$, $\forall\ (x_i, y_i), \ldots (x_k, y_k)$

$\in \mathbb{N}_0^2$. Now, $\forall\, i$ $(m_i = x_i \wedge y_i)$

$$\begin{aligned}\left|p_{x_i y_i} - p^o_{x_i y_i}\right| &= \Bigg|\sum_{a=0}^{m_i}\left(p^A_a p^B_{x_i-a} p^C_{y_i-a} - p^{A,o}_a p^{B,o}_{x_i-a} p^{C,o}_{y_i-a}\right)\Bigg| \\ &\qquad -p^{A,o}_a p^{B,o}_{x_i-a} p^C_{y_i-a} + p^{A,o}_a p^{B,o}_{x_i-a} p^C_{y_i-a} - p^{A,o}_a p^{B,o}_{x_i-a} p^{C,o}_{y_i-a}\Big| \\ &\leq \sum_{a=0}^{m_i}\left|p^A_a - p^{A,o}_a\right| + \sum_{a=0}^{m_i}\left|p^B_{x_i-a} - p^{B,o}_{x_i-a}\right| + \sum_{a=0}^{m_i}\left|p^C_{y_i-a} - p^{C,o}_{y_i-a}\right| \\ &= \sum_{a=0}^{m_i}\left|p^A_a - p^{A,o}_a\right| + \sum_{j=x_i-m_i}^{x_i}\left|p^B_j - p^{B,o}_j\right| + \sum_{j=y_i-m_i}^{y_i}\left|p^C_j - p^{C,o}_j\right| \text{ a.s. } \Pi_2,\end{aligned}$$

as $p^{A,o}_i, p^{B,o}_i, p^{C,o}_i \in [0,1]$ and $p^A_i, p^B_i, p^C_i \in [0,1]$ a.s. Π_2, $\forall\, i$.

By construction, the sequences $\{p^A_i,\, i \in \mathbb{N}_0\}$, $\{p^B_i,\, i \in \mathbb{N}_0\}$, and $\{p^C_i,\, i \in \mathbb{N}_0\}$ are independent beta-Stacy processed on $\mathbb{N}_0$, so

$$\begin{aligned}\Pi_2\left[\left|p_{x_i y_i} - p^o_{x_i y_i}\right| < \varepsilon,\, i = 1,\ldots,k\right] \geq\ & \Pi_2\left[\sum_{j=0}^{m_i}\left|p^A_j - p^{A,o}_j\right| < \frac{\varepsilon}{3},\, i = 1,\ldots,k\right] \\ & \Pi_2\left[\sum_{j=0}^{x_i}\left|p^B_j - p^{B,o}_j\right| < \frac{\varepsilon}{3},\, i = 1,\ldots,k\right] \\ & \Pi_2\left[\sum_{j=0}^{y_i}\left|p^C_j - p^{C,o}_j\right| < \frac{\varepsilon}{3},\, i = 1,\ldots,k\right].\end{aligned}$$

Consider, for instance, the first element of the product.

Let $M = \bigvee_{i=1}^k m_i$. If

$$\left|p^A_j - p^{A,o}_j\right| < \frac{\varepsilon}{3(M+1)} = \varepsilon_M, \quad j = 0,\ldots,M,$$

then

$$\sum_{j=0}^{m_i}\left|p^A_j - p^{A,o}_j\right| < \frac{\varepsilon}{3},\, i = 1,\ldots,k,$$

so that

$$\Pi_2\left[\sum_{j=0}^{m_i}\left|p^A_j - p^{A,o}_j\right| < \frac{\varepsilon}{3},\, i = 1,\ldots,k\right] \geq \Pi_2\left[\left|p^A_j - p^{A,o}_j\right| < \varepsilon_M, j = 0,\ldots,M\right]. \tag{B.1}$$

Since $\left(p^A_0,\ldots,p^A_M\right)$ has a Generalized Dirichlet distribution with parameters $\alpha^A_j, \beta^A_j > 0,\ j = 0,1,\ldots,M$ (see Walker and Muliere (1997)), for the full support of this distribution, the result in Proposition 3 of Ferguson (1973) can be generalized to the discrete beta-Stacy process and the last probability in (B.1) is strictly positive.

Lemma 6. *If* $X \sim Beta\,(\alpha, \beta)$, $\alpha > 0$, $\beta > 0$, *then*

$$E\left[-\log X\right] = \frac{\beta}{\alpha}\left[\frac{1}{\alpha+\beta} + \alpha\sum_{n=0}^{\infty}\frac{1}{\alpha+n+1}\frac{1}{\alpha+\beta+n+1}\right] < +\infty.$$

Proof. We write $B(x;\alpha,\beta)$ for the incomplete beta function and note that $B(\alpha,\beta) = B(1;\alpha,\beta) = (\Gamma(\alpha)\Gamma(\beta))/\Gamma(\alpha+\beta)$ and $I(x;\alpha,\beta) = B(x;\alpha,\beta)/B(\alpha,\beta)$. Now

$$\begin{aligned} E\left[-\log X\right] &= \int_0^1 -\log x\frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}\,dx \\ &= \left|-\log x\, I(x;\alpha,\beta)\right|_0^1 + \int_0^1 \frac{1}{x}\, I(x;\alpha,\beta)\,dx. \qquad \text{(B.2)} \end{aligned}$$

The first term is 0 as

$$\begin{aligned} \lim_{x\to 0}\log x\, B(x;\alpha,\beta) &= \lim_{x\to 0}\frac{\log x}{B(x;\alpha,\beta)^{-1}} \\ &= \lim_{x\to 0}\frac{\frac{1}{x}}{-B(x;\alpha,\beta)^{-2}x^{\alpha-1}(1-x)^{\beta-1}} \qquad \text{(B.3)} \\ &= \lim_{x\to 0} -\frac{B(x;\alpha,\beta)^2}{x^{\alpha}} \\ &= \lim_{x\to 0} -\frac{2\,B(x;\alpha,\beta)x^{\alpha-1}(1-x)^{\beta-1}}{\alpha x^{\alpha-1}} = 0, \qquad \text{(B.4)} \end{aligned}$$

where (B.3) and (B.4) apply De l'Hôpital's rule.

According to Abramowitz and Stegun (1964),

$$I(x;\alpha,\beta) = \frac{x^{\alpha}(1-x)^{\beta}}{\alpha B(\alpha,\beta)}\left[1 + \sum_{n=0}^{\infty}\frac{B(\alpha+1,n+1)}{B(\alpha+\beta,n+1)}\,x^{n+1}\right],$$

So (B.2) reduces to

$$\begin{aligned} &\int_0^1 \frac{x^{\alpha-1}(1-x)^{\beta}}{\alpha B(\alpha,\beta)}\left[1 + \sum_{n=0}^{\infty}\frac{B(\alpha+1,n+1)}{B(\alpha+\beta,n+1)}\,x^{n+1}\right]dx \\ &= \frac{1}{\alpha}\left[\frac{B(\alpha,\beta+1)}{B(\alpha,\beta)} + \sum_{n=0}^{\infty}\frac{B(\alpha+1,n+1)}{B(\alpha+\beta,n+1)B(\alpha,\beta)}\int_0^1 x^{\alpha+n}(1-x)^{\beta}dx\right] \\ &= \frac{1}{\alpha}\left[\frac{\beta}{\alpha+\beta} + \sum_{n=0}^{\infty}\frac{B(\alpha+1,n+1)B(\alpha+n+1,\beta+1)}{B(\alpha+\beta,n+1)B(\alpha,\beta)}\right] \end{aligned}$$

$$= \frac{1}{\alpha}\left[\frac{\beta}{\alpha+\beta} + \sum_{n=0}^{\infty}\frac{\alpha}{\alpha+n+1}\frac{\beta}{\alpha+\beta+n+1}\right].$$

Since $\sum_{n=0}^{\infty}(\alpha+n+1)^{-1}(\alpha+\beta+n+1)^{-1} < +\infty$, we are finished.

Recall that $T(\alpha,\beta) = \sum_{n=0}^{\infty}(\alpha+n+1)^{-1}(\alpha+\beta+n+1)^{-1}$, $\alpha,\beta > 0$, is a well defined function.

Proof of Proposition 4. Under Π, for $j \in \mathbb{N}_0$, $p_j = u_j \prod_{k=0}^{j-1}[1-u_k]$, where $u_k \sim Beta(\alpha_k, \beta_k)$ and independent. Thus

$$\begin{aligned}
E\Big[-\sum_{j=0}^{\infty} p_j^o \log p_j\Big] &= \sum_{j=0}^{\infty} p_j^o E\left[-\log p_j\right] \\
&= \sum_{j=0}^{\infty} p_j^o E\left[-\log u_j\right] + \sum_{j=0}^{\infty}\sum_{k=0}^{j-1} p_j^o E\left[-\log\left(1-u_k\right)\right] \\
&= \sum_{j=0}^{\infty} p_j^o E\left[-\log u_j\right] + \sum_{k=0}^{\infty} \overline{p_k^o} E\left[-\log\left(1-u_k\right)\right] \\
&= \sum_{j=0}^{\infty} p_j^o \left[\frac{\beta_j}{\alpha_j(\alpha_j+\beta_j)} + \beta_j\, T(\alpha_j,\beta_j)\right] \\
&\quad + \sum_{j=0}^{\infty} \overline{p_j^o}\left[\frac{\alpha_j}{\beta_j(\alpha_j+\beta_j)} + \alpha_j\, T(\beta_j,\alpha_j)\right],
\end{aligned}$$

by Lemma 6.

Conditions (3.7) and (3.8) imply the finiteness of the expected value and

$$-\sum_{j=0}^{\infty} p_j^o \log p_j < +\infty \quad \text{a.s. } \Pi.$$

Since $\sum_{j=0}^{\infty} p_j^o \log(p_j^o/p_j) = \sum_{j=0}^{\infty} p_j^o \log p_j^o - \sum_{j=0}^{\infty} p_j^o \log p_j < \varepsilon$, it is sufficient to require

$$\sum_{j=0}^{M} p_j^o \log p_j^o - \sum_{j=0}^{M} p_j^o \log p_j < \frac{\varepsilon}{2}. \tag{B.5}$$

But $(p_0,\ldots,p_M)$, under Π, has a generalized Dirichlet distribution with parameters α_j,β_j, $j=0,\ldots,M$. Since this distribution has full support on the $M+1$-dimensional sub-simplex, and the event above is implied by $\{p_j > p_j^o/\exp(\epsilon/2)$, $j=0,\ldots,M\}$, (B.5) has strictly positive probability.

Proof of Proposition 5. Compute

$$E\Big[\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} -p^o_{ij}\log\sum_{a=0}^{i\wedge j} p^A_a p^B_{i-a} p^C_{j-a}\Big] = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} p^o_{ij} E\Big[-\log\sum_{a=0}^{i\wedge j} p^A_a p^B_{i-a} p^C_{j-a}\Big]$$
$$\leq \sum_{i=0}^{+\infty}\sum_{j=0}^{\infty} p^o_{ij} E\left[-\log p^A_0 p^B_i p^C_j\right]$$
$$= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} p^o_{ij} E\left[-\log p^A_0\right] + \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} p^o_{ij} E\left[-\log p^B_i\right] + \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} p^o_{ij} E\left[-\log p^C_j\right]$$
$$= E\left[-\log p^A_0\right] + \sum_{i=0}^{\infty} p^o_{i\cdot} E\left[-\log p^B_i\right] + \sum_{j=0}^{\infty} p^o_{\cdot j} E\left[-\log p^C_j\right].$$

Hence if α^B_j, β^B_j and α^C_j, β^C_j satisfy the above hypotheses, the expected value is finite.

Now

$$-\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} p^o_{ij}\log\sum_{a=0}^{i\wedge j} p^A_a p^B_{i-a} p^C_{j-a} < +\infty \quad \text{a.s. } \Pi_2.$$

To complete the proof, we can just show

$$\sum_{i=0}^{M}\sum_{j=0}^{M} p^o_{ij}\log p^o_{ij} - \sum_{i=0}^{M}\sum_{j=0}^{M} p^o_{ij}\log\sum_{a=0}^{i\wedge j} p^A_a p^B_{i-a} p^C_{j-a} < \frac{\varepsilon}{2} = \varepsilon' \tag{B.6}$$

has positive probability.

Now, as $p^o_{ij} = \sum_{a=0}^{i\wedge j} p^{A,o}_a p^{B,o}_{i-a} p^{C,o}_{j-a}$, $\forall\, i,j \in \mathbb{N}^2_0$, the event above is implied by

$$p^A_i > \frac{p^{A,o}_i}{\delta} \quad i = 0,\ldots,M, \tag{B.7}$$
$$p^B_i > \frac{p^{B,o}_i}{\delta} \quad i = 0,\ldots,M, \tag{B.8}$$
$$p^C_i > \frac{p^{C,o}_i}{\delta} \quad i = 0,\ldots,M, \tag{B.9}$$

where $\delta = \exp(\varepsilon'/3) > 1$. But, under Π_2, $\big(p^A_0,\ldots,p^A_M\big)$, $\big(p^B_0,\ldots,p^B_M\big)$ and $\big(p^C_0,\ldots,p^C_M\big)$ have independent generalized Dirichlet distribution with parameters α^A_j, β^A_j, $\alpha^B_j, \beta^B_j, \alpha^C_j$, $\beta^C_j > 0$, $j = 0,\ldots,M$, respectively. The independent events in (B.7), (B.8) and (B.9) define subsets of the $M+1$-dimensional sub-simplex (the support of the above distribution), so that they have positive probability under Π_2 as well as does (B.6).

References

Abramowitz, M and Stegun, I. A. (1964). *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, volume 55 of *National Bureau of Standards Applied Mathematics Series*. Superintendent of Documents, U.S. Government Printing Office, Washington, D.C..

Campbell, G. and Földes, A. (1982). Large-sample properties of nonparametric bivariate estimators with censored data. In *Nonparametric statistical inference, Vol. I, II (Budapest, 1980)*, volume 32 of *Colloq. Math. Soc. János Bolyai*, 103-121. North-Holland, Amsterdam.

Campbell, G. (1981)Nonparametric bivariate estimation with randomly censored data. *Biometrika*, **68**, 417-422.

Dabrowska, D. M. (1988). Kaplan-Meier estimate on the plane. *Ann. Statist.*, **16**, 1475-1489.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1**, 209-230.

Hanley J. A. and Parnes, M. N. (1983). Nonparametric estimation of a multivariate distribution in the presence of censoring. *Biometrics* **39**, 129-139.

Hjort, N. L. (1990). Nonparametric Bayes Estimators Based on Beta Processes in Models for Life History Data. *Ann. Statist.* **18**, 1259-1294.

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

Muliere, P., Secchi, P. and Walker, S. G. (2000). Urn schemes and reinforced random walks. *Stochastic Process. Appl.* **88**, 59-78.

Oakes, D. (1989). Bivariate survival models induced by frailties. *J. Amer. Statist. Assoc.* **84**, 487-493.

Prentice, R. L. and Cai, J. (1992). Covariance and survivor function estimation using censored multivariate failure time data. *Biometrika* **79**, 495-512.

Pruitt, R. C. (1991). On negative mass assigned by the bivariate Kaplan-Meier estimator. *Ann. Statist.* **19**, 443-453.

Schwartz, L. (1965). On Bayes procedures. *Z. Wahrsch. Verw. Gebiete* **4**, 10-26.

Tsai, W. Y., Leurgans, S. and Crowley, J. (1986). Nonparametric estimation of a bivariate survival function in the presence of censoring. *Ann. Statist.* **14**, 1351-1365.

van der Laan, M. J. (1996). Nonparametric estimation of the bivariate survival function with truncated data. *J. Multivariate Anal.* **58**, 107-131.

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

Wang W. and Wells, M. T. (1997). Nonparametric estimators of the bivariate survival function under simplified censoring conditions. *Biometrika* **84**, 863-880.

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

IMQ Università Bocconi, viale Isonzo 25, 20136 Milano, Italy.

E-mail: paolo.bulla@unibocconi.it

IMQ Università Bocconi, viale Isonzo 25, 20136 Milano, Italy.

E-mail: pietro.muliere@unibocconi.it

IMSAS University of Kent, Canterbury, CT2 7NZ, U.K.

E-mail: S.G.Walker@kent.ac.uk

(Received July 2005; accepted May 2006)