View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Warwick Research Archives Portal Repository

University of Warwick institutional repository: http://go.warwick.ac.uk/wrap

This paper is made available online in accordance with publisher policies. Please scroll down to view the document itself. Please refer to the repository record for this item and our policy information available from the repository home page for further information.

To see the final version of this paper please visit the publisher's website. Access to the published version may require a subscription.

Author(s): F Rigat and P Muliere
Article Title: Beta-Stacy survival regression models
Year of publication: 2007
Link to published article:
http://www2.warwick.ac.uk/fac/sci/statistics/crism/research/2007/paper07-6/
Publisher statement: None

Beta-Stacy survival regression models

Fabio Rigat*and Pietro Muliere†

April 11th 2007

Abstract

This paper introduces a class of survival models for discrete time-to-event data with random right censoring. Flexible distributions for the survival times are constructed by modelling the random survival functions as discrete-time beta-Stacy processes (DBS) and by introducing the regression effects via their prior means. Identifiability is attained by defining the DBS precision parameters as appropriate functions of the regression coefficients. By the conjugacy of the beta-Stacy process with respect to random right censoring, marginal posterior inferences for all model parameters can be efficiently approximated using the standard Gibbs sampler. The latter also yields a Monte Carlo approximation for the predictive distributions of the survival probabilities for future covariate profiles. We provide three clinical applications of the DBS survival regression framework comparing its estimates with those of parametric, semiparametric and non-parametric survival models.

Keywords: survival analysis, random right censoring, beta-Stacy process, Bayesian hierarchical models, Markov chain Monte Carlo, melanoma, cerebral palsy.

1 Introduction

This paper introduces a class of hierarchical regression models for discrete univariate time-to-event data with random right censoring. The distinctive feature of our approach is that the regression coefficients, the survival probabilities and their precision are represented by random parameters introduced via a hierarchical model structure. Our approach offers a flexible

*(Corresponding author) CRiSM, Department of Statistics, University of Warwick, U.K.;email: f.rigat@warwick.ac.uk; tel: +44(0) 24 7657 5754; fax: +44(0)24 7652 4532

†IMQ, Universita'Bocconi, Italy

CRiSM Paper No. 07-06v2, www.warwick.ac.uk/go/crism

alternative to mixture modelling (Farewell [1982], Farewell [1986], Jewell [1982], Ibrahim et al. [2001]) and to semiparametric models (Cox [1972], Kalbfleisch [1978], Sinha and Dey [1996], Kleinman and Ibrahim [1998], Walker and Mallick [1999], Kottas and Gelfand [2001], De Blasi and Hjort [2007]). An advantage of the models presented here with respect to mixture models is that inference and prediction are carried out within a fixed parameter space using the standard Gibbs sampler (Gelfand and Smith [1990]). Unlike the semiparametric proportional hazards model, the structure of our models is non-separable in the regression coefficients and in the random survival processes. This assumption is relaxed by modelling the latter as a set of discrete time beta-Stacy processes (DBS) whose prior mean incorporates a regression component (Cifarelli et al. [1981]). The beta-Stacy process is a generalization of the Dirichlet process, in that more flexible prior beliefs are able to be represented and, unilke the Dirichlet process, is conjugate to right censored observations. Connections with beta process introduced by Hjort [1990] are immediate (Walker and Muliere [1997]). To ensure identifiability of the models thus constructed we define the DBS precision parameter so that large regression effects correspond to posterior survival processes concentrated around their mean and vice versa. By the conjugacy of the DBS process the joint posterior distribution of our models' parameters factors into the product of the prior distributions for the regression coefficients, the marginal likelihood of the survival data and the conditional posterior density of the survival probabilities. Similar results are obtained under a Dirichlet process prior in absence of censored observations in Mira and Petrone [1996], Dominici and Parmigiani [2001] and Carota and Parmigiani [2002], following Antoniak [1974] and Cifarelli and Regazzini [1978].

The hierarchical DBS framework is used in this paper to analyse clinical survival data recording the occurrence of events related to a patient's status at discrete times such as days, weeks or months. In this context, the survival times can be thought of as taking values on a finite time grid of time points $\{\tau_1, \ldots, \tau_K\}$ fixed by design. Under this assumption, each individual's survival distribution is a finite-dimensional random process defined by the random heights of its jumps at the grid points. We show that as long as all the observed survival times are included in the grid the posterior estimates are not sensitive to changes of its resolution.

The paper is organized as follows. The DBS prior and the marginal distribution of the survival times are presented in Section 2. The derivation of the latter is reported in the Appendix. In Section 3 the hierarchical survival regression models are introduced along with a Markov chain algorithm to sample from the joint posterior distribution of their parameters

CRiSM Paper No. 07-06v2, www.warwick.ac.uk/go/crism

and from the predictive distributions under a Weibull prior mean. Section 4.1 illustrates the analysis of the photocarcinogenic study of Grieve [1987] using both proportional hazards and accelerated failure time Weibull DBS models. The inferences obtained using the proportional hazards DBS model are compared with those of Dellaportas and Smith [1993] and with the non-parametric Kaplan-Meier (Kaplan and Meier [1958]) estimates. In Section 4.2 the Danish melanoma dataset of Andersen et al. [1993] is analysed. Unlike the Cox proportional hazards model (Cox [1972]), the DBS analysis of this dataset uncovers the significance of the skin resistance to the tumor infiltration, the tumor thickness and the patients' age at surgery as independent prognostic factors. Section 4.3 illustrates the results of a DBS analysis of the cerebral palsy survival data of Hutton and Pharoah [2002]. We confirm their findings with regard to the non-linear effect of birth weight on the survival probability of the cerebral palsy patients. We also produce summaries of the predictive survival probabilities for the seven covariate profiles not included in the dataset. Should the next patient display one such covariate profiles, these predictions can be used for medical and legal purposes. Section 5 concludes the paper with a critical discussion of the DBS approach and by noting two possible generalizations.

2 The discrete beta-Stacy process

The beta-Stacy is a Lévy process introduced in Walker and Muliere [1997] as a non-parametric prior for the survival function of randomly right-censored survival times. Here we recall the constructive definition of its discrete-time version and we summarize its relevant properties for this work.

A scalar random variable $Y \in (0, \epsilon)$ has a beta-Stacy distribution, with non-negative parameters $(\alpha, \beta, \epsilon)$, if its probability density function is

$$f(Y \mid \alpha, \beta, \epsilon) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{Y^{\alpha-1}(\epsilon - Y)^{\beta-1}}{\epsilon^{\alpha+\beta-1}}.$$

Its mean and variance are

$$\begin{aligned} E(Y \mid \alpha, \beta, \epsilon) &= \frac{\alpha}{\alpha+\beta}\frac{1}{\epsilon}, \\ V(Y \mid \alpha, \beta, \epsilon) &= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}\frac{1}{\epsilon^2}. \end{aligned}$$

If Y is beta-Stacy then $Z = \frac{Y}{\epsilon}$ has a beta (α, β) distribution.

CRiSM Paper No. 07-06v2, www.warwick.ac.uk/go/crism

When each element of the ordered K-dimensional random vector $\{Y_k\}_{k=1}^K$ is conditionally independent beta-Stacy with parameters $(\alpha_k, \beta_k, 1-\sum_{j<k} Y_j)$, their joint distribution is generalized Dirichlet (Connor and Mosimann [1969]) with probability density function

$$f_Y(Y_1, \ldots, Y_K \mid \alpha_1, \ldots, \alpha_k, \beta_1, \ldots, \beta_K) \propto \prod_{k=1}^{k} Y_k^{\alpha_k - 1} \frac{(1 - \sum_{j \leq k} Y_j)^{\beta_k - 1}}{(1 - \sum_{j<k} Y_j)^{\alpha_k + \beta_k - 1}}. \quad (1)$$

The standard Dirichlet density obtains if $\beta_{k-1} = \beta_k + \alpha_k$ for $k = 1, ..., K$.

If $\{Y_k\}_{k=1}^K$ is jointly generalised Dirichlet and $S_k = 1 - \sum_{j \leq k} Y_j$, then $\{S_k\}_{k=1}^K$ is a discrete-time beta-Stacy process (DBS) with parameters $\{\alpha_k, \beta_k\}_{k=1}^K$, written as

$$\{S_k\}_{k=1}^K \sim \mathcal{DBS}\left(\{\alpha_k, \beta_k\}_{k=1}^K\right).$$

In what follows the DBS process is used as the prior distribution for the survival probabilities $\{S_k\}_{k=1}^K$ having random jumps $\{Y_k\}_{k=1}^K$ at the fixed time points $\tau = \{\tau_1, \ldots, \tau_K\}$.

Let $\underline{t} = \{t_i\}_{i=1}^N$ and $\underline{\delta} = \{\delta_i\}_{i=1}^N$ represent a sample of univariate random survival times with $\delta_i = 1$ if t_i is observed exactly and $\delta_i = 0$ if t_i is right-censored at random. Let τ be such that $\sum_k 1_{\{t_i = \tau_k\}} = 1$ for $i = 1, \ldots, N$. If the survival probabilities of each individual are a priori $\{S_{i,k}\}_{k=1}^K \sim \mathcal{DBS}(\{\alpha_{i,k}, \beta_{i,k}\}_{k=1}^K)$, by Theorem 1 of Walker and Muliere [1997] their posterior distributions are

$$\{S_{i,k}\}_{k=1}^K \mid t_i, \delta_i \sim \mathcal{DBS}\left(\{\alpha_{i,k} + n_{i,k}, \beta_{i,k} + m_{i,k}\}_{k=1}^K\right),$$

with

$$\begin{aligned} n_{i,k} &= 1_{\{t_i = \tau_k, \delta_i = 1\}}, \\ m_{i,k} &= 1_{\{t_i \geq \tau_k, \delta_i = 0\}} + 1_{\{t_i > \tau_k, \delta_i = 1\}}. \end{aligned}$$

The conjugacy of the beta-Stacy process with respect to random right censoring also yields a closed form expression for their marginal posterior predictive survival probabilities, that is

$$p(t_{N+1} = \tau_k \mid t_i, \delta_i, \{\alpha_{i,j}, \beta_{i,j}\}_{j=1}^k) = \frac{\alpha_{i,k} + n_{i,k}}{\alpha_{i,k} + \beta_{i,k} + n_{i,k} + m_{i,k}} \prod_{j<k} \frac{\beta_{i,j} + m_{i,j}}{\alpha_{i,j} + \beta_{i,j} + n_{i,j} + m_{i,j}}.$$

2.1 Marginal likelihood for DBS survival models

To construct a class of survival regression models using the discrete DBS process prior, we follow Walker and Muliere [1997] by letting $\{\alpha_{i,k}, \beta_{i,k}\}$ be

$$\begin{aligned} \alpha_{i,k} &= \nu_i(G_{i,k} - G_{i,k-1}), & (2) \\ \beta_{i,k} &= \nu_i(1 - G_{i,k}). & (3) \end{aligned}$$

CRiSM Paper No. 07-06v2, www.warwick.ac.uk/go/crism

for $i = 1, \ldots, N$ and $k = 1, \ldots, K$. Here $G_{i,k}$ represents the prior mean cumulative distribution function (CDF) at time τ_k for the survival time of individual i and the coefficient $\nu_i \geq 0$ represents their prior precision. When the DBS parameters are defined as in (2) and (3) we write

$$\{S_{i,k}\}_{k=1}^{K} \sim \mathcal{DBS}\left(\nu_i, \{G_{i,k}\}_{k=1}^{K}\right).$$

Under this parametrization the prior mean and variance of the increments of the survival function $Y_{i,k} = (S_{i,k} - S_{i,k-1}) = P(t_i = \tau_k)$ are (Connor and Mosimann [1969])

$$E(Y_{i,k} \mid G_{i,k}, G_{i,k-1}) = G_{i,k} - G_{i,k-1}, \tag{4}$$

$$V(Y_{i,k} \mid \nu_i, \{G_{i,j}\}_{j=1}^{k}) = (G_{i,k} - G_{i,k-1}) \times$$
$$\times \left(\tfrac{1+\nu_i(G_{i,k}-G_{i,k-1})}{\nu_i(1-G_{i,k-1})+1} \prod_{j<k} \tfrac{1+\nu_i(1-G_{i,j})}{1+\nu_i(1-G_{i,j-1})} - (G_{i,k} - G_{i,k-1}) \right), \tag{5}$$

and for $l < k$ their covariance is

$$Cov(Y_{i,l}, Y_{i,k} \mid \nu_i, \{G_{i,j}\}_{j=1}^{k}) = (G_{i,k} - G_{i,k-1}) \times$$
$$\times \left(\tfrac{1+\nu_i(G_{i,l}-G_{i,l-1})}{\nu_i(1-G_{i,l-1})+1} \prod_{j<l} \tfrac{1+\nu_i(1-G_{i,j})}{1+\nu_i(1-G_{i,j-1})} - (G_{i,l} - G_{i,l-1}) \right).$$

By equation (4) the prior survival probability for individual i at time τ_k is centered around its prior mean $(1 - G_{i,k})$. The right-hand side of equation (5) is decreasing in ν_i, which motivates the interpretation of the latter as the prior precision for the survival probabilities. By (2) and (3) we have $\alpha_{i,k} + \beta_{i,k} = \beta_{i,k-1}$, so that the survival probabilities for each individual are assigned a discrete Dirichlet process prior and have a discrete beta-Stacy posterior. Within this framework the prior coefficient ν_i can be interpreted as a measure of the prior strenght of belief in model $G_{i,k}$ (Ferguson [1974], Antoniak [1974], Ferguson and Phadia [1979], Kuo [1983], Brunner and Lo [1989], Muliere and Tardella [1998], Escobar [1994]).

Let $\mathbf{G} = \{\{G_{i,k}\}_{k=1}^{K}\}_{i=1}^{N}$ be a $N \times K$ matrix including the prior mean CDF values for all samples at τ. The marginal probability mass function of the survival times is

$$p(\underline{t} \mid \underline{\delta}, \mathbf{G}, \tau) = \prod_{i,k} \left(\tfrac{(G_{i,k}-G_{i,k-1})^{\delta_i}(1-G_{i,k})^{1-\delta_i}}{(1-G_{i,k-1})} \prod_{j<k} \tfrac{(1-G_{i,j})}{(1-G_{i,j-1})} \right)^{1_{\{t_i=\tau_k\}}}. \tag{6}$$

Equation (6) is derived in the Appendix. The analogy between (6) and parametric models which do not allow for the randomness of the survival function itself can be seen when $\tau_1 = 0$ and $\tau_k - \tau_{k-1} = \Delta_K = \frac{1}{K}$ for all $k = 2, ..., K$. When the prior mean CDF $G(\cdot)$ has a density with respect

CRiSM Paper No. 07-06v2, www.warwick.ac.uk/go/crism

to Lebesgue measure, if $K \to \infty$ the right-hand side of (6) approximates the likelihood function of the conditionally independent right-censored data $\underline{t}$ under model $G(\cdot)$.

3 Hierarchical DBS regression models

When the covariate profiles $X_i = [X_{i,1}, \ldots, X_{i,q}]$ are given for each sample unit $i = 1, \ldots, N$, a regression component is incorporated in the prior mean model $G(\cdot)$ by letting

$$G_{i,k}(X_i, \theta, \eta) = G(t_i \leq \tau_k \mid X_i, \theta, \eta),$$

where θ represents a $q \times 1$ vector or regression effects and η are additional parameters indexing the prior mean CDF. We propose constructing flexible survival models by letting the distribution of the survival data t_i at the time points τ be represented by the DBS random survival probabilities $\{S_{i,k}\}_{k=1}^{K}$ centered around their mean survival function $1 - G_{i,k}(X_i, \theta, \eta)$ with precision ν_i. However, if both the coefficients (θ, η) indexing $G(\cdot)$ and the DBS precision parameters $(\nu_1, ..., \nu_N)$ are unknown, such models are not identifiable from the data. This can be seen by considering that under (2) and (3) the generalised Dirichlet density for the survival probabilities of individual i may not be integrable with respect to ν_i. The same issue has been noted by Dominici and Parmigiani [2001] and by Carota and Parmigiani [2002] with regard to the total mass parameter of their Dirichlet process priors. In this paper we identify the hierarchical DBS models by defining the covariate-specific precisions as

$$\nu_i = (X_i \theta)^2. \tag{7}$$

By (7) when the regression coefficients are large in absolute value the DBS random survival probabilities are concentrated around their mean and vice versa. Therefore, under (7) the random survival probabilities have a prominent role to fit the survival data only when their relationship with the covariates is weak under model $G(\cdot)$.

The hierarchy of the DBS survival regression models is completed by the specification of the prior distributions for the coefficients (θ, η). By letting $\mathbf{S} = \{\{S_{i,k}\}_{k=1}^{K}\}_{i=1}^{N}$ and by assuming that the survival times $\underline{t}$ are independent random variables conditionally on their survival probabilities,

CRiSM Paper No. 07-06v2, www.warwick.ac.uk/go/crism

the hierarchy can be written as

$$
\begin{aligned}
(\theta, \eta) &\sim f_{\theta,\eta}(\theta, \eta), \\
\{S_{i,k}\}_{k=1}^{K} &\sim \mathcal{DBS}\left((X_i\theta)^2, \{G(t_i \leq \tau_k \mid X_i, \theta, \eta)\}_{k=1}^{K}\right) \text{ for } i = 1, \ldots, N, \\
P(\underline{t} \mid \underline{\delta}, \mathbf{S}, \tau) &= \prod_{i,k} \left((S_{i,k-1} - S_{i,k})^{\delta_i} S_{i,k}^{1-\delta_i}\right)^{1_{\{t_i=\tau_k\}}}, \qquad (8)
\end{aligned}
$$

with $S_{i,0} = 1$ for all $i = 1, ..., N$. We adopt two alternative formulations for the regression component under a Weibull prior mean survival function G, namely

$$G_{PH}(t \leq \tau \mid X\theta, \eta) = 1 - e^{-e^{X\theta}\tau^{\eta}}, \qquad (9)$$

$$G_{AFT}(t \leq \tau \mid X\theta, \eta) = 1 - e^{-\left(e^{X\theta}\tau\right)^{\eta}}, \qquad (10)$$

where X is the $N \times q$ matrix which ith row X_i represents the covariate profile for individual i. Here the Weibull distribution is adopted for the availability of a closed form expression for the survival function $1 - G(\cdot)$. Equation (9) assumes that the prior mean cumulative hazard processess for individuals having different covariate profiles are proportional at different time points (Cox [1972]). Equation (10) adopts an accelerated failure time regression (Prentice and Kalbfleisch [1979]) where the survival time is shifted along the time axis proportionally to a stress factor. The latter is represented by the exponential of the linear predictor $X\theta$ as in Walker and Mallick [1999]. The coefficient η in (9) and in (10) is the Weibull index parameter, which determines the convexity of the prior mean survival distribution for all individuals over time.

3.1 Parameter estimation

Under (9) and (10) the unknown parameters of the DBS survival models are $(\theta, \eta, \mathbf{S})$. A key feature of (8) is that the joint posterior density factors as

$$f(\theta, \eta, \mathbf{S} \mid \underline{t}, \underline{\delta}, X) = f(\theta, \eta) p(\underline{t} \mid \underline{\delta}, \mathbf{G}(X\theta, \eta), \tau) f_{\mathbf{S}}(\mathbf{S} \mid \underline{t}, \underline{\delta}, \mathbf{G}(X\theta, \eta)), \qquad (11)$$

where $\mathbf{G}(X\theta, \eta) = \{\{G_{i,k}(X_i\theta, \eta)\}_{k=1}^{K}\}_{i=1}^{N}$, $p(\underline{t} \mid \underline{\delta}, \mathbf{G}(X\theta, \eta), \tau)$ is the marginal likelihood of the survival data (6), $f(\theta, \eta)$ represents the joint prior density for (θ, η) and $f_{\mathbf{S}}(\mathbf{S} \mid \underline{t}, \underline{\delta}, \mathbf{G}(X\theta, \eta))$ is the product of the generalized Dirichlet posterior densities for the survival probabilities of the N conditionally independent samples. Similar factorizations can be found in Mira and Petrone [1996], Carota and Parmigiani [2002] and Dominici and Parmigiani [2001]. The analogy between their models and the hierarchy (8) is the availability

CRiSM Paper No. 07-06v2, www.warwick.ac.uk/go/crism

of a closed-form expression for the marginal likelihood of the data conditionally on the parameters of their random distributions. In previous works the marginal likelihood for exact data is obtained using a Dirichlet process prior whereas in this work equation (6) was derived for discrete randomly right-censored survival times using the conjugacy of the DBS process.

In what follows we let the priors for the coefficients θ and η be respectively Gaussian $N(m, s)$ with fixed mean m and standard deviation s and gamma $Ga(\frac{a^2}{b}, \frac{b}{a})$, having mean a and variance b. To elicit the prior hyper parameters (m, s, a, b) we consider the corresponding marginal prior predictive survival processes. Since the generalised Dirichlet density cannot be marginalised analytically with respect to the priors for (θ, η) we use a Monte Carlo strategy. For given values of (m, s, a, b) we sample an array of realisations of (θ, η) from their priors and we generate corresponding realisations of the DBS prior survival processes for each distinct covariate profile. The sample distribution of the generated survival probabilities approximates their marginal prior distribution, thus representing the marginal effect of a given configuration (m, s, a, b) on the prior predictive survival functions. In what follows a set of values for (m, s, a, b) is adopted if the corresponding prior predictive mean survival probabilities at τ and their 95% prior probability intervals are considered appropriate for the data to be analysed.

The conditional posterior densities for (θ, η) are log-concave but they are not available for exact sampling, so that approximate marginal posterior inferences can be computed via the Gibbs sampler (Gelfand and Smith [1990], Tierney [1998]) using a Metropolis-Hastings (MH) rejection step (Hastings [1970]). The posterior inferences reported in Section 4 were computed using a component-wise random walk MH update for (θ, η). For each component of θ we employ a Gaussian proposal density centerd at its current value whereas for the Weibull index η we use a gamma random walk proposal of the form $Ga(\frac{\eta_w^2}{C_\eta}, \frac{C_\eta}{\eta_w})$ where η_w represents the current value of η and C_η is a fixed coefficient. Under this parametrization the proposal mean is η_w and its variance is C_η. The survival probabilities $\mathbf{S}$ are updated exactly within the Gibbs sampler using the constructive definition of the DBS process of Section 2. When the covariate profiles of all N samples are distinct, the parameters of their generalized Dirichlet conditional posterior densities are

$$\begin{aligned}
\alpha^*_{i,k} &= (X_i\theta)^2 \left(G_{i,k}(X_i\theta, \eta) - G_{i,k-1}(X_i\theta, \eta)\right) + \delta_i 1_{\{t_i=\tau_k\}}, \\
\beta^*_{i,k} &= (X_i\theta)^2 \left(1 - G_{i,k}(X_i\theta, \eta)\right) + \delta_i 1_{\{t_i>\tau_k\}} + 1_{\{t_i\geq\tau_k, \delta_i=0\}}.
\end{aligned}$$

If X is a design matrix defining groups of observations, in (8) all observations within the same group share the same survival probabilities. In such a case

CRiSM Paper No. 07-06v2, www.warwick.ac.uk/go/crism

let $\{g(i)\}_{i=1}^{N}$ be the group label of the ith observation, with $1 \leq g(i) \leq N$ for all values of i. The posterior parameters of the DBS process for the survival probabilities for group $g(i)$ are

$$\alpha^*_{g(i),k} = (X_{g(i)}\theta)^2 \left(G_{i,k}(X_{g(i)}\theta, \eta) - G_{i,k-1}(X_{g(i)}\theta, \eta)\right) + \sum_{i=g(i)} \delta_i 1_{\{t_i=\tau_k\}}, \quad (12)$$

$$\beta^*_{g(i),k} = (X_{g(i)}\theta)^2 \left(1 - G_{i,k}(X_{g(i)}\theta, \eta)\right) + \sum_{i=g(i)} \delta_i 1_{\{t_i>\tau_k\}} + 1_{\{t_i\geq\tau_k, \delta_i=0\}}. \quad (13)$$

3.2 Predictions

Let t_{N+1} be the unknown exact survival time for a future sample with covariates $X_{g(N+1)}$. According to the hierarchy (8), if $X_{g(N+1)} = X_{i^*}$ for some $i^* \in [1, N]$, the distribution of the survival probabilities $(S_{N+1,1}, ..., S_{N+1,K})$ coincides with that of $(S_{i^*,1}, ..., S_{i^*,K})$, so that summaries of their Gibbs sampler draws provide approximate marginal posterior predictions for t_{N+1} at the time points τ. When $X_{g(N+1)}$ does not coincide with any of the observed covariate profiles, approximate marginal posterior predictions can be computed using the Gibbs sampler draws $\{\theta_w, \eta_w\}_{w=1}^{M}$. For instance, within the DBS framework the marginal posterior predictions for the survival time of patient $(N + 1)$ with covariates $X_{g(N+1)}$ at time τ_k are summaries of the random variable $S_{N+1,k}$ given $(\underline{t}, \underline{\delta}, X, X_{g(N+1)})$. Its conditional posterior distribution can be sampled exactly by drawing a realization of the increments $(S_{N+1,1}, S_{N+1,2} - S_{N+1,1}, ..., S_{N+1,k} - S_{N+1,k-1})$ for each couple (θ_w, η_w) using their generalised Dirichlet joint posterior distribution. Summaries of the sequence $\{S_{N+1,k}(w)\}_{w=B+1}^{M}$ thus generated provide a Monte Carlo approximation of its marginal posterior moments. These summaries can be used to evaluate the survival perspective of a future patient which clinical profile $X_{g(N+1)}$ has not been observed in the past. If $X_{g(N+1)}$ includes covariates which value can be controlled, such as treatments, estimates of $S_{N+1,k}$ under alternative scenarios indicate which combinations of values of $X_{g(N+1)}$ correspond to the highest survival rates at time τ_k having observed $(\underline{t}, \underline{\delta}, X)$.

4 Applications of the model

4.1 Analysis of the mice dataset

The study reported in Grieve [1987] includes the survival times of 80 mice divided in four groups of 20 individuals. Each group was given a different

CRiSM Paper No. 07-06v2, www.warwick.ac.uk/go/crism

photocarcinogenic treatment to assess whether pre-treatment with a test substance (8-MOP) shortens the time to occurrence of skin tumors (Grieve [1987]). This data have been analysed via a proportional hazards Weibull survival model by Grieve [1987] using the numerical integration techniques of Naylor and Smith [1982] and by Dellaportas and Smith [1993], who fitted the same model using the Gibbs sampler. In this Section we analyse the mice data via the DBS Weibull survival models using the PH and AFT regressions (9) and (10) under two specifications of the time grid τ. The latter include respectively the months $(1, ..., 40)$ and all time points between 1 and 40 months equally spaced by 0.1 months. The first specification of τ matches the resolution and range of the survival data, which facilitates the comparison between the DBS estimates of the survival probabilities and the nonparametric Kaplan-Meier (KM; Kaplan and Meier [1958]) estimates. The second time grid represents a ten-fold increase in the number of time points over the same range. The comparison of the posterior estimates for the two specifications of τ informs on the sensitivity of the posterior distributions of $(\theta, \eta, \mathbf{S})$ with respect to the resolution of the time grid. Posterior sampling was carried out for fifty thousand iterations via the Gibbs sampler described in Section 3. All posterior estimates were computed using the last twenty five thousand samples.

Figure 1 shows the KM estimates of the survival curves for the four groups along with their 95% point-wise confidence intervals. The estimated group medians in months are respectively $21(13, 23)$ for group 1 (irradiated controls), $22(15, 26)$ for group 2 (test substance: 8-methoxypsoralen), $18(17, 22)$ for group 3 (positive controls) and $30(27, 32)$ for group 4 (vehicle controls), implying that the survival perspective for the mice belonging to the vehicle control group is better that those of the other groups.

Figure 2 displays the Monte Carlo summaries of the prior predictive mean survival probabilities for all groups and their 95% highest probability density (HPD) intervals under the PH and the AFT regressions. Prior summaries were computed for two sets of values of (m, s, a, b), that are respectively $(0, 1, 1, 1)$ for the plots in the first row and $(-3, 1, 1, 0.1)$ for the second row. The top plots in Figure 2 indicate that under the first prior setting all mice are expected to die within two months from the beginning of the experiment. This emphasizes that an apparently weakly informative prior for (θ, η) can have unforseen consequences for the prior distribution of the DBS survival probabilities. For the following analyses we adopt the second set of values

CRiSM Paper No. 07-06v2, www.warwick.ac.uk/go/crism

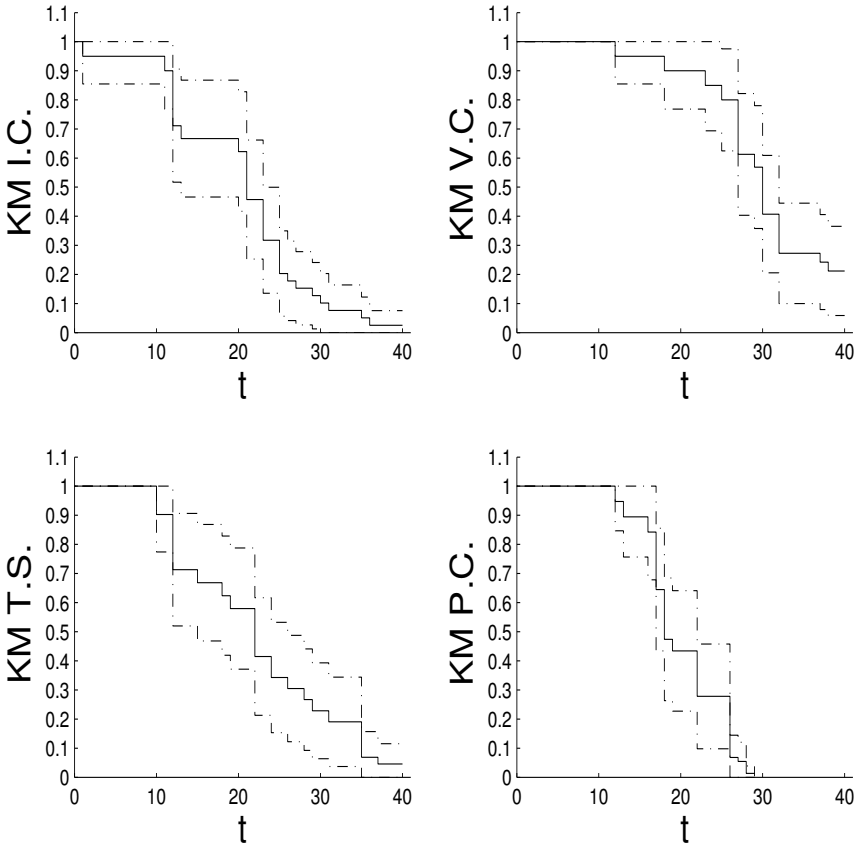

Figure 1: Kaplan-Meier estimates of the survival probabilities for the four groups composing the mice dataset (I.C. = irradiated control, T.S. = test substance (8-MOP), P.C. = positive control, V.C. = vehicle control). Their 95% confidence intervals were computed using Greenwood's formula. The estimates show that the survival perspective for the mice belonging to the vehicle control group is better that those of the other groups.

for (m, s, a, b), which allows an average $10 - 15\%$ of all mice to survive until month 40 irrespectively of their treatment.

Tables 1 and 2 report the posterior estimates for (θ, η) and for the median survival times of the four mice groups. Point estimates are represented by the means of their Gibbs sampler draws and the interval estimates include 95% of the sampled values. The ceontrla column in table 1 and the left-

CRiSM Paper No. 07-06v2, www.warwick.ac.uk/go/crism

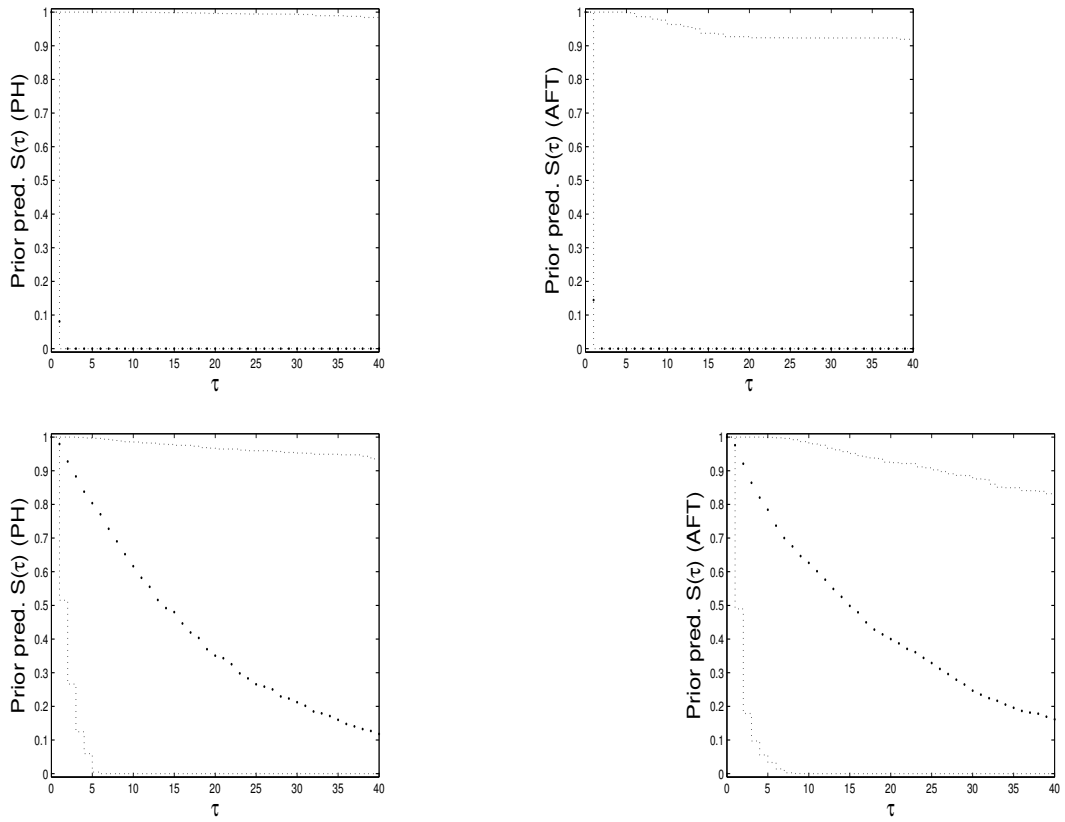

Figure 2: prior predictive sample mean survival probabilities and their 95% prior predictive HPDs under two alternative values of the prior hyper-parameters for (θ, η) (respectively $m = 0, s = 1, a = 1, b = 1$ in the first row and $m = -2, s = 1, a = 1, b = 0.1$ in the second row). The prior predictive summaries were computed using samples of size 10000 from the marginal priors for (θ, η). The second prior setting was adopted for the DBS analysis of the mice data.

most column in table 2 correspond to the coarser time grid. The difference between the estimates reported in these two tables reflects the structural difference between the PH and the AFT regressions. Under both regressions positive values of the regression coefficients θ worsen the mean survival probabilities and vice versa. Table 1 also reports the posterior estimates obtained using the Weibull survival regression model of Dellaportas and Smith [1993] (DS) using the same priors for (θ, η). Under all models the estimates for the membership to the vehicle control group are larger in absolute value with

CRiSM Paper No. 07-06v2, www.warwick.ac.uk/go/crism

respect to those of the other three groups, implying that the survival probabilities of the individuals belonging to group 4 are higher than those of the other groups. This conclusion is consistent with the KM estimates shows in Figure 1. The large overlapping between the estimated 95% posterior intervals reported in Tables 1 and 2 suggests that the higher resoultion of the time grid τ does not affect significantly the posterior estimates of (θ, η). Figures 3 through 6 illustrate the posterior estimates of the survival proba-

	PH	PH_{10}	DS
$\theta_{I.C.}$	$-4.65(-5.41, -3.96)$	$-4.46(-5.24, -3.77)$	$-4.63(-5.41, -3.84)$
$\theta_{T.S.}$	$-4.90(-5.77, -4.09)$	$-4.66(-5.43, -3.90)$	$-4.86(-5.69, -4.04)$
$\theta_{P.C.}$	$-4.60(-5.33, -3.89)$	$-4.40(-5.20, -3.95)$	$-4.60(-5.37, -3.80)$
$\theta_{V.C.}$	$-5.42(-6.22, -4.60)$	$-5.17(-6.10, -4.45)$	$-5.37(-6.26, -4.53)$
η	$1.47(1.26, 1.68)$	$1.40(1.21, 1.60)$	$1.92(1.41, 2.44)$
I.C.	$23(16, 27)$	$22(16, 27)$	$18.75(13.95, 25.13)$
T.S.	$24(18, 30)$	$24(17.6, 31)$	$21.95(16.04, 30.13)$
P.C.	$21(17, 26)$	$20.6(17, 26)$	$18.15(13.26, 24.81)$
V.C.	$32(27, 40)$	$32(27, 40)$	$31.47(22.41, 44.50)$

Table 1: posterior estimates of (θ, η) and posterior predictive median survival times obtained using the DBS model with PH regression and the DS model. The estimated regression coefficient $\theta_{V.C.}$ is lower than the estimates of the other groups, implying that the posterior mean survival probabilities of the individuals belonging to the vehicle control group are higher than those of the other groups.

bilities for the four mice groups. The central dots represent their estimated posterior means and the dashed lines mark the end-points of their estimated 95% posterior intervals. Despite of the difference between the spread of the prior predictive distributions between regression formulations shown in Figure 2, the posterior spread within each group is similar across regressions and across resolutions of τ, indicating that the values of (m, s, a, b) chosen for our analysis did not overwhelm the data. Figures $3 - 6$ also show that the DBS estimates of the survival probabilities follow the shape of their Weibull centering distribution over the range of τ when no deaths are observed whereas they display large jumps at the time points where exact failures occurr. Therefore, unlike models which do not allow for the randomness of the survival function itself, the DBS models allow for a different degree of smoothness of the fitted survival probabilities depending on the location of the observed survival times. Unlike semiparametric survival regressions, when no data are observed the DBS survival probabilities follow

CRiSM Paper No. 07-06v2, www.warwick.ac.uk/go/crism

	AFT	AFT_{10}
$\theta_{I.C.}$	$-3.25(-3.46,-3.05)$	$-3.26(-3.47,-3.08)$
$\theta_{T.S.}$	$-3.38(-3.62,-3.17)$	$-3.39(-3.62,-3.19)$
$\theta_{P.C.}$	$-3.15(-3.37,-2.95)$	$-3.18(-3.40,-2.98)$
$\theta_{V.C.}$	$-3.69(-3.93,-2.95)$	$-3.70(-3.94,-3.48)$
η	$2.92(1.85,2.78)$	$2.33(1.87,2.83)$
I.C.	$23(20,27)$	$23(20,27.2)$
T.S.	$26(19,32)$	$26(19.3,31)$
P.C.	$22(18,26)$	$22(18,26)$
V.C.	$32(28,40)$	$32(28.1,40)$

Table 2: posterior estimates of (θ,η) and posterior predictive median survival times obtained using the DBS model with AFT regression. Consistently with the estimates of the PH model, the AFT model confirms that the mean mortality of the vehicle control group is the lowest.

their covariate-dependent Weibull centering distribution.

4.2 Analysis of the Danish mealanoma survival data

Andersen et al. [1993] describe a Danish prospective study run between 1962 and 1978 reporting the survival times of 225 patients whose malignant melanomas were removed by surgery. For most patients a set of fixed-time predictors were also successfully measured. The predictors are the tumor depth (3 ordered levels), the level of resistance to the tumor infiltration (4 ordered levels) presence or absence of epithelioid cells, the presence or absence of skin ulceration, the tumor thickness (ranging from 0.10mm to 17.42mm), the patients' gender and their age at the time of surgery. The dataset is affected by heavy right censoring with 57 exact observations and 148 right-censored observations. A censored observation was recorded when a patient had not yet perished as a consequence of the melanoma at the end of the study. The 14 patients belonging to this sample who deceased

CRiSM Paper No. 07-06v2, www.warwick.ac.uk/go/crism

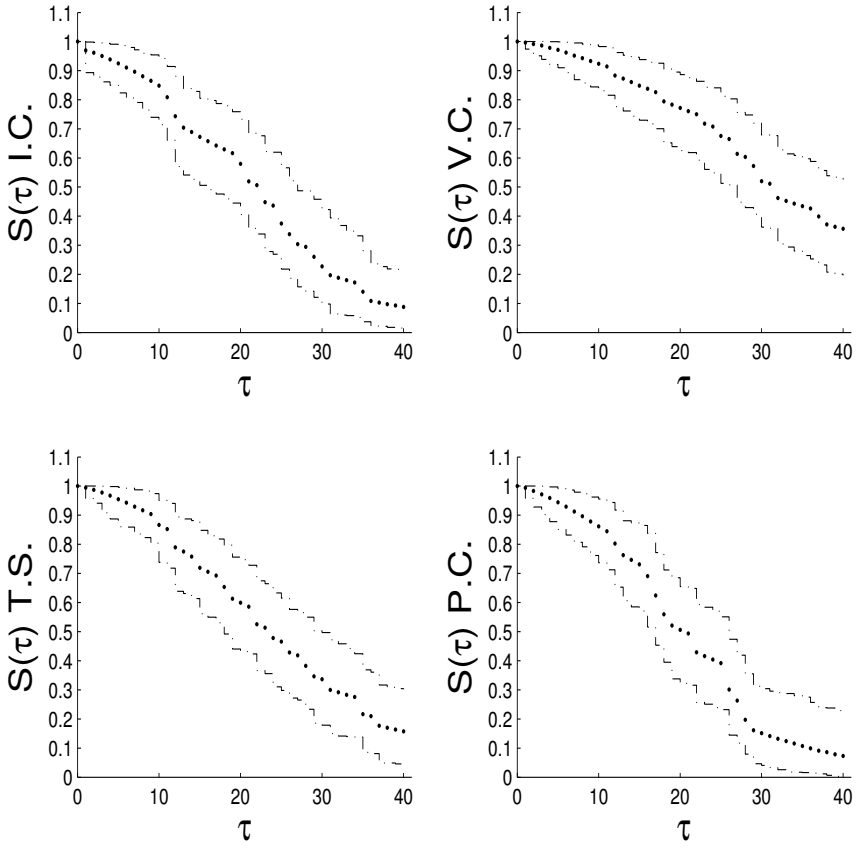

Figure 3: posterior estimates for the DBS survival probabilities for the four mice groups under PH regression and coarser time grid.

for causes other than the melanoma are considered censored in the following analysis. This data has been extensively studied in Andersen et al. [1993] emphasizing a worse survival perspective of the males versus females, an increased mortality for older patients, for patients having thick tumors and for those having skin ulceration. Tumor thickness, the skin ulceration status and the level of resistance are among the factors considered by the current American Joint Committee on Cancer (AJCC) staging system. The tumor thickness appears also to be a relevant predictor of the risk of local recurrence for cutaneous melanomas (Thompson et al. [2005]). Andersen et al. also emphasize that the hazards among different patients groups are not proportional over time.

CRiSM Paper No. 07-06v2, www.warwick.ac.uk/go/crism

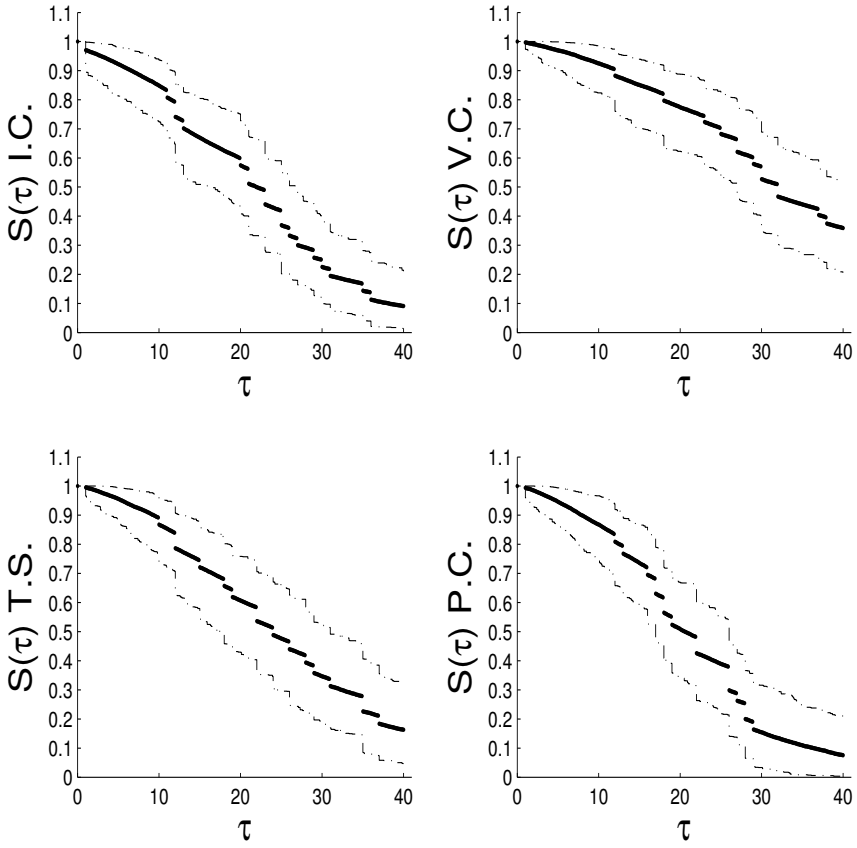

Figure 4: posterior estimates for the DBS survival probabilities for the four mice groups under PH regression and finer time grid.

In this Section we analyse the survival data of the 201 patients having a complete clinical profile using the DBS model with PH regression (9). We let τ concide with the observed survival times in months plus years 1 through 20. Posterior sampling was carried out using the Gibbs sampler described in Section 3 for fifty thousand iterations. Posterior estimates were computed using the last ten thousand draws. The patients' age at surgery and tumor thickness were treated as factors with two levels each. The cutoffs defining these levels are their sample medians, which are respectively 54 years of age and 194 millimeters thickness. The remaining covariates were collapsed into two ordered categories (low/high). The prior parameters for (θ, η) are $m = -1, \sigma = 5, a = 1, b = 0.1$ corresponding to an approximately uniform

CRiSM Paper No. 07-06v2, www.warwick.ac.uk/go/crism

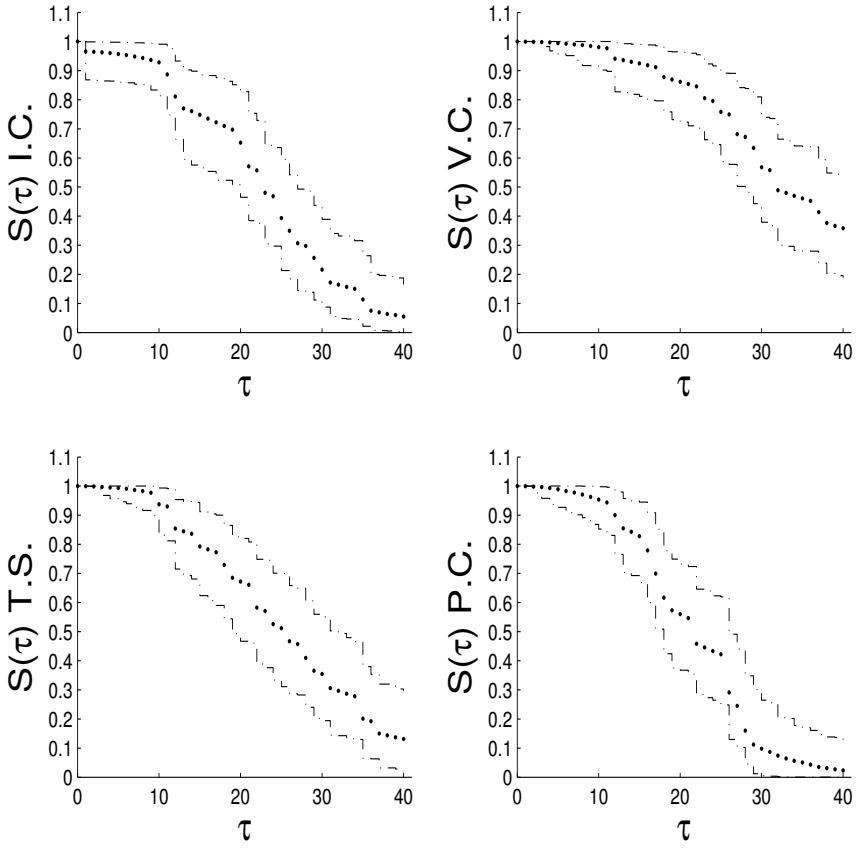

Figure 5: posterior estimates for the DBS survival probabilities for the four mice groups under AFT regression and coarser time grid.

distribution for the prior predictive medians for the survival times of all individuals over the range of the data ($(0, 183)$ months).

The central column in table 3 reports the estimated posterior means and the end-points of the estimated 95% posterior intervals for the regression parameters θ and for the Weibull index η. Its right-most column reports the estimates for the regression effects of the semiparametric proportional hazards model (Cox [1972]). and the end-points of their 95% confidence intervals. All covariates but the ulceration status are significant predictors of the survival time according to the DBS estimates, whereas only the tumor depth and the patients' age are found significant by the Cox model. This difference is partially explained by the fact that the hazards among different

CRiSM Paper No. 07-06v2, www.warwick.ac.uk/go/crism

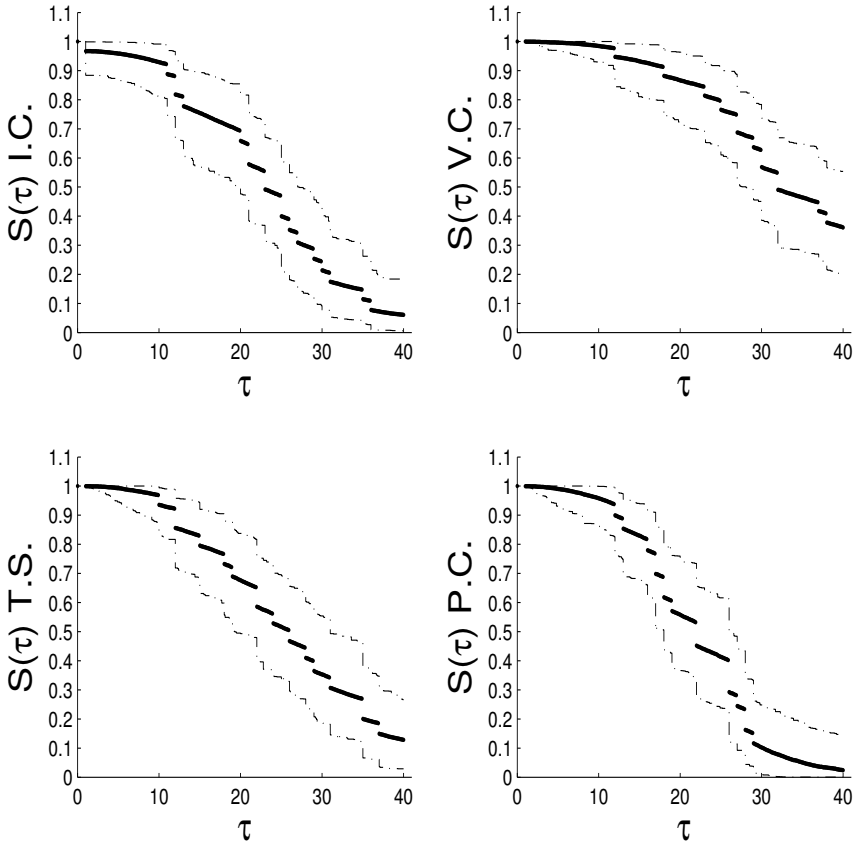

Figure 6: posterior estimates for the DBS survival probabilities for the four mice groups under AFT regression and finer time grid.

data groups may not be proportional, as pointed out in Andersen et al. [1993].

The lowest DBS estimated median survival probability at five years after surgery corresponds to deep thick uninfiltrated melanomas with no epithelioid cells for young male patients ($0.42(0.2, 0.82)$) whereas the highest survival corresponds to superficial thin uninfiltrated tumors with no epithelioid cells for young female patients ($0.94(0.87, 1.00)$). Figure 7 shows the posterior predictive median survival probabilities for four hypothetical future patients. Should any of such patients represent the next actual case of melanoma, the estimates depicted in Figure 7 provide an indication of their likely survival prognosis based on the available past observations. The four

CRiSM Paper No. 07-06v2, www.warwick.ac.uk/go/crism

patients are young females presenting uninfiltrated thick melanomas without epithelioid cells. The predictive survival probabilities are higher for patients displaying deep tumors with respect to cases of superficial melanomas. The highest median survival at five years after surgery correspond to deep melanomas with or without skin ulceration (respectively $0.65(0.43, 0.97)$ and $0.87(0.50, 0.99)$) whereas the lowest estimated survival probabilities correspond to superficial tumors with or without skin ulceration (respectively $0.03(0.01, 0.10)$ and $0.06(0.02, 0.22)$).

Coefficient	DBS post. est.	Cox PH
θ_{depth}	$-0.44(-0.92, -0.01)$	$-0.54(-1.03, -0.05)$
$\theta_{resistance}$	$-1.16(-1.53, -0.78)$	$-0.01(-0.35, 0.35)$
$\theta_{epithelioid}$	$-1.14(-1.55, -0.75)$	$-0.16(-0.52, 0.19)$
$\theta_{ulceration}$	$0.10(-0.35, 0.53)$	$0.06(-0.30, 0.43)$
θ_{thick}	$0.61(0.06, 1.15)$	$0.87(-0.36, 1.38)$
θ_{male}	$-0.75(-1.20, -0.34)$	$-0.08(-0.44, 0.28)$
θ_{age}	$-0.91(-1.36, -0.53)$	$-0.79(-1.15, -0.42)$
η	$0.22(0.18, 0.26)$	-

Table 3: DBS posterior estimates of (θ, η) for the melanoma survival data under the PH regression and the Cox semiparametric survival regression model. According to the DBS estimates all covariates but the ulceration status are significant predictors of the survival time, whereas only the tumor depth and the patients' age are found significant by the Cox model.

4.3 Cerebral palsy survival times

Hutton and Pharoah [2002] present an accelerated failure time analysis of a set of cerebral palsy survival times of patients born between 1966 and 1989 in Merseyside and Cheshire. We illustrate an analysis of the same dataset using the DBS model with Weibull AFT regression (10). We consider the 1585 survival times for which the five available fixed-time covariates are recorded: ambulatory (a), manual (ma), mental (me) impairment, sight quality (s) and birth weight. The right-censored times amount to the 84.5% of this data. The mean survival time for all patients is 24.5 years and 95% of their survival times are included in the interval $(2.10, 39.1)$ years. We evaluate the DBS survival probabilities over a time grid τ including all observed survival times plus the years $[1, 2, ..., 40]$. To control for a non-linear

CRiSM Paper No. 07-06v2, www.warwick.ac.uk/go/crism

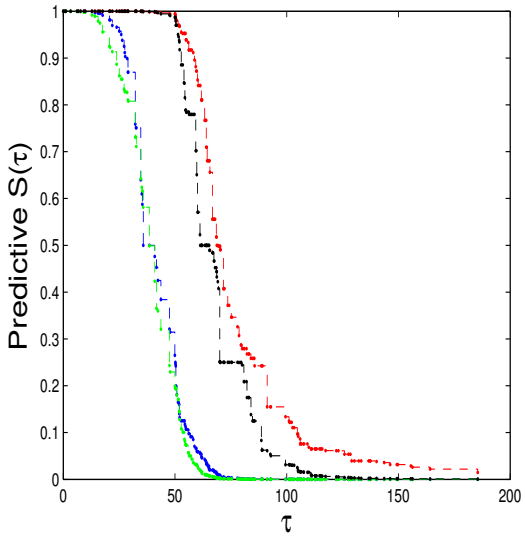

Figure 7: estimated marginal posterior predictive median survival probabilities at five years from surgery for four hypothetical future patients whose covariate profiles are not included among the 201 melanoma patients. The four patients are young females presenting uninfiltrated thick melanomas without epithelioid cells. The highest median survival probabilities correspond to deep melanomas with $(0.65(0.43, 0.97))$ or without $(0.87(0.50, 0.99))$ skin ulceration whereas the lowest estimated survival probabilities correspond to superficial tumors with $(0.03(0.01, 0.10))$ or without $(0.06(0.02, 0.22))$ skin ulceration.

effects of birth weight found by Hutton and Pharoah [2002] this predictor was categorized in three classes, based on its 33rd and 66th percentiles, which are: $[580, 2264)$, $[2264, 3147)$ and $[3147, 5260)$ grams. The prior for the regression parameters θ and for the Weibull index η are respectively $N(-1, 1)$ and $Ga(1, 0.1)$. As in the previous example, these priors correspond to approximately uniformly distributed prior predictive median survival times over the range of the observed survival times. Posterior sampling was carried out for twenty five thousand iterations using the Gibbs sampler described in Section 3. Posterior estimates and predictions were computed using the last ten thousand iterations.

Table 4 reports the estimated posterior means and the estimated 95% posterior intervals for the regression coefficients θ and for the Weibull index η. The occurrence of any type of impairment has an adverse effect on the mean survival probabilities. All other predictors being constant, the sur-

CRiSM Paper No. 07-06v2, www.warwick.ac.uk/go/crism

vival probabilities for individuals with weight at birth between $[3147, 5260)$ grams is the lowest whereas those of individuals with birth weight between $[580, 2264)$ grams is the highest. The difference between estimates of the birth weight categories indicate a non-linear effect on the mean survival probabilities consistenly with the results of Hutton and Pharoah [2002]. The estimated Weibull index parameter η indicates that an exponential mean model is adequate for this dataset. Figure 8 displays the posterior

Coefficient	Post.Estimates
θ_a	$0.87(0.41, 1.34)$
θ_{ma}	$1.18(0.73, 1.58)$
θ_{me}	$0.60(0.20, 1.00)$
θ_s	$0.85(0.57, 1.14)$
θ_{low}	$-6.65(-7.06, -6.27)$
θ_{med}	$-6.48(-6.86, -6.13)$
θ_{high}	$-6.43(-6.79, -6.08)$
η	$1.08(0.99, 1.18)$

Table 4: posterior estimates of the regression parameters and for the Weibull index for the cerebral palsy dataset. The occurrence of any impairment worsens the survival perspective, with manual impairment being the most severe. The effect of the birth weight on the mean survival probabilities appears to be non-linear. All other factors being constant, individuals with low birth weight ($[580, 2264)$ grams) have the best survival whereas those with high birth weight ($[3147, 5260)$ grams) have the worse perspective.

mean survival probabilities and their 95% posterior intervals for the best and the worst case in the dataset. The latter displays ambulatory, manual and mental impairments and has medium birth weigth whereas the former does not present any impairment at the same birth weight.

Among the possible combinations of covariate values, seven profiles do not correspond to any of the 1585 recorded patients. Table 6 reports the posterior estimates of the survival probabilities for all seven cases at 24.5 years from birth, which coincides with the mean survival time for all patients. The rows of Table 6 are ordered so that the predicted median survival is non-increasing. Should the next patient display one of such covariate profiles, these predictions provide an indication of her likely survival time based on the available past observations. When the occurrence of cerebral palsy

CRiSM Paper No. 07-06v2, www.warwick.ac.uk/go/crism

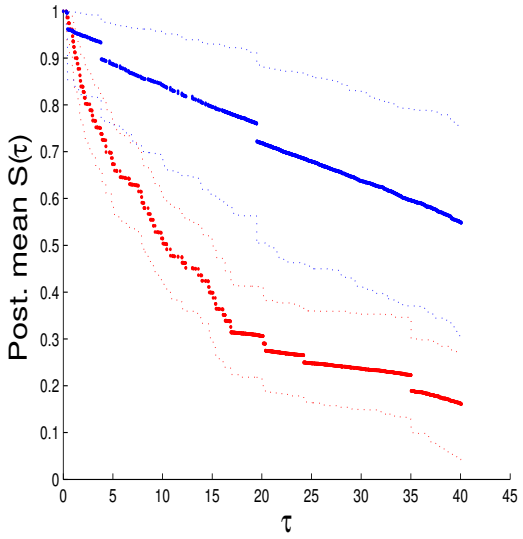

Figure 8: best and worst estimated survival probabilities among the 1585 cerabral palsy patients. The central dots represent the estimated posterior mean survival probabilities at the grid points τ. The dashed lines represent the survival probabilities' estimated 95% posterior intervals. The worse survival prespective is associated to ambulatory, manual and mental impairments and has medium birth weigth whereas the covariate profile corresponding to the best survival does not present any impairment at the same birth weight.

is due to admitted medical malpractice, these predictions can provide a reference to determine the quantum of a compensation based on the child's survival probabilities. The highest predicted survival corresponds to future individuals having ambulatory impairment only and birth weight within $[3147, 5260)$ grams, whereas the worse predictions correspond to individuals within the same birth weight categoty and displaying manual, mental and sight impairments.

5 Discussion

This paper introduces the hierarchical DBS survival regressions as a flexible and interpretable modelling framework for right-censored survival times. The DBS regression effects have analogous interpretations to those of simpler survival regressions such as that of Dellaportas and Smith [1993]. The random DBS survival probabilities at the time points τ effectively represent

CRiSM Paper No. 07-06v2, www.warwick.ac.uk/go/crism

Profile	Est. $S(24.5)$
a, $[3147, 5260)$gr	$0.90(0.73, 0.98)$
a,s, $[580, 2264)$gr	$0.74(0.46, 0.93)$
a,s, $[3147, 5260)$gr	$0.69(0.41, 0.89)$
ma,s, $[580, 2264)$gr	$0.68(0.36, 0.88)$
ma,s, $[2264, 3147)$gr	$0.62(0.29, 0.86)$
ma,me,s, $[580, 2264)$gr	$0.38(0.08, 0.70)$
ma,me,s, $[3147, 5260)$gr	$0.29(0.03, 0.66)$

Table 5: posterior predictive median survival probabilities at 24.5 years from birth and end-points of their 95% HPDs for the 7 profiles for which no observation is available. The highest predicted survival corresponds to future individuals having ambulatory impairment only and birth weight within $[3147, 5260)$ grams, whereas the worse predictions correspond to individuals within the same birth weight categoty and with manual, mental and sight impairments.

time and covariate dependent frailty parameters conferring flexibility to the survival processes.

The DBS framework represents an intermediate modelling framework between parametric and non-parametric survival models. While sharing with the former an interpretable parameter structure, the number of DBS random survival probabilities can be large as it is typically the case for semi-parametric and non-parametric models. Having a potentially large number of unknown parameters is not computationally cumbersome in our work because, by the conjugacy of the discrete beta-Stacy process, all survival probabilities are updated exactly in one step within the Gibbs sampler.

In Section 2 we noted that when the DBS precision coefficients and the parameters indexing the prior mean survival functions are unknown, the DBS models are not identifiable. An analogous point was noted in Dominici and Parmigiani [2001] and Carota and Parmigiani [2002] for model heirarchies including a Dirichlet process prior centered around a parametric backbone. Model identifiability is attained in these two papers via an informative prior for the Dirichlet process precision parameter shared by all samples. In this work we identify the DBS regressions by defining the covariate-dependent DBS precision parameters as a quadratic function of their linear predictors, implying that large regression parameters induce smooth survival processes and vice versa. This solution does not depend on specific assumptions on the prior for the precision parameters and it allows for the smoothness of each survival process to depend on its corresponding

CRiSM Paper No. 07-06v2, www.warwick.ac.uk/go/crism

linear predictor.

The example of Section 4.1 shows that the posterior estimates of the survival probabilities and those of the regession parameters are consistent with those of Dellaportas and Smith [1993] and with the non-parametric Kaplan-Meier estimates. The posterior estimates are also found to be not significantly affected by the resolution of τ as long as all the observed survival times are included. The examples of Sections 4.2 and 4.3 demonstrate the relevance of the DBS framework for clinical applications. The results presented in Section 4.2 support those of Andersen et al. [1993] and the recent directives included in the American Joint Committee on Cancer Staging Manual, empsasizing the key role of the tumor thickness, the skin resistance to tumor infiltration and the patients' age at surgery as independent prognostic factors of survival. The analysis of the cerebral palsy data of Section 4.3 confirms a non-linear effect of the birth weight on survival and it provides flexible predictions for the survival times associated to the seven covariate profiles not included in the dataset.

Throughout this paper we let the resolution of τ be fixed by design. When the position of some of the time points of τ cannot be fixed in advance, since the beta-Stacy is a Lévy process the algorithms of Walker and Damien [1998] and Wolpert and Ickstadt [1998] can be used to efficiently generate draws from their conditional posterior distributions. If also the number of jumps needs to be a priori unknown, a reversible jump step (Green [1995]) can be added to their samplers.

A second generalization of the DBS paradigm beyond the scope of this work allows for the covariate-dependent grouping structure of the different samples $\{g(i)\}_{i=1}^{N}$ to be a priori unknown. In the current model formulation, different individuals share a common survival process only if their covariate profiles are identical. However, when their covariate profiles are similar it might be possible to associate to all such individuals a common survival process. This extension of the DBS paradigm can thus lead to the construction of more parsimonious models when the covariate profiles of several groups of individuals are similar among themselves.

Appendix

The marginal likelihood of the survival data $\underline{t}$ given ther covariates X, the censoring indicators $\underline{\delta}$, the coefficients (θ, η) and the time grid τ can be obtained by integrating the likelihood function with respect to the joint discrete beta-Stacy prior of the survival probabilities $(S_{1,1}, ..., S_{N,k})$. Let

CRiSM Paper No. 07-06v2, www.warwick.ac.uk/go/crism

$\alpha = \{\{\alpha_{i,k}\}_{i=1}^N\}_{k=1}^K$, $\beta = \{\{\beta_{i,k}\}_{i=1}^N\}_{k=1}^K$ and $S_i = \{S_{i,k}\}_{k=1}^K$. By assuming that the survival times of the N samples are independent conditionally on their survival probabilities, this integral can be written as

$$p(\underline{t} \mid \underline{\delta}, \tau, \alpha, \beta) = \prod_{i=1}^N \int_{S_i} \prod_{k=1}^K \left((S_{i,k-1} - S_{i,k})^{\delta_i} S_{i,k}^{1-\delta_i} \right)^{1_{\{t_i=\tau_k\}}} \times$$
$$\times \frac{\Gamma(\alpha_{i,k}+\beta_{i,k})}{\Gamma(\alpha_{i,k})\Gamma(\beta_{i,k})} (S_{i,k-1} - S_{i,k})^{\alpha_{i,k}-1} \frac{S_{i,k}^{\beta_{i,k}-1}}{S_{i,k-1}^{\alpha_{i,k}+\beta_{i,k}-1}} dS_i.$$

Let now $Y_{i,k} = S_{i,k-1} - S_{i,k}$ and $Y_i = \{Y_{i,k}\}_{k=1}^K$. As shown in Section 2, if the joint prior for the survival probabilities S_i is a discrete beta-Stacy process, it follows that $Y_{i,k}$ is conditionally distributed as $BS(\alpha_{i,k}, \beta_{i,k}, 1 - \sum_{j<k} Y_{i,j})$ and vice versa. Therefore, the integral can be rewritten as a function of the random jumps of the survival function Y as

$$p(\underline{t} \mid \underline{\delta}, \tau, \alpha, \beta) = \prod_{i=1}^N \int_{Y_i} \prod_{k=1}^K \left(Y_{i,k}^{\delta_i} (1 - \textstyle\sum_{j \le k} Y_{i,j})^{1-\delta_i} \right)^{1_{\{t_i=\tau_k\}}} \times$$
$$\times \frac{\Gamma(\alpha_{i,k}+\beta_{i,k})}{\Gamma(\alpha_{i,k})\Gamma(\beta_{i,k})} Y_{i,k}^{\alpha_{i,k}-1} \frac{(1-\sum_{j \le k} Y_{i,j})^{\beta_{i,k}-1}}{(1-\sum_{j<k} Y_{i,j})^{\alpha_{i,k}+\beta_{i,k}-1}} dY_i. \tag{14}$$

The K-dimensional integral on the right-hand side of equation (14) can be solved with respect to each of its coordinates $Y_{i,k}$ in turn, starting from $Y_{i,K}$. The expression of the marginal likelihood (6) follows by observing that

$$E(Y_{i,k} \mid \alpha_{i,k}, \beta_{i,k}, \delta_i) = \frac{\alpha_{i,k}^{\delta_i} \beta_{i,k}^{1-\delta_i}}{\alpha_{i,k} + \beta_{i,k}} \prod_{j<k} \frac{\beta_{i,j}}{\alpha_{i,j} + \beta_{i,j}},$$

and by substituting the expressions for the beta-Stacy hyperparameters (2) and (3).

Acknowledgements

The first author acknowledges CRiSM and EURANDOM for their support during the development of this paper and Doctor Pierpaolo De Blasi for providing the dataset analysed in Section 4.2 in digital format. The implementation of the Weibull DBS models used throughout this paper and the three analysed datasets can be made available upon request to the first author in form of an independent Matlab toolbox.

CRiSM Paper No. 07-06v2, www.warwick.ac.uk/go/crism

References

P.K. Andersen, O. Borgan, R.D. Gill, and N. Keiding. *Statistical Models Based on Counting Processes*. Springer, 1993.

C. Antoniak. Mixtures of Dirichlet processes with applications to Bayesian Nonparametric problems. *The Annals of Statistics*, 2:1152–1174, 1974.

L.Y. Brunner and A.Y. Lo. Bayes methods for a symmetric unimodal density and its mode. *The Annals of Statistics*, 17:1550–1566, 1989.

C. Carota and G. Parmigiani. Semiparametric regression for count data. *Biometrika*, 89:265–281, 2002.

D.M. Cifarelli and E. Regazzini. Problemi statistici non parametrici in condizioni di scambiabilita'parziale. Impiego di medie associative. *Quaderni dell'Istituto di Matematica Finanziaria dell'Universita'di Torino*, III, 12:1–36, 1978.

D.M. Cifarelli, P. Muliere, and M. Scarsini. Il modello lineare nell'approccio Bayesiano non parametrico. *Technical Report, Istituto Matematico G. Castelnuovo, Universitá di Roma*, 1981.

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

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

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

P. Dellaportas and A.F.M. Smith. Bayesian inference for Generalised Linear and Proportional Hazards models via Gibbs sampling. *Applied Statistics*, 42:443–459, 1993.

F. Dominici and G. Parmigiani. Bayesian semiparametric analysis of developmental toxicology data. *Biometrics*, 57:150–157, 2001.

M.D. Escobar. Estimating Normal means with a Dirichelt process prior. *Journal of the American Statistical Association*, 89:268–277, 1994.

V.T. Farewell. The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics*, 38:1041–1046, 1982.

V.T. Farewell. Mixture models in survival analysis: are they worth the risk? *The Canadian Journal of Statistics*, 14:257–262, 1986.

CRiSM Paper No. 07-06v2, www.warwick.ac.uk/go/crism

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

T.S. Ferguson and E.G. Phadia. Bayesian non parametric estimation based on censored data. *Annals of Statistics*, 7:163–186, 1979.

A.E. Gelfand and A. F. M. Smith. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85:398–409, 1990.

P.J. Green. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82:711–732, 1995.

A.P. Grieve. Applications of Bayesian software: two examples. *The Statistician*, 36:283–288, 1987.

W.K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57-1:97–109, 1970.

N.L. Hjort. Non parametric Bayes estimators based on Beta processes in models for life history data. *Annals of Statistics*, 18:1259–1294, 1990.

J.L. Hutton and P.O.D. Pharoah. Effects of cognitive, motor and sensory disabilities on survival in cerabral palsy. *Archives of Disease in Childhood*, 86:84–89, 2002.

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

N.P. Jewell. Mixtures of exponential distributions. *The Annals of Statistics*, 10: 479–484, 1982.

J.D. Kalbfleisch. Non-parametric Bayesian analysis of survival time data. *Journal of the Royal Statistical Society Series B*, 40:214–221, 1978.

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

K.P. Kleinman and J.G. Ibrahim. A semiparametric Bayesian approach to the random effects model. *Biometrics*, 54:921–938, 1998.

A. Kottas and A.E. Gelfand. Bayesian semiparametric median regression modeling. *Journal of the American Statistical Association*, 96:1458–1468, 2001.

L. Kuo. Bayesian bioassay design. *The Annals of Statistics*, 11:886–895, 1983.

A. Mira and S. Petrone. Bayesian hierarchical nonparametric inference for change-point problems. *Bayesian Statistics 5, J.O. Berger et al. Editors*, pages 693–703, 1996.

CRiSM Paper No. 07-06v2, www.warwick.ac.uk/go/crism

P. Muliere and L. Tardella. Approximating distributions of random functionals of Ferguson-Dirichlet priors. *The Canadian Journal of Statistics*, 26:283–297, 1998.

J.C. Naylor and A.F.M. Smith. Applications of a method for the efficient computation of posterior distributions. *Applied Statistics*, 31:214–225, 1982.

R.L. Prentice and J.D. Kalbfleisch. Hazard rate models with covariates. *Biometrics*, 35:25–39, 1979.

D. Sinha and D.K. Dey. The use of mixture models for the analysis of survival data with long-term survivors. *Journal of the American Statistical Association*, 92: 1195–1212, 1996.

J.F. Thompson, R.A. Scolyer, and R.F. Kefford. Cutaneous melanoma. *The Lancet*, 365:687–701, 2005.

L. Tierney. A note for Metropolis-Hastings kernels for general state-spaces. *The Annals of Probability*, 8:1–9, 1998.

S. Walker and P. Damien. A full Bayesian non parametric analysis involving a neutral to the right process. *Scandinavian Journal of Statistics*, 25:669–680, 1998.

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

S. G. Walker and B.K. Mallick. A Bayesian semiparametric accelerated failure time model. *Biometrics*, 55:477–483, 1999.

R.L. Wolpert and K. Ickstadt. Poisson/Gamma random field models for spatial statistics. *Biometrika*, 85:251–267, 1998.

CRiSM Paper No. 07-06v2, www.warwick.ac.uk/go/crism