BIOMETRICS 55, 477–483
June 1999

A Bayesian Semiparametric Accelerated Failure Time Model

Stephen Walker

Department of Mathematics, Imperial College,
180 Queen's Gate, London SW7 2BZ, U.K.

and

Bani K. Mallick

Department of Statistics, Texas A & M University,
College Station, Texas 77843-3143, U.S.A
email: bmallick@stat.tamu.edu

SUMMARY. A Bayesian semiparametric approach is described for an accelerated failure time model. The error distribution is assigned a Pólya tree prior and the regression parameters a noninformative hierarchical prior. Two cases are considered: the first assumes error terms are exchangeable; the second assumes that error terms are partially exchangeable. A Markov chain Monte Carlo algorithm is described to obtain a predictive distribution for a future observation given both uncensored and censored data.

KEY WORDS: Markov chain Monte Carlo; Partial exchangeability; Pólya trees.

1. Introduction

Considerable attention has focused on the role of covariates in survival studies. This paper is directed exclusively at a Bayesian semiparametric approach for an accelerated failure time model. The accelerated failure time model assumes that failure times, $T_1, T_2, \ldots, T_n$, arise according to the model

$$T_i = \exp(-\mathbf{X}_i\boldsymbol{\beta})V_i, \qquad i = 1, \ldots, n, \tag{1}$$

which in log scale becomes the linear model

$$\log T_i = -\mathbf{X}_i\boldsymbol{\beta} + \Theta_i, \qquad i = 1, \ldots, n, \tag{2}$$

where $\mathbf{X}_i = (x_{i1}, x_{i2}, \ldots, x_{ip})$ is a vector of known explanatory variables for the ith individual, $\boldsymbol{\beta}$ is a vector of p unknown regression coefficients, and $\Theta_i = \log V_i$ is the error term. Usually the distribution of the error term is assumed to be a member of some parametric family. This parametric assumption may be too restrictive and so we propose a Bayesian nonparametric approach to model the distribution of the error terms. We assume initially that the error terms are independently and identically distributed (i.i.d.) from some unknown distribution F. Later we relax this condition by introducing a partially exchangeable model.

A classical analysis of an accelerated failure time model is described by Kalbfleisch and Prentice (1980). A semi-Bayesian analysis of this model is given by Christensen and Johnson (1988), where they model V_i as being i.i.d. from F, with F chosen from a Dirichlet process (Ferguson, 1973). The Dirichlet process is popular due to the simple interpretation of its parameters, a base measure with associated precision. However, troubles arise due to the discrete nature of the Dirichlet process, which is discussed in Johnson and Christensen (1989). Even in the uncensored case, a full Bayesian analysis is very difficult, let alone in the presence of censored observations. Christensen and Johnson present a semi-Bayesian approach in the sense that they first obtain a marginal estimate for $\boldsymbol{\beta}$, after which the analysis is straightforward.

Instead of the Dirichlet process, we propose a Pólya tree distribution (Lavine, 1992; Mauldin, Sudderth, and Williams, 1992) as the prior for the unknown distribution of Θ. The advantages over the Dirichlet process are (i) the conjugate nature of Pólya trees makes the analysis uncomplicated; (ii) under some sufficient conditions, Pólya tree priors assign probability one to the set of continuous distributions; (iii) it is easy to constrain a random Pólya tree distribution to have median zero and hence to consider a median regression model for (2) (see Ying, Jung, and Weis [1995] for a frequentist version); (iv) it is easy to sample a random Pólya tree distribution, so samples from posterior functionals of the survival curve will be available, permitting a full Bayesian analysis; and (v) the Dirichlet process and the beta process (Hjort, 1990) are a special case of Pólya trees, so we are generalizing previous work.

There is a large amount of literature about the use of Markov chain Monte Carlo (MCMC) to model with a Dirichlet process (Escobar, 1994) and a mixture of Dirichlet processes but, as yet, there is very little literature for Pólya trees other than Lavine's (1994) work. In this paper. we show how to implement MCMC for model (2) with application for survival analysis.

15410420, 1999, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.0006-341X.1999.00477.x by Indiana University Libraries Technical Services/Acquisitions, Wiley Online Library on [13/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

In Section 2, we describe Pólya tree prior distributions; in Section 3, we describe the Markov chain Monte Carlo algorithm for analysis of model (2) for both censored and uncensored observations and also present the predictive distribution for a new observation; Section 4 contains some numerical examples; and Section 5 considers a more flexible form of the model in which the error terms are assumed to be partially exchangeable instead of the more usual assumption of exchangeability.

2. Pólya Tree Distributions

Pólya tree distributions for random probability measures were recently studied by Lavine (1992, 1994) and by Mauldin et al. (1992), although an original description was given by Ferguson (1974). Let Ω be a separable, measurable space (though we will only be considering $\Omega = (-\infty, \infty)$) and let (B_0, B_1) be obtainted splitting Ω into two pieces. Similarly, B_0 splits into (B_{00}, B_{01}) and B_1 splits into (B_{10}, B_{11}). Continue in this fashion *ad infinitum*. Let, for some m, $\epsilon = \epsilon_1 \cdots \epsilon_m$, with the $\epsilon_k \in \{0, 1\}$, $k = 1, \ldots, m$, so that each ϵ defines a unique set B_ϵ. The number of sets at the mth level is 2^m. Thus, in general, B_ϵ splits into $B_{\epsilon 0}$ and $B_{\epsilon 1}$. Degenerate splits are allowed so that we could have $B_{\epsilon 0} = B_\epsilon$ and $B_{\epsilon 1} = \emptyset$.

DEFINITION 1 (Lavine, 1992): A random probability measure F on Ω is said to have Pólya tree distribution, or a Pólya tree prior, with parameter $(\Pi, \mathcal{A})$, written $F \sim PT(\Pi, \mathcal{A})$, if there exist nonnegative numbers $\mathcal{A} = (\alpha_0, \alpha_1, \alpha_{00}, \ldots)$ and random variables $\mathcal{C} = (C_0, C_{00}, C_{10}, \ldots)$ such that the following hold:

(i) all the random variables in $\mathcal{C}$ are independent;
(ii) for every ϵ, $C_{\epsilon 0} \sim \text{beta}(\alpha_{\epsilon 0}, \alpha_{\epsilon 1})$;
(iii) for every $m = 1, 2, \ldots$ and every $\epsilon = \epsilon_1 \cdots \epsilon_m$,

$$F(B_{\epsilon_1 \cdots \epsilon_m}) = \left(\prod_{j=1; \epsilon_j = 0}^{m} C_{\epsilon_1 \cdots \epsilon_{j-1} 0} \right) \times \left(\prod_{j=1; \epsilon_j = 1}^{m} (1 - C_{\epsilon_1 \cdots \epsilon_{j-1} 0}) \right), \quad (3)$$

where the first terms, i.e., for $j = 1$, are interpreted as C_0 and $1 - C_0$.

A random probability measure $F \sim PT(\Pi, \mathcal{A})$ is sampled by sampling $\mathcal{C}$ as indicated in Definition 1. Since $\mathcal{C}$ is an infinite set, an approximate probability measure from $PT(\Pi, \mathcal{A})$ is sampled by terminating the process at a finite level M. Lavine (1992) refers to this as a partially specified Pólya tree. Let this finite set be denoted by $\mathcal{C}_M$. From the sampled variates of $\mathcal{C}_M$, we define $F(B_{\epsilon_1 \cdots \epsilon_M})$ for each $\epsilon = \epsilon_1 \cdots \epsilon_M$ according to (3). For analysis, our attention is restricted to the $r = 2^M$ partitions given by $\pi_M = (B_1, \ldots, B_r)$, where π_M is the collection of partitions at the level M. If Θ is an observation from F, then we are only interested in which $B \in \pi_M$ that Θ is observed. For predictive inference, we will also only be interested in the r probabilities $p(\Theta \in B)$ for each $B \in \pi_M$. Assesssing the closeness of a random Pólya tree distribution from a partially specified Pólya tree to an exact Pólya tree is given in Lavine (1992), where a criterion for the selection of M is given.

We center the Pólya tree prior on a particular probability distribution G by taking the partitions to coincide with percentiles of G and then taking $\alpha_{\epsilon 0} = \alpha_{\epsilon 1}$ for each ϵ. This involves setting $B_0 = (-\infty, G^{-1}(1/2))$, $B_1 = [G^{-1}(1/2), \infty)$, and, at level m, setting, for $j = 1, \ldots, 2^m$, $B_j = [G^{-1}((j-1)/2^m), G^{-1}(j/2^m))$, with $G^{-1}(0) = -\infty$ and $G^{-1}(1) = +\infty$, where $(B_j; j = 1, \ldots, 2^m)$ correspond in order to the 2^m partitions of level m. It is then straightforward to show that $\mathrm{E}[F(B_\epsilon)] = G(B_\epsilon)$ for all ϵ. Note that here G defines Π. Therefore, we need to specify G, $\mathcal{A}$, and M, and we discuss the choice of these parameters in Section 2.1.

Given an observation Θ_1 from F, the posterior Pólya tree distribution is easily obtained (Lavine, 1992). We write $F \mid \Theta_1 \sim PT(\Pi, \mathcal{A} \mid \Theta_1)$ and with $\mathcal{A} \mid \Theta_1$ given by

$$\alpha_\epsilon \mid \Theta_1 = \begin{cases} \alpha_\epsilon + 1 & \text{if } \Theta_1 \in B_\epsilon \\ \alpha_\epsilon & \text{otherwise.} \end{cases}$$

For n independent observations given by data equal $(\Theta_1, \ldots, \Theta_n)$, $\mathcal{A} \mid \text{data}$ is given by $\alpha_\epsilon \mid \text{data} = \alpha_\epsilon + n_\epsilon$, where n_ϵ is the number of observations from $(\Theta_1, \ldots, \Theta_n)$ in B_ϵ. It is easy to obtain samples from $p(F(B_\epsilon) \mid \text{data})$ by sampling $\mathcal{C}_M$ and using (3). This permits a full Bayesian analysis.

In many applications, it is the posterior predictive distribution for the next observation that is of interest, i.e., $p(\Theta_{n+1} \in B_\epsilon \mid \text{data})$ for some ϵ. Let $\epsilon = \epsilon_1 \cdots \epsilon_M$, then

$$\begin{aligned} p\bigl(\Theta_{n+1} \in B_\epsilon \mid \text{data}\bigr) &= \mathrm{E}\bigl(F(B_\epsilon) \mid \text{data}\bigr) \\ &= \frac{\alpha_{\epsilon_1} + n_{\epsilon_1}}{\alpha_0 + \alpha_1 + n} \frac{\alpha_{\epsilon_1 \epsilon_2} + n_{\epsilon_1 \epsilon_2}}{\alpha_{\epsilon_1 0} + \alpha_{\epsilon_1 1} + n_{\epsilon_1}} \cdots \\ &\quad \times \frac{\alpha_{\epsilon_1 \cdots \epsilon_m} + n_{\epsilon_1 \cdots \epsilon_M}}{\alpha_{\epsilon_1 \cdots \epsilon_{M-1} 0} + \alpha_{\epsilon_1 \cdots \epsilon_{M-1} 1} + n_{\epsilon_1 \cdots \epsilon_{M-1}}}. \end{aligned} \quad (4)$$

This result is straightforward to derive from Definition 1.

2.1 *Assigning G, $\mathcal{A}$, and M*

Assigning G. It seems reasonable to take G as a normal distribution with mean (median) zero. This is usually the choice for an error distribution in a parametric framework. We found that fixing a large variance proved satisfactory in practice.

Assigning $\mathcal{A}$. In practice, we do not wish to assign a separate α_ϵ for each ϵ. It is convenient therefore to take $\alpha_\epsilon = c_m$ whenever ϵ defines a set at level m. For the higher levels (m small), it is not necessary for $F(B_{\epsilon 0})$ and $F(B_{\epsilon 1})$ to be close; on the contrary, it is desirable for a large amount of variability. However, as we move down the levels (m large), we wish $F(B_{\epsilon 0})$ and $F(B_{\epsilon 1})$ to be close to reflect beliefs in the underlying continuity of F. This can be achieved by allowing c_m to be small for small m and allowing c_m to increase as m increases, e.g., $c_m = cm^2$ for some $c > 0$. According to Ferguson (1974), $c_m = m^2$ implies F is absolutely continuous with probability one and therefore, according to Lavine (1992), this would often be a sensible canonical choice. Therefore, we choose cm^2 for the α's and we can consider a prior for c. Note the Dirichlet process arises when $c_m = c/2^m$, which means that $c_m \to 0$ as $m \to \infty$ and F is discrete with probability one (Blackwell, 1973).

Assigning M. We can either fix M or assign it a prior distribution, e.g., a Poisson distribution. This is not difficult to do. However, we found that fixing M such that the parti-

15410420, 1999, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.0006-341X.1999.00477.x by Indiana University Libraries Technical Services/Acquisitions, Wiley Online Library on [13/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

tions around zero were adequately small (in the sense that, given an observation in one of these partitions, we would not want to locate it any more precisely) was satisfactory. If observations do occur in larger partitions, then M can be increased.

3. Prediction via MCMC

Here we describe MCMC algorithms for obtaining samples from posterior distributions of $\boldsymbol{\beta}$ and F for both censored and uncensored observations. We then derive the predictive survival curve for a future patient with a particular set of covariate values.

Consider the linear model (2). For our semiparametric analysis of the model, we will assign F with a Pólya tree prior and for the parameter $\boldsymbol{\beta}$ a p-dimensional multivariate normal distribution with mean μ and covariance matrix Σ. *A priori*, we take F and $\boldsymbol{\beta}$ to be independent.

If F is allowed to be completely arbitrary, then a source of confounding is the intercept term of $\boldsymbol{\beta}$ with the location of F. To solve this problem, we fix the median of F at $G^{-1}(1/2)$ by defining $F(B_0) = F(B_1) = 1/2$. This does not affect the centering of F. If G is normal with median at zero, then F has its median at zero almost surely. Here (2) becomes a median regression model with $\text{med}(Y_i) = -\mathbf{X}_i\boldsymbol{\beta}$ instead of the more usual mean regression model. This could also be easily extended to a quantile regression model if required. Newton, Czado, and Chappell (1996) describe a similar procedure for the Dirichlet process.

A frequentist semiparametric inference for median regression model is described by Ying et al. (1995). They do not assume that the error terms are i.i.d. As a consequence, their approach only obtains an estimate for $\boldsymbol{\beta}$ and hence only a predictive median survival time.

Interest for us is in obtaining posterior distributions for $\boldsymbol{\beta}$ and F that will lead to a predictive distribution for the survival curve. This is analytically intractable and hence a Markov chain Monte Carlo (MCMC) method (Tierney, 1994) is utilized in order to obtain samples from the relevant posterior distributions.

We will consider two situations. The first is when all the Y_i's are observed exactly and where y_i represents the observed value of Y_i. The second situation is when some of the Y_i's are subjected to random right censoring, i.e., it is only known that $Y_i \geq y_i$ for some observed value y_i. In this case, the data are represented as

$$(y_1, \delta_1), \ldots, (y_n, \delta_n),$$

with the implication that $Y_i = y_i$ if $\delta_i = 1$ and $Y_i \geq y_i$ if $\delta_i = 0$.

3.1 *Uncensored Observations*

Here we assume that the survival times for the n patients are known exactly so that the data is given by $(y_1, \ldots, y_n)$. We implement MCMC to sample $\boldsymbol{\beta}$ and F from the posterior.

As mentioned previously, we will only be considering the distributions for F up to a finite level M for which there will be the partitions $\pi_M = \{B_j : j = 1, \ldots, r\}$, where $r = 2^M$. Θ is drawn from F, so the likelihood of Θ is given by $F(B_\Theta)$, for some $B_\Theta \in \pi_M$, where $\Theta \in B_\Theta$. Likewise, as Y is an observation, given $\boldsymbol{\beta}$, from an individual with covariate vector $\mathbf{X}$, the likelihood of $\boldsymbol{\beta}$ is again given by $F(B_\Theta)$, where $\Theta = Y + \mathbf{X}\boldsymbol{\beta}$.

Obtaining the posterior distribution of $p(F \mid \boldsymbol{\beta}, \text{data})$ for the Pólya tree is straightforward since $\Theta_i = Y_i + \mathbf{X}_i\boldsymbol{\beta}$ are i.i.d. observations from F. This involves the sampling of a random probability measure F, given the data and the current value of $\boldsymbol{\beta}$, by sampling $\mathcal{C}_M$ (see Definition 1) from which we will obtain the random probabilities $\{\mu_j = F(B_j) : j = 1, \ldots, r\}$.

To sample $\boldsymbol{\beta}$, we first observe that

$$p(\boldsymbol{\beta} \mid F, \text{data}) \propto p(\text{data} \mid F, \boldsymbol{\beta})p_0(\boldsymbol{\beta}),$$

where $p_0(\boldsymbol{\beta})$ is the prior for $\boldsymbol{\beta}$ and $p(\text{data} \mid F, \boldsymbol{\beta}) \propto \Pi_{i=1}^n F(B_{\Theta_i})$. We use a Metropolis–Hastings (Hastings, 1970) algorithm to obtain the required sample.

At the kth iteration of the Markov chain, we obtain the samples $\{\mu_{jk} : j = 1, \ldots, r\}$ and $\boldsymbol{\beta}_k$. After a suitable burn-in period, the samples are kept, say for $l = 1, \ldots, L$, from which posterior summaries can be obtained.

3.2 *Censored Observation*

For censored observations, the extra step required in the MCMC algorithm, as descibed in Section 3.1, is the sampling of the missing data corresponding to those observations for which $\delta_i = 0$. Assume without loss of generality that $\delta_i = 0$ for $i = 1, \ldots, i'$ and that $\delta_i = 1$ for $i = i'+1, \ldots, n$ for some i'. The required extra steps are the sampling of random variates from

$$p(Y_i \mid Y_{(i)}, \boldsymbol{\beta}, F, Y_i \geq y_i) = p(Y_i \mid F, \boldsymbol{\beta}, Y_i \geq y_i),$$
$$i = 1, \ldots, i', \tag{5}$$

where $Y_{(i)} = (Y_1, \ldots, Y_{i-1}, Y_{i+1}, \ldots, Y_n)$.

Let the partition points of the Pólya tree structure Π at level M be $\{a_j : j = 1, \ldots, r-1\}$. For each $i = 1, \ldots, i'$, select the $a_{j(i)}$ closest to and below $y_i + \mathbf{X}_i\boldsymbol{\beta}$ and let F_i^* represent the probability measure F restricted to $[a_{j(i)}, \infty)$. If F_i^* is defined by the probabilities μ_{ij}^* for $j = j(i), \ldots, r$, then $\mu_{ij}^* = \mu_j/\mu_{(i)}$, where $\mu_{(i)} = \Sigma_{j=j(i)}^r \mu_j$. We then take Θ_i from F_i^* by sampling a set B from $\{B_j : j = j(i), \ldots, r\}$ with corresponding probabilities $\{\mu_{ij}^* : j = j(i), \ldots, r\}$ and then take Θ_i uniformly from B. If B is one of the outer partitions, it is necessary for M to be increased until all of these B's are not outer partitions. Finally, we take $Y_i = \Theta_i - \mathbf{X}_i\boldsymbol{\beta}$ to be the random variate from (5). Interval-censored observations can be dealt with in essentially the same way. Here we would need to consider F restricted to a set $[a_{j(.)}, b_{j(.)})$.

3.3 *Prediction of Survival Curve*

Here we consider the predictive survival curve for a new patient with covariate matrix $\mathbf{X}$. First, consider the uncensored data. We are interested in

$$p(T > t \mid \text{data}) = 1 - p(T \leq t \mid \text{data})$$

for any $t > 0$. Now

$$p(T \leq t \mid \text{data}) = \int p(T \leq t \mid \text{data}, \boldsymbol{\beta})p(\boldsymbol{\beta} \mid \text{data})d\boldsymbol{\beta},$$

which we approximate using the samples $\{\boldsymbol{\beta}_l\}$ obtained from the simulated Markov chain, by

$$L^{-1}\sum_{l=1}^{L} p(T \leq t \mid \text{data}, \boldsymbol{\beta}_l).$$

15410420, 1999, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.0006-341X.1999.00477.x by Indiana University Libraries Technical Services/Acquisitions, Wiley Online Library on [13/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Now $p(T \leq t \mid \text{data}, \beta_l) = p(\Theta \leq \log t + \mathbf{X}\beta_l \mid \text{data}, \beta_l)$, which is given by

$$\sum_{j=1}^{\eta} \lambda_j(\beta_l),$$

where $\eta = j(\log t + \mathbf{X}\beta_l)$, $\lambda_j(\beta_l) = \mathrm{E}(F(B_j) \mid \Theta_{1l}, \ldots, \Theta_{nl})$, $\Theta_{il} = Y_i + \mathbf{X}_i\beta_l$, and $j(s) \in (1, \ldots, r)$ such that $s \in B_{j(s)}$. Each λ is evaluated from (4). The predictive survival curve is then given by

$$\hat{S}(t) = 1 - \frac{1}{L}\sum_{l=1}^{L}\sum_{j=1}^{\eta} \lambda_j(\beta_l). \tag{6}$$

For censored data, it is straightforward to show that the predictive survival curve is also given by (6) except for $\lambda_j(\beta_l) = \mathrm{E}(F(B_j) \mid \Theta_{1l}, \ldots, \Theta_{nl})$, where now $\Theta_{il} = Y_{il} + \mathbf{X}_i\beta_l$ for $i = 1, \ldots, i'$ and $\Theta_{il} = Y_i + \mathbf{X}_i\beta_l$ for $i = i' + 1, \ldots, n$. It is straightforward to evaluate (6) with no extra computation since the $\lambda_j(\beta_l)$ can be obtained at each iteration of the chain.

3.4 *Posterior Distributions*

It is also possible to obtain samples from the posterior distribution

$$p(S(t) \mid \text{data}, \mathbf{X})$$

for any time point t and explanatory variable $\mathbf{X}$. Using the samples $\{\mu_{jl}\}$ from the Markov chain output,

$$S_l = 1 - \sum_{j=1}^{\eta} \mu_{jl}$$

is a sample from the required posterior distribution.

4. Examples

Here two numerical examples are given using real data sets. For all the examples, we took a noninformative hierarchical normal prior for β (see Wakefield et al., 1994, for example). For the Pólya tree prior, we took the prior expected probability measure G (see Section 2) to be normal with mean zero and standard deviation of 10 and took $\mathcal{A}$ such that, at level m, all the α's are equal to $0.1 \times m^2$. We took M to be eight, so the number of partitions considered is 2^8. The MCMC algorithm was first run in order to ascertain at which point suitable convergence was obtained, and then a further chain was run in order to obtain 3000 samples for inference.

1. *Uncensored data.* For this example, we consider the data set presented by Feigl and Zelen (1965), which involves 33 patients suffering from leukaemia. Each patient's design matrix $\mathbf{X}_i$ is given by (x_{i1}, x_{i2}, x_{i3}), where $x_{i1} = 1$, x_{i2} is an indicator function that is zero if the patient's Auer rods and/or granulature (AG) factor is positive and one if the patient's AG factor is negative, and x_{i3} is the natural logarithm of the patient's white blood cell count. The uncensored observations, $\{Y_i\}$, are taken to be the natural logarithm of the patients' survival times in weeks.

The Bayes estimates of β and the 95% credible intervals (the Bayesian analog of classical confidence intervals) are given by $\hat{\beta}_1 = -5.26\ (-6.7, -3.69)$, $\hat{\beta}_2 = 1.46\ (-0.61, 3.07)$, and $\hat{\beta}_3 = 0.64\ (0.31, 0.98)$.

Two sets of predictive survival curves are presented—the first for new patients with positive AG factor (Figure 1) and the second for new patients with negative AG factor (Figure 2). The three curves are for patients whose covariate values coincide with the quartiles of the observed covariate values.

Also with this data set, we took five of the patients (with positive AG factor) and simulated censoring times. The new Bayes estimates of β and the standard errors are given by $\hat{\beta}_1 = -5.33\ (-6.14, -3.92)$, $\hat{\beta}_2 = 1.22\ (-0.49, 3.12)$, and $\hat{\beta}_3 = 0.56\ (0.19, 1.01)$.

The estimated failure times for the five patients alongside the true values are given in Table 1.

2. *Censored data.* For the censored example, we use the data set presented by Ying et al. (1995). This involves 121 patients suffering small-cell lung cancer and each undertaking one of two treatments, arm A (62 patients) or arm B (59) patients. The survival times are given in days, with 98 patients providing exact survival times and the remainder providing right-censored survival times. The covariates are the treatment type, zero or one, and the natural logarithm of the entry age of the patient. The Bayes estimates of β and 95% credible intervals are given by $\hat{\beta}_1 = -7.57\ (-8.92, -6.87)$, $\hat{\beta}_2 = 0.30\ (0.08, 0.91)$, and $\hat{\beta}_3 = 0.31\ (0.07, 0.78)$. The results agree with those of Ying et al. (1995).

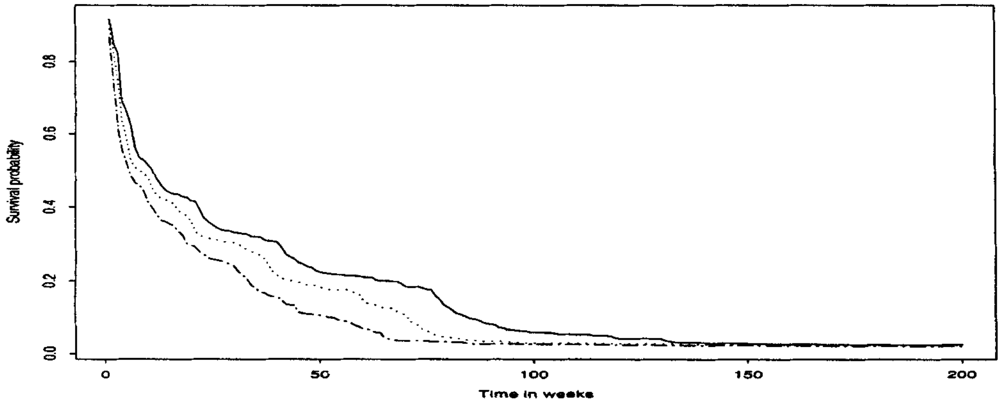

Figure 1. Predictive survival curves for three new patients with positive AG factor.

15410420, 1999, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.0006-341X.1999.00477.x by Indiana University Libraries Technical Services/Acquisitions, Wiley Online Library on [13/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

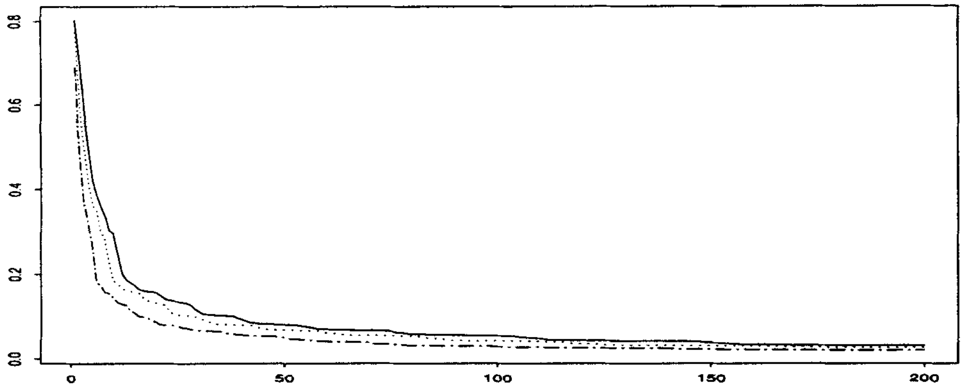

Figure 2. Predictive survival curves for three new patients with negative AG factor.

5. Partially Exchangeable Model

Here we propose an alternative and more flexible formulation for model (2). So far, it has been assumed that, given F, all the error terms are i.i.d. F. In some instances, this may be too strong an assumption. In particular, we consider experiments in which individuals can be put into like groups, e.g., according to a categorical covariate associated with each individual.

The approach considered up to this point assumes the error terms to be exchangeable. The assumption of exchangeability may in some situations be too strong, e.g., when one suspects that the individuals may have varying degrees of similarity or likeness. A more flexible approach, which allows individuals to be thought of as being more alike and unlike, consists of entertaining several partial exchangeability structures for the error terms $\{\Theta_1, \ldots, \Theta_n\}$ and then combining the corresponding inferences by summarising over the posterior beliefs in these structures.

Briefly, the method can be thought of as partitioning the experimental set, $\{i = 1, \ldots, n\}$, to give d subsets denoted by $S_1, \ldots, S_d$. The basic assumption is to regard as exchangeable only the Θ_i's associated with individuals belonging to the same partition subset S_k while the Θ_i's associated with individuals in different subsets are taken to be independent. The advantage of the analysis lies essentially in its ability to borrow strength from related individuals without imposing a prespecified dependence structure on the error terms. (See Mallick and Walker (1997) for more details.)

Survival data typically arises as a result of a study comparing two treatment types that would have the effect of splitting the patient population undergoing therapy into two groups (cf., the data set of Ying et al., 1995). Therefore, an alternative model to (2) is the partially exchangeable model, in which it is assumed that the individuals on one of the treatments have error terms that are i.i.d. F_1 and the rest have error terms that are i.i.d. F_2. Therefore, instead of the treatment types causing a median shift in the distribution of the error terms, a possible distribution shift is now being allowed (which is shifting the medians, too). The assumption on F_1 and F_2 is they are i.i.d. from a Pólya tree prior distribution. Represent the model given by (2) as M_0 and the new model by M_1.

These two models can be thought of as competing models. However, instead of selecting one of them as being optimal in some sense, we follow Draper (1994) and assign prior weights to the two models to reflect prior opinion concerning their relative plausibility. Inference will then be made by averaging over the posterior weights.

Predictive probabilities are taken as a mixture of the predictive probabilities obtained from each model, i.e., as

$$p(T > t \mid \text{data}) = p(T > t \mid \text{data}, M_0)p(M_0 \mid \text{data}) + p(T > t \mid \text{data}, M_1)p(M_1 \mid \text{data}).$$

Therefore, two analyses are performed under M_0 and M_1. For the final inference, the posterior weights $p(M_0 \mid \text{data})$ and $p(M_1 \mid \text{data})$ need to be evaluated.

These posterior weights can be obtained in several ways, but we have used the simplest method. From Bayes theorem, we obtain

$$p(M_j \mid \text{data}) = \frac{p(\text{data} \mid M_j)}{\sum_{i=0}^{1} p(\text{data} \mid M_i)p(M_i)}$$

for $j = 0, 1$. $p(\text{data} \mid M_j) = \int p(\text{data} \mid M_j, \theta_j)p(\theta_j \mid M_j)d\theta_j$, where θ_j is the vector of parameters under M_j for $j = 0, 1$. We ran both of the models seperately and have m samples $\theta_j^{(i)}$, $i = 1, \ldots, m$, $j = 0, 1$, drawn from the posterior distributions of θ_j. Now $p(\text{data} \mid M_j)$ can be approximated by $\{(1/m)\sum_{i=1}^{m} p(\text{data} \mid M_j, \theta_j^{(i)})\}^{-1}$, the harmonic mean of the likelihood values evaluted at posterior samples (Newton and Raftery, 1994).

The censored data example was rerun with the new mixture model with M_1 being the partially exchangeable model partitioned into the two treatment groups. The prior weights for each of the models was 1/2, and the posterior weights obtained were 0.19 for M_0 and 0.81 for M_1. This suggests that the data "prefers" the distribution shift as opposed to the median shift for modelling the effect of the treatment types.

Table 1
Estimated and true failure times

Patient	Estimated	True
1	60.3	65.0
2	164.0	156.0
3	99.5	100.0
4	134.3	134.0
5	38.9	16.0

15410420, 1999, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.0006-341X.1999.00477.x by Indiana University Libraries Technical Services/Acquisitions, Wiley Online Library on [13/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

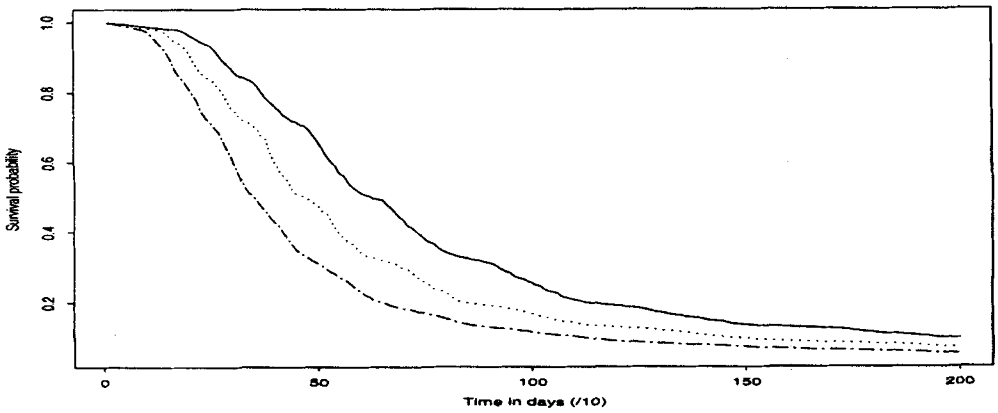

Figure 3. Predictive survival curves for three new patients with treatment arm A (exchangeable model).

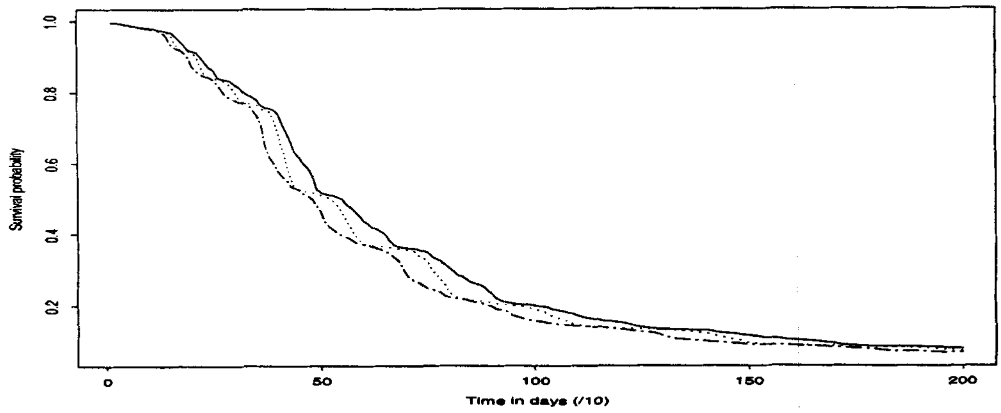

Figure 4. Predictive survival curves for three new patients with treatment arm A (partially exchangeable model).

Two predictive survival curves are presented, the first for a new patient with arm A treatment for exhangeable model (Figure 3) and the second for a new patient with arm A treatment for partially exchangeable model (Figure 4).

6. Conclusions

We have proposed a semiparametric Bayesian accelerated failure time model. An adavantage of this method is that credible intervals for the parameters as well as the predictive survival curves can be obtained. The Bayesian inference framework with a sampling-based implementation offers a relatively straightforward fitting technique.

Acknowledgements

The authors thank an associate editor and a referee for very helpful remarks. The second author's research was partially supported by grant CA 57030 from the National Cancer Institute.

Résumé

On décrit ici une approche semi-paramétrique bayésienne pour un modèle de défaillance accélérée. La distribution a priori des erreurs est supposée être un arbre de Pólya, et on suppose un a priori non informatif hiérarchique sur les paramètres de régression. Deux cas sont envisagés: dans le premier on suppose que les termes d'erreur sont échangeables; dans le second on suppose que les termes d'erreur sont partiellement échangeables. Un algorithme MCMC est décrit pour obtenir la distribution prédictive d'une observation future, étant données des observations censurées et non censurées.

References

Blackwell, D. (1973). The discreteness of Ferguson selections. *The Annals of Statistics* **1,** 356–358.

Christensen, R. and Johnson, W. (1988). Modelling accelerated failure time with a Dirichlet process. *Biometrika* **75,** 693–704.

Draper, D. (1994). Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society, Series B* **57,** 45–98.

Escobar, M. D. (1994). Estimating normal means with a Dirichlet process prior. *Journal of the American Statistical Association* **89,** 268–277.

Feigl, P. and Zelen, M. (1965). Estimation of exponential probabilities with concomitant information. *Biometrics* **21,** 826–838.

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

15410420, 1999, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.0006-341X.1999.00477.x by Indiana University Libraries Technical Services/Acquisitions, Wiley Online Library on [13/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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

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

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

Johnson, W. and Christensen, R. (1989). Nonparametric Bayesian analysis of the accelerated failure time model. *Statistics and Probability Letters* **7,** 179–184.

Kalbfleisch, J. D. and Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data.* New York: Wiley.

Lavine, M. (1992). Some aspects of Pólya tree distributions for statistical modelling. *The Annals of Statistics* **20,** 1222–1235.

Lavine, M. (1994). More aspects of Pólya tree distributions for statistical modelling. *The Annals of Statistics* **22,** 1161–1176.

Mallick, B. K. and Walker, S. G. (1997). Combining information from several experiments with nonparametric priors. *Biometrika* **84,** 697–706.

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

Newton, M. A. and Raftery, A. E. (1994). Approximate Bayesian inference by the weighted likelihood bootstrap (with discussion). *Journal of the Royal Statistical Society, Series B* **56,** 1–48.

Newton, M. A., Czado, C., and Chappell, R. (1996). Bayesian inference for semiparametric binary regression. *Journal of the American Statistical Association* **91,** 142–153.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *The Annals of Statistics* **22,** 1701–1762.

Wakefield, J. C., Smith, A. F. M., Racine-Poon, A., and Gelfand, A. E. (1994). Bayesian analysis of linear and nonlinear population models using the Gibbs sampler. *Applied Statistics* **43,** 201–221.

Ying, Z., Jung, S. H., and Wei, L. J. (1995). Survival analysis with median regression models. *Journal of the American Statistical Association* **90,** 178–184.

Received April 1997. *Revised August* 1998.
Accepted August 1998.

15410420, 1999, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.0006-341X.1999.00477.x by Indiana University Libraries Technical Services/Acquisitions, Wiley Online Library on [13/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License