

BAYESIAN FRAILTY MODELS BASED ON BOX-COX TRANSFORMED HAZARDS

Guosheng Yin and Joseph G. Ibrahim

The University of Texas and The University of North Carolina at Chapel Hill

Abstract: Due to natural or artificial clustering, multivariate failure time data often arise in biomedical research. To account for the intracluster correlation, we propose a novel class of frailty models by imposing the Box-Cox transformation on the hazard functions. This class of models generalizes the relationships between the baseline hazard and the hazard functions, which includes the proportional and the additive hazards frailty models as two special cases. Since hazards cannot be negative, complex multidimensional nonlinear parameter constraints must be imposed in the model formulation. To facilitate a tractable computational algorithm, the joint priors are constructed through a conditional-marginal specification. The conditional distribution of the prior specification is univariate and absorbs the parameter constraints, while the marginal part is free of constraints. We propose a Markov chain Monte Carlo (MCMC) computational scheme for sampling from the posterior distribution of the parameters. We derive an MCMC approximation for the conditional predictive ordinate to assess model adequacy, and illustrate the proposed method with a dataset.

Key words and phrases: Additive hazards, Bayesian inference, Box-Cox transformation, constrained parameter, frailty model, Gibbs sampling, proportional hazards.

1. Introduction

In biomedical research, multivariate failure time data occur in the presence of natural or artificial clustering effects. For example, in family studies of genetic diseases or litter-matched mice studies, failure times for subjects in the same cluster may be dependent. An example is the Diabetic Retinopathy Study (DRS Research Group (1985)), which involved paired failure time data on patients' eyes. The study focused on the investigation of abnormalities in the microvasculature within the retina of the eye associated with diabetes mellitus. In the United States, diabetic retinopathy is the leading cause of legal blindness in people under the age of 65, and it remains a significant health problem worldwide. There were 197 patients in the dataset, which was a 50% random sample of the patients with "high-risk" diabetic retinopathy, as defined by the DRS criteria. One eye of each subject was randomly selected for laser photocoagulation treatment and the other eye was untreated, for ethical reasons. The event of interest was time from

the initiation of treatment to the occurrence of severe visual loss (blindness). Censoring was on the basis of death, dropout, or end of the study. The primary objective was to study the effectiveness of laser photocoagulation in delaying the onset of blindness. The failure times of the two eyes on the same subject might be correlated, which needs to be taken into account in statistical modeling and inference.

A common approach to accommodating the intraclass correlation is to incorporate an unobserved random effect, or frailty, into the Cox (1972) proportional hazards model. For the k th subject in the i th cluster, with a possibly external time-dependent covariate vector $\mathbf{Z}_{ik}(t)$ ($i = 1, \dots, n; k = 1, \dots, K_i$) as defined by Kalbfleisch and Prentice (2002), the usual Cox shared frailty model is

$$\lambda(t|\mathbf{Z}_{ik}, W_i) = \lambda_0(t)W_i \exp\{\boldsymbol{\beta}'\mathbf{Z}_{ik}(t)\}, \quad (1.1)$$

where $\lambda_0(t)$ is the unspecified baseline hazard function, $\boldsymbol{\beta}$ is the $p \times 1$ unknown regression coefficient vector, and W_i is the unobservable frailty for the i th cluster. Conditional on W_i , the failure times are assumed to be independent. The most studied parametric distribution for W_i is the gamma distribution, referred to as gamma frailty. For the case of no covariates, the consistency and asymptotic distribution of the maximum likelihood estimator for the gamma frailty model have been well studied by Murphy (1994, 1995) and, for the case with covariates, by Parner (1998). In the Bayesian framework, extensive research has been carried out on the gamma frailty model (1.1), for example, in the work of Clayton (1991), Sinha (1993) and Gustafson (1997), among others. The positive stable distribution is another commonly used assumption for W_i (Hougaard (2000)) to account for heterogeneity among the intensity processes in different subjects. As opposed to (1.1), an alternative based on the Cox proportional hazards model is

$$\lambda(t|\mathbf{Z}_{ik}, \mathbf{X}_{ik}, \mathbf{b}_i) = \lambda_0(t) \exp\{\boldsymbol{\beta}'\mathbf{Z}_{ik}(t) + \mathbf{b}_i'\mathbf{X}_{ik}(t)\}, \quad (1.2)$$

where the $q \times 1$ random effect vectors \mathbf{b}_i ($i = 1, \dots, n$) are usually assumed to be independent and identically distributed (i.i.d.) from a zero-mean q -dimensional normal distribution, and $\mathbf{X}_{ik}(t)$ is the associated covariate vector. It is easy to see that (1.1) and (1.2) are equivalent when $\mathbf{X}_{ik}(t) \equiv 1$ and W_i in (1.1) is assumed to be log-normal.

However, the proportional hazards assumption might not hold in many practical situations. The underlying true relation of the hazards could be parallel, instead of proportional. Through a linear relationship between the covariate and the hazard function, Lin and Ying (1994) proposed the additive hazards model for independent survival data. For multivariate failure time data, we propose the additive hazards frailty model

$$\lambda(t|\mathbf{Z}_{ik}, \mathbf{X}_{ik}, \mathbf{b}_i) = \lambda_0(t) + \boldsymbol{\beta}'\mathbf{Z}_{ik}(t) + \mathbf{b}_i'\mathbf{X}_{ik}(t). \quad (1.3)$$

Apparently, one major difficulty in (1.3) is the non-negative hazard constraint, i.e., $\lambda_0(t) + \beta' \mathbf{Z}_{ik}(t) + \mathbf{b}'_i \mathbf{X}_{ik}(t) \geq 0$.

The Box-Cox transformation (Box and Cox (1964)) has been widely used in the linear regression model when the normality assumption of the error terms is not satisfied,

$$Y^{(\gamma)} = \begin{cases} (Y^\gamma - 1)/\gamma & \gamma \neq 0 \\ \log(Y) & \gamma = 0, \end{cases}$$

where γ is the transformation parameter. The application of this family of transformations in survival analysis is very limited, and has only been applied to independent failure time data. Breslow and Storer (1985) and Barlow (1985) applied this power transformation to the covariate structure to model the relative risk $R(\mathbf{Z}_i)$, $\log R(\mathbf{Z}_i) = \{(1 + \beta' \mathbf{Z}_i)^\gamma - 1\}/\gamma$. Aranda-Ordaz (1983) and Breslow (1985) argued the desirability of a compromise between the proportional and additive hazards models where, for grouped data, they imposed the transformation on the conditional probability $-\log\{1 - \Pr(t_{i-1} < T \leq t_i | T > t_{i-1})\}$, using the m partitions of the time scale $0 = t_0 < t_1 < \dots < t_m < \infty$. Particularly, the British Doctors Study (Breslow and Day (1987)) which was concerned with the effects of cigarette smoking on mortality, suggested that the hazard difference increases in t whereas the hazard ratio decreases. Thus a compromise between the proportional and the additive hazards models might capture this interesting phenomenon.

In this article, we propose a class of frailty models based on the Box-Cox transformation for clustered survival data. This family of transformation models is general and broad, including models (1.2) and (1.3) as special cases. By adding only one pre-specified power parameter, the modeling structure allows for a much richer class of hazard patterns. In many multivariate survival applications, where the hazards are neither proportional nor parallel, the proposed Box-Cox transformation frailty model provides a unified and flexible methodology.

In Section 2, we introduce this class of frailty models based on the Box-Cox transformed hazards, and derive the corresponding likelihood function within the Bayesian paradigm. In Section 3, we propose a prior specification that is suitable for the inherent constrained parameter problem. In Section 4, we derive the full conditional distributions needed for Gibbs sampling. In Section 5, we study the model selection criterion based on the conditional predictive ordinate (Geisser (1993)) for correlated data, and derive its corresponding MCMC approximation. We illustrate the proposed methods with the DRS data in Section 6.

2. A class of frailty models

Let T_{ik} ($i = 1, \dots, n; k = 1, \dots, K_i$) be the failure time for the k th subject in the i th cluster, C_{ik} be the censoring variable, and $Y_{ik} = \min(T_{ik}, C_{ik})$ be the

observed time. The failure time indicator $\nu_{ik} = 1$ if $Y_{ik} = T_{ik}$, and $\nu_{ik} = 0$ otherwise. Let $\mathbf{Z}_{ik}(t)$ be the corresponding $p \times 1$, and $\mathbf{X}_{ik}(t)$ be the $q \times 1$, bounded and possibly external time-dependent covariate vectors, where $\mathbf{X}_{ik}(t)$ usually contains 1 and part of $\mathbf{Z}_{ik}(t)$. Assume that T_{ik} and C_{ik} are conditionally independent given $\mathbf{Z}_{ik}(t)$ and $\mathbf{X}_{ik}(t)$. Within each cluster, $\{(T_{ik}, C_{ik}, \mathbf{Z}_{ik}(t), \mathbf{X}_{ik}(t)), k = 1, \dots, K_i\}$ may be dependent but are exchangeable.

Under this setup, we propose a class of frailty models by imposing the Box-Cox transformation on the hazards,

$$\frac{\lambda(t|\mathbf{Z}_{ik}, \mathbf{X}_{ik}, \mathbf{b}_i)^\gamma - 1}{\gamma} = \frac{\lambda_0(t)^\gamma - 1}{\gamma} + \boldsymbol{\beta}'\mathbf{Z}_{ik}(t) + \mathbf{b}_i'\mathbf{X}_{ik}(t). \quad (2.1)$$

Though γ is an identifiable parameter, we consider γ to be fixed and known throughout, since there is usually little information in the data to precisely estimate γ . The extension of the transformation on the hazard functions is natural due to their non-negativity. It is easy to see that model (2.1) reduces to (1.2) as $\gamma \rightarrow 0$, and reduces to (1.3) as $\gamma = 1$. By pre-specifying a set of γ 's, the rigidity of modeling structures (either proportional or parallel hazards) is completely relaxed and various hazard shapes become possible. Our primary interest has γ in $[0, 1]$, a compromise between the proportional ($\gamma = 0$) and the additive ($\gamma = 1$) hazards models, though γ can theoretically take any value on the real line. When $\gamma = 0$, it is reasonable to assume the \mathbf{b}_i 's are i.i.d. random vectors from $N_q(\mathbf{0}, \boldsymbol{\Sigma}_b)$. When $\gamma \in (0, 1]$, due to the non-negativity of the hazard function, it is more natural to assign a distribution with support on the positive real line to help alleviate the burden of the hazard constraint. Here, we take the \mathbf{b}_i 's to be i.i.d. from the first-order autoregressive correlated gamma distributions. Let $\mathbf{b}_i = (b_{[i,1]}, \dots, b_{[i,q]})'$. For $0 < \gamma \leq 1$, consider modeling the components of \mathbf{b}_i through a sequence of one-dimensional gamma distributions,

$$b_{[i,1]} \sim Ga(\zeta, \xi), \quad (b_{[i,k]}|b_{[i,k-1]}) \sim Ga(\zeta b_{[i,k-1]}, \xi), \quad k = 2, \dots, q. \quad (2.2)$$

This novel structure of modeling the random effects provides a very flexible and general class of distributions on the positive real line.

The hazard function based on model (2.1) is

$$\lambda(t|\mathbf{Z}_{ik}, \mathbf{X}_{ik}, \mathbf{b}_i) = [\lambda_0(t)^\gamma + \gamma\{\boldsymbol{\beta}'\mathbf{Z}_{ik}(t) + \mathbf{b}_i'\mathbf{X}_{ik}(t)\}]^{1/\gamma}. \quad (2.3)$$

For $0 < \gamma < 1$, $\boldsymbol{\beta}$, $\lambda_0(t)$ and $\mathbf{b} = (\mathbf{b}_1, \dots, \mathbf{b}_n)'$ are intertwined together under the power of $1/\gamma$, and hence cannot be separated in (2.3). The usual frequentist martingale and counting process techniques (Fleming and Harrington (1991)) might not be directly applicable in this situation. Therefore, we propose to conduct inference with this transformation frailty model using a Bayesian approach.

Toward this goal, a piecewise constant hazard is assumed for $\lambda_0(t)$ (Ibrahim, Chen and Sinha (2001)). Let J denote the number of partitions of the time

axis, i.e., $0 < s_1 < \dots < s_J$. For $J = 1$, namely with no partition, we obtain a parametric exponential model. By increasing J , we allow more flexibility in modeling the underlying baseline hazard. The piecewise constant hazard model assumes that $\lambda_0(y) = \lambda_j$ for $y \in (s_{j-1}, s_j]$, $j = 1, \dots, J$. The partition of the time axis needs to balance the number of failures among all the intervals, as well as guarantee that there is at least one failure in each interval. Define $\delta_{ikj} = 1$ if subject k in cluster i fails or is censored in interval j , and 0 otherwise. The likelihood function is constructed as follows. Let y_{ik} be the observed time for the k th subject in the i th cluster, $\mathbf{y} = (y_{11}, \dots, y_{1K_1}; \dots; y_{n1}, \dots, y_{nK_n})'$, $\boldsymbol{\nu} = (\nu_{11}, \dots, \nu_{1K_1}; \dots; \nu_{n1}, \dots, \nu_{nK_n})'$, the $(N \times p)$ -dimensional data matrix $\mathbf{Z}(t) = (\mathbf{Z}_1(t), \dots, \mathbf{Z}_p(t))$, and $(N \times q)$ -dimensional $\mathbf{X}(t) = (\mathbf{X}_1(t), \dots, \mathbf{X}_q(t))$, where N is the total sample size, i.e., $N = \sum_{i=1}^n K_i$. The hazard in the j th interval is $\lambda_j(t|\mathbf{Z}_{ik}, \mathbf{X}_{ik}, \mathbf{b}_i) = [\lambda_j^\gamma + \gamma\{\boldsymbol{\beta}'\mathbf{Z}_{ik}(t) + \mathbf{b}_i'\mathbf{X}_{ik}(t)\}]^{\delta_{ikj}/\gamma}$. Let $D = (N, \mathbf{y}, \mathbf{Z}(t), \mathbf{X}(t), \boldsymbol{\nu})$ denote the observed data and $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_J)'$. For ease of exposition, let $\mathbf{Z}_{ik} \equiv \mathbf{Z}_{ik}(t)$, $\mathbf{X}_{ik} \equiv \mathbf{X}_{ik}(t)$. Thus, the likelihood function is

$$L(\boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{b}|D) = \prod_{i=1}^n L_i(\boldsymbol{\beta}, \boldsymbol{\lambda}|\mathbf{b}_i, D)\pi(\mathbf{b}_i),$$

where

$$\begin{aligned} &L_i(\boldsymbol{\beta}, \boldsymbol{\lambda}|\mathbf{b}_i, D) \\ &= \prod_{k=1}^{K_i} \prod_{j=1}^J \{\lambda_j^\gamma + \gamma(\boldsymbol{\beta}'\mathbf{Z}_{ik} + \mathbf{b}_i'\mathbf{X}_{ik})\}^{\delta_{ikj}\nu_{ik}/\gamma} \exp[-\delta_{ikj}\{(\lambda_j^\gamma + \gamma(\boldsymbol{\beta}'\mathbf{Z}_{ik} + \mathbf{b}_i'\mathbf{X}_{ik}))^{1/\gamma} \\ &\quad \times (y_{ik} - s_{j-1}) + \sum_{l=1}^{j-1} (\lambda_l^\gamma + \gamma(\boldsymbol{\beta}'\mathbf{Z}_{ik} + \mathbf{b}_i'\mathbf{X}_{ik}))^{1/\gamma} (s_l - s_{l-1})\}], \end{aligned} \tag{2.4}$$

and $\pi(\mathbf{b}_i)$ is the density,

$$\begin{aligned} \pi(\mathbf{b}_i) &= \frac{1}{\sqrt{2\pi}|\boldsymbol{\Sigma}_b|^{1/2}} \exp\left(-\frac{1}{2}\mathbf{b}_i'\boldsymbol{\Sigma}_b^{-1}\mathbf{b}_i\right), \quad \text{when } \gamma = 0, \\ \pi(\mathbf{b}_i) &= \pi(b_{[i,1]}) \prod_{k=2}^q \pi(b_{[i,k]}|b_{[i,k-1]}), \quad \text{when } \gamma \in (0, 1], \end{aligned}$$

where $b_{[i,1]} \sim Ga(\zeta, \xi)$, $(b_{[i,k]}|b_{[i,k-1]}) \sim Ga(\zeta b_{[i,k-1]}, \xi)$, $k = 2, \dots, q$. It is appealing for \mathbf{b}_i to have a correlated gamma distribution for $\gamma \neq 0$, to alleviate the burden from the complex nonlinear constraints in $(\boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{b})$.

3. Prior Distributions

The joint prior distribution of $(\boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{b})$ needs to accommodate the non-negativity constraint for the hazard function, that is,

$$\lambda_j^\gamma + \gamma(\boldsymbol{\beta}'\mathbf{Z}_{ik} + \mathbf{b}_i'\mathbf{X}_{ik}) \geq 0 \quad (i = 1, \dots, n; k = 1, \dots, K_i; j = 1, \dots, J). \tag{3.1}$$

Bayesian computation and analysis become quite complicated with constrained parameter problems (Gelfand, Smith and Lee (1992), Chen and Shao (1998) and Chen, Shao and Ibrahim (2000, Chap.6)). If (3.1) is violated, the likelihood function and the posterior density are not well defined. For $\gamma = 0$ we assume that $\mathbf{b}_i \sim N_q(\mathbf{0}, \boldsymbol{\Sigma}_b)$, where $\boldsymbol{\Sigma}_b^{-1} \sim \text{Wishart}(\eta, (\eta\boldsymbol{\Omega})^{-1})$ with the density,

$$\pi(\boldsymbol{\Sigma}_b^{-1}) \propto \frac{|\boldsymbol{\Sigma}_b^{-1}|^{(\eta-q-1)/2}}{|\eta^{-1}\boldsymbol{\Omega}^{-1}|^{\eta/2}} \exp\left\{-\frac{1}{2}\text{trace}(\eta\boldsymbol{\Omega}\boldsymbol{\Sigma}_b^{-1})\right\},$$

$\boldsymbol{\Omega}$ is a $q \times q$ symmetric and positive definite parameter matrix, and $\eta > 0$. For $0 < \gamma \leq 1$, we take \mathbf{b}_i to be a set of first-order autoregressive correlated gamma random variables, $b_{[i,1]} \sim \text{Ga}(\zeta, \xi)$, and $(b_{[i,k]}|b_{[i,k-1]}) \sim \text{Ga}(\zeta b_{[i,k-1]}, \xi)$, $k = 2, \dots, q$. We fix ζ as a constant and give ξ a gamma prior as $\xi \sim \text{Ga}(a, d)$. This prior formulation can be easily generalized to higher-order autoregressive correlated gamma distributions.

One way to satisfy (3.1) is to specify an appropriately truncated joint prior distribution for $(\boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{b})$, for example a truncated multivariate normal prior for $(\boldsymbol{\beta}|\boldsymbol{\lambda}, \mathbf{b})$. This would lead to a prior distribution of the form

$$\pi(\boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{b}) = \pi(\boldsymbol{\beta}|\boldsymbol{\lambda}, \mathbf{b})\pi(\boldsymbol{\lambda}, \mathbf{b})I\{\lambda_j^\gamma + \gamma(\boldsymbol{\beta}'\mathbf{Z}_{ik} + \mathbf{b}'_i\mathbf{X}_{ik}) \geq 0, \text{ for all } i, k, j\},$$

where $I(\cdot)$ is the indicator function. Following this route, suppose that $(\boldsymbol{\beta}|\boldsymbol{\lambda}, \mathbf{b})$ has a p -dimensional truncated normal prior distribution with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}_\beta$. The normalizing constant under the constrained space,

$$c(\boldsymbol{\lambda}, \mathbf{b}) = \int \cdots \int_{\lambda_j^\gamma + \gamma(\boldsymbol{\beta}'\mathbf{z}_{ik} + \mathbf{b}'_i\mathbf{x}_{ik}) \geq 0, \text{ for all } i, k, j} \exp\left(-\frac{1}{2}\boldsymbol{\beta}'\boldsymbol{\Sigma}_\beta^{-1}\boldsymbol{\beta}\right) d\beta_1 \cdots d\beta_p,$$

does not have an analytic closed form, and thus the full conditionals are not tractable.

To circumvent the multivariate constrained parameter problem, we reduce it to a one-dimensional integral such that the normalizing constant can be obtained in a closed form. Without loss of generality, assume that all the covariates are positive. Let $\mathbf{Z}_{(-g)}$ denote the covariate matrix with the g th column deleted, and $\boldsymbol{\beta}_{(-g)}$ denote the $(p-1)$ -dimensional parameter vector with the g th component of $\boldsymbol{\beta}$ removed, thus $\mathbf{Z}_{(-g)} = (\mathbf{Z}_1, \dots, \mathbf{Z}_{g-1}, \mathbf{Z}_{g+1}, \dots, \mathbf{Z}_p)$ and $\boldsymbol{\beta}_{(-g)} = (\beta_1, \dots, \beta_{g-1}, \beta_{g+1}, \dots, \beta_p)'$. We propose a joint prior for $(\boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{b})$ of the form

$$\begin{aligned} \pi(\boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{b}) &= \pi(\beta_g|\boldsymbol{\beta}_{(-g)}, \boldsymbol{\lambda}, \mathbf{b})\pi(\boldsymbol{\beta}_{(-g)}, \boldsymbol{\lambda}, \mathbf{b}) \\ &\times I\left(\beta_g \geq -\frac{\lambda_j^\gamma + \gamma(\boldsymbol{\beta}'_{(-g)}\mathbf{Z}_{[ik,(-g)]} + \mathbf{b}'_i\mathbf{X}_{ik})}{\gamma Z_{[ik,g]}}\right), \end{aligned}$$

where $\mathbf{Z}_{[ik,(-g)]}$ and $Z_{[ik,g]}$ are covariates $\mathbf{Z}_{(-g)}$ and \mathbf{Z}_g corresponding to subject k in cluster i , respectively. This specification involves only one parameter β_g in the constraints, and leaves all the other parameters free of constraints. Specifically, we take $(\beta_g|\boldsymbol{\beta}_{(-g)}, \boldsymbol{\lambda}, \mathbf{b})$ to have a truncated normal distribution,

$$\begin{aligned} \pi(\beta_g|\boldsymbol{\beta}_{(-g)}, \boldsymbol{\lambda}, \mathbf{b}) &= c^{-1}(\boldsymbol{\beta}_{(-g)}, \boldsymbol{\lambda}, \mathbf{b}) \exp\left(-\frac{\beta_g^2}{2\sigma_g^2}\right) \\ &\times I\left\{\beta_g \geq -\frac{\lambda_j^\gamma + \gamma(\boldsymbol{\beta}'_{(-g)}\mathbf{Z}_{[ik,(-g)]} + \mathbf{b}'_i\mathbf{X}_{ik})}{\gamma Z_{[ik,g]}}\right\}, \end{aligned} \tag{3.2}$$

where the normalizing constant is given by

$$c(\boldsymbol{\beta}_{(-g)}, \boldsymbol{\lambda}, \mathbf{b}) = \sqrt{2\pi}\sigma_g \left[1 - \Phi\left(-\min_{i,k,j} \left\{\frac{\lambda_j^\gamma + \gamma(\boldsymbol{\beta}'_{(-g)}\mathbf{Z}_{[ik,(-g)]} + \mathbf{b}'_i\mathbf{X}_{ik})}{\gamma Z_{[ik,g]}\sigma_g}\right\}\right)\right], \tag{3.3}$$

and $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. For ease of exposition, we can take $\boldsymbol{\beta}_{(-g)}$, $\boldsymbol{\lambda}$ and \mathbf{b} to be independent a priori in (3.2), and we assume that the components of $\boldsymbol{\lambda}$ have independent gamma priori distributions.

4. Full Conditionals and Gibbs Sampling

When $\gamma = 0$, we have the Cox-type random effects model and the likelihood function is log-concave in $(\beta_1, \dots, \beta_p; \lambda_1, \dots, \lambda_J; \mathbf{b}_1, \dots, \mathbf{b}_n)$ (for details see Ibrahim *et al.* (2001, Chap.4)). For $0 < \gamma \leq 1$, the full conditionals of $(\beta_1, \dots, \beta_p)$ are log-concave, in which case the adaptive rejection sampling (ARS) proposed by Gilks and Wild (1992) is readily applicable. Due to the non-log-concavity of the full conditionals of $(\lambda_1, \dots, \lambda_J; \mathbf{b}_1, \dots, \mathbf{b}_n)$, a Metropolis step is required within the Gibbs iterations, referred to as the adaptive rejection Metropolis sampling (ARMS) proposed by Gilks, Best and Tan (1995). For each Gibbs sampling step, the parameter space needs to be set to satisfy the constraint (3.1), such that the likelihood function always has valid support.

The likelihood function given \mathbf{b} is $L(\boldsymbol{\beta}, \boldsymbol{\lambda}|\mathbf{b}, D) = \prod_{i=1}^n L_i(\boldsymbol{\beta}, \boldsymbol{\lambda}|\mathbf{b}_i, D)$. The full conditional distributions are $\pi(\beta_g|\boldsymbol{\beta}_{(-g)}, \boldsymbol{\lambda}, \mathbf{b}, D) \propto L(\boldsymbol{\beta}, \boldsymbol{\lambda}|\mathbf{b}, D)\pi(\beta_g)$, $\pi(\beta_l, l \neq g|\boldsymbol{\beta}_{(-l)}, \boldsymbol{\lambda}, \mathbf{b}, D) \propto L(\boldsymbol{\beta}, \boldsymbol{\lambda}|\mathbf{b}, D)\pi(\beta_l)c^{-1}(\boldsymbol{\beta}_{(-g)}, \boldsymbol{\lambda}, \mathbf{b})$, and $\pi(\lambda_j|\boldsymbol{\beta}, \boldsymbol{\lambda}_{(-j)}, \mathbf{b}, D) \propto L(\boldsymbol{\beta}, \boldsymbol{\lambda}|\mathbf{b}, D)\pi(\lambda_j)c^{-1}(\boldsymbol{\beta}_{(-g)}, \boldsymbol{\lambda}, \mathbf{b})$, where $\pi(\beta_l) \propto \exp\{-\beta_l^2/(2\sigma_l^2)\}$, and $\pi(\lambda_j) \propto \lambda_j^{\alpha-1} \exp(-\phi\lambda_j)$, for $i = 1, \dots, n$, $j = 1, \dots, J$ and $l = 1, \dots, p$. When $\gamma = 0$,

$$\begin{aligned} \pi(\mathbf{b}_i|\boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{b}_{(-i)}, \boldsymbol{\Sigma}_b, D) &\propto L_i(\boldsymbol{\beta}, \boldsymbol{\lambda}|\mathbf{b}_i, D)c^{-1}(\boldsymbol{\beta}_{(-g)}, \boldsymbol{\lambda}, \mathbf{b}) \exp\left(-\frac{1}{2}\mathbf{b}'_i\boldsymbol{\Sigma}_b^{-1}\mathbf{b}_i\right), \\ \pi(\boldsymbol{\Sigma}_b^{-1}|\mathbf{b}, D) &\propto \text{Wishart}\left(\eta + n, \left(\sum_{i=1}^n \mathbf{b}_i\mathbf{b}'_i + \eta\boldsymbol{\Omega}\right)^{-1}\right). \end{aligned}$$

When $\gamma \in (0, 1]$,

$$\pi(b_{[i,1]}|\boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{b}_{(-[i,1])}, \xi, D) \propto L_i(\boldsymbol{\beta}, \boldsymbol{\lambda}|\mathbf{b}_i, D)c^{-1}(\boldsymbol{\beta}_{(-g)}, \boldsymbol{\lambda}, \mathbf{b})b_{[i,1]}^{\zeta-1} \exp(-\xi b_{[i,1]}),$$

and, for $k = 2, \dots, q$,

$$\pi(b_{[i,k]}|\boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{b}_{(-[i,k])}, \xi, D) \propto L_i(\boldsymbol{\beta}, \boldsymbol{\lambda}|\mathbf{b}_i, D)c^{-1}(\boldsymbol{\beta}_{(-g)}, \boldsymbol{\lambda}, \mathbf{b})b_{[i,k]}^{\zeta b_{[i,k-1]}-1} \exp(-\xi b_{[i,k]}),$$

$$\pi(\xi|\mathbf{b}, D) \propto Ga\left(\zeta \sum_{i=1}^n (1 + b_{[i,1]} + \dots + b_{[i,q-1]}) + a, \sum_{i=1}^n (b_{[i,1]} + \dots + b_{[i,q]}) + d\right).$$

These full conditionals have nice tractable structures since $c(\boldsymbol{\beta}_{(-g)}, \boldsymbol{\lambda}, \mathbf{b})$ has a closed form. Remarkably, the posterior estimation is very robust with respect to the choice of g in (3.2).

5. Model Assessment

It is critical to compare a class of competing models for a given dataset and select the model that best fits the data. A suitable model selection criterion is based on the conditional predictive ordinate (CPO) statistics (Geisser (1993), Gelfand, Dey and Chang (1992) and Dey, Chen and Chang (1997)). Closely related to cross-validation, the CPO statistic is defined for each observation in the dataset by obtaining the conditional predictive density given the deletion of that observation.

Let $\mathbf{y}^{(-i)}$ denote the response vector with $\mathbf{y}_i = (y_{i1}, \dots, y_{iK_i})'$ deleted, where $\mathbf{y} = \{\mathbf{y}'_i, \mathbf{y}^{(-i)}\}'$. Let $\boldsymbol{\nu}^{(-i)}$ be the $(N - K_i) \times 1$ failure time indicator vector with $\boldsymbol{\nu}_i = (\nu_{i1}, \dots, \nu_{iK_i})'$ deleted, $\mathbf{Z}^{(-i)}$ be the $(N - K_i) \times p$ covariate matrix with the K_i rows of observations associated with the i th cluster deleted, and $\mathbf{X}^{(-i)}$ be defined similarly. The resulting observed data can be written as $D^{(-i)} = \{(N - K_i), \mathbf{y}^{(-i)}, \mathbf{Z}^{(-i)}, \mathbf{X}^{(-i)}, \boldsymbol{\nu}^{(-i)}\}$, i.e., the data with all the observations in the i th cluster deleted. The conditional density function of y_{ik} ($k = 1, \dots, K_i$; $i = 1, \dots, n$) is denoted by $f(y_{ik}|\mathbf{Z}_{ik}, \mathbf{X}_{ik}, \boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{b}_i)$. Define CPO_i to be the posterior predictive density of \mathbf{y}_i given $\mathbf{Z}_i = (\mathbf{Z}_{i1}, \dots, \mathbf{Z}_{iK_i})'$, $\mathbf{X}_i = (\mathbf{X}_{i1}, \dots, \mathbf{X}_{iK_i})'$ and $D^{(-i)}$, which can be written as $\text{CPO}_i = f(\mathbf{y}_i|\mathbf{Z}_i, \mathbf{X}_i, D^{(-i)})$. A more convenient form for CPO_i is

$$\text{CPO}_i = f(\mathbf{y}_i|\mathbf{y}^{(-i)}) = \frac{\int \int \int \frac{f(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{b}_i)\pi(\boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{b}_i)}{\pi(\boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{b}_i|\mathbf{y})f(\mathbf{y})} \pi(\boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{b}_i|\mathbf{y}) d\boldsymbol{\beta} d\boldsymbol{\lambda} d\mathbf{b}_i}{\int \int \int \frac{f(\mathbf{y}^{(-i)}|\boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{b}_i)\pi(\boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{b}_i)}{\pi(\boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{b}_i|\mathbf{y})f(\mathbf{y})} \pi(\boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{b}_i|\mathbf{y}) d\boldsymbol{\beta} d\boldsymbol{\lambda} d\mathbf{b}_i}.$$

Due to the conditional independence, we have

$$\text{CPO}_i = \left\{ \int \int \int \frac{\pi(\boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{b}_i|D)}{\prod_{k=1}^{K_i} f(y_{ik}|\mathbf{Z}_{ik}, \mathbf{X}_{ik}, \boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{b}_i)} d\boldsymbol{\beta} d\boldsymbol{\lambda} d\mathbf{b}_i \right\}^{-1}.$$

A Monte Carlo approximation of CPO_i for the proposed model, is given by

$$\widehat{CPO}_i = \left\{ \frac{1}{M} \sum_{m=1}^M \frac{1}{L_i(\boldsymbol{\beta}_{[m]}, \boldsymbol{\lambda}_{[m]} | \mathbf{b}_{i,[m]}, \mathbf{y}_i, \mathbf{Z}_i, \mathbf{X}_i)} \right\}^{-1},$$

where M is the number of Gibbs samples after burn-in, and

$$\begin{aligned} & L_i(\boldsymbol{\beta}_{[m]}, \boldsymbol{\lambda}_{[m]} | \mathbf{b}_{i,[m]}, \mathbf{y}_i, \mathbf{Z}_i, \mathbf{X}_i) \\ = & \prod_{k=1}^{K_i} \prod_{j=1}^J (\lambda_{j,[m]}^\gamma + \gamma(\boldsymbol{\beta}'_{[m]} \mathbf{Z}_{ik} + \mathbf{b}'_{i,[m]} \mathbf{X}_{ik}))^{\delta_{ikj} \nu_{ik} / \gamma} \\ & \times \exp[-\delta_{ikj} \{(\lambda_{j,[m]}^\gamma + \gamma(\boldsymbol{\beta}'_{[m]} \mathbf{Z}_{ik} + \mathbf{b}'_{i,[m]} \mathbf{X}_{ik}))^{1/\gamma} \\ & \times (y_{ik} - s_{j-1}) + \sum_{l=1}^{j-1} (\lambda_{l,[m]}^\gamma + \gamma(\boldsymbol{\beta}'_{[m]} \mathbf{Z}_{ik} + \mathbf{b}'_{i,[m]} \mathbf{X}_{ik}))^{1/\gamma} (s_l - s_{l-1})\}]. \end{aligned}$$

Here $\boldsymbol{\beta}_{[m]} = (\beta_{1,[m]}, \dots, \beta_{p,[m]})'$, $\boldsymbol{\lambda}_{[m]} = (\lambda_{1,[m]}, \dots, \lambda_{J,[m]})'$ and $\mathbf{b}_{i,[m]}$ ($i = 1, \dots, n$) are the samples of the m th Gibbs iteration. We summarize the CPO statistics across all the clusters with $B = \sum_{i=1}^n \log(CPO_i)$, where a larger value of B indicates a better fit of a model.

6. The DRS Example

In many applications, one can apply (2.1) to the data with a set of pre-specified γ 's and choose the best fitting model according to a suitable model selection criterion. As an illustration, we applied the proposed Box-Cox transformation frailty model to the DRS example. The response variable was time to blindness (in months), which could be right-censored. The covariates in this analysis were treatment (laser photocoagulation and control), the type of diabetes (juvenile with the age at diagnosis < 20 years, and adult), and age (a standardized continuous variable).

Without loss of generality, we considered $q = 1$ in this example. For $\gamma = 0$, we defined $\tau = \sigma_b^{-2}$ and assumed $\tau \sim Ga(0.001, 0.001)$ in order to obtain a noninformative prior. For $\gamma \in (0, 1]$, we assumed $b_i \sim Ga(\zeta, \xi)$, where we fixed the shape parameter at $\zeta = 10^{-6}$, and assigned a gamma prior for the scale parameter, $\xi \sim Ga(a, d)$. The unconditional prior mean and variance of b_i are given by,

$$\begin{aligned} E(b_i) &= \frac{\zeta d}{a - 1}, \\ \text{var}(b_i) &= \frac{\zeta d^2}{(a - 1)(a - 2)} + \zeta^2 \left\{ \frac{d^2}{(a - 1)(a - 2)} - \left(\frac{d}{a - 1} \right)^2 \right\}. \end{aligned}$$

Hence, we chose the hyperparameter $a = 3$ and $d = 10^5$ to obtain a vague prior distribution for b_i , with $\text{var}(b_i) \approx 5,000$.

We constrained the regression coefficient for laser treatment (β_1) to have a truncated normal prior. The priors for $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)'$ and $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_J)'$ were taken to be noninformative, where $(\beta_1 | \beta_2, \beta_3, \boldsymbol{\lambda}, \mathbf{b})$ had the truncated $N(0, 10^4)$ prior as defined in (3.2), β_2 and β_3 were independent a priori and taken to have $N(0, 10^4)$ distributions, and $\lambda_j \sim Ga(\alpha, \phi)$ with $\alpha = 2$ and $\phi = 0.001$, and independent for $j = 1, \dots, J$. We specified priors in such a way that the likelihood clearly dominated the posterior distribution, which would allow for a fair comparison between different models, particularly for $\gamma = 0$ and $\gamma \in (0, 1]$. The shape and flexibility of the baseline hazard function is controlled by the choice of J . The finer the partition of the time axis, the more general the pattern of the hazard that is captured. However, by increasing J , we introduce more unknown parameters (the λ_j 's) to be estimated, hence there should be some optimal J due to this trade-off. The parameter γ also directly affects the shapes of the hazard functions, and there is much interplay between J and γ in controlling the shapes of the hazards. To search for a suitable model, we set $J = (1, 2, 3, 4, 5)$ and $\gamma = (0, 0.25, 0.5, 0.75, 1)$. In this two-dimensional grid, we would locate the point (J, γ) which yielded the largest B statistic, and the corresponding model was then deemed to be the best fitting one.

The posterior computations were based on 30,000 Gibbs iterations with a burn-in of 2,000 samples. The B statistics for model selection are summarized in Table 1. Clearly, $J = 4$ and $\gamma = 0.5$ yields the largest CPO statistic ($B = -827.10$) and the corresponding model is deemed to be the best fitting one. Table 2 summarizes the posterior mean, standard deviation, and the 95% HPD interval for $J = 1$ and $J = 4$. Across these competing models, the signs of the three regression parameters are the same and they all show a significant effect of laser treatment and a nonsignificant effect of age. However, there are some discrepancies on the effect of the type of diabetes among different models. The selected best fitting model ($J = 4$ and $\gamma = 0.5$) shows a significant effect of the type of diabetes on the time to blindness. Patients with the adult type of diabetes had significantly longer vision survival than those with the juvenile type.

Table 1. Grid search for the largest B statistic based on the CPO's with respect to J and γ in the DRS data.

		J				
		1	2	3	4	5
γ	0	-828.88	-831.56	-834.60	-837.21	-843.15
	0.25	-1032.43	-828.40	-828.46	-829.41	-834.85
	0.5	-977.92	-829.00	-828.46	-827.10	-832.75
	0.75	-990.16	-827.61	-827.17	-827.38	-828.38
	1	-987.80	-827.51	-827.53	-827.21	-828.86

Table 2. Posterior means, standard deviations, and 95% HPD intervals for the DRS data.

J	γ	Covariate	Mean	Std. Dev.	95% HPD Interval
1	0	Treatment	0.8997	0.1819	(0.5504, 1.2636)
		Type of Diabetes	-0.6094	0.4647	(-1.5639, 0.2586)
		Age	0.3082	0.2284	(-0.1250, 0.7384)
	0.25	Treatment	0.2655	0.0545	(0.1578, 0.3724)
		Type of Diabetes	-0.0808	0.1019	(-0.2824, 0.1184)
		Age	0.0443	0.0545	(-0.0649, 0.1491)
	0.5	Treatment	0.0772	0.0160	(0.0455, 0.1086)
		Type of Diabetes	-0.0321	0.0316	(-0.0961, 0.0268)
		Age	0.0110	0.0174	(-0.0220, 0.0456)
	0.75	Treatment	0.0225	0.0046	(0.0135, 0.0314)
		Type of Diabetes	-0.0137	0.0078	(-0.0290, 0.0013)
		Age	0.0040	0.0048	(-0.0052, 0.0136)
1	Treatment	0.0062	0.0013	(0.0037, 0.0088)	
	Type of Diabetes	-0.0039	0.0017	(-0.0070, -0.0003)	
	Age	0.0008	0.0012	(-0.0012, 0.0033)	
4	0	Treatment	0.5502	0.1741	(0.2050, 0.8900)
		Type of Diabetes	-1.3971	0.4818	(-2.4140, -0.4892)
		Age	0.5008	0.2224	(-0.0795, 0.9564)
	0.25	Treatment	0.1830	0.0521	(0.0797, 0.2839)
		Type of Diabetes	-0.2518	0.1131	(-0.4756, -0.0292)
		Age	0.0899	0.0585	(-0.0275, 0.2017)
	0.5	Treatment	0.0593	0.0156	(0.0288, 0.0894)
		Type of Diabetes	-0.0724	0.0287	(-0.1333, -0.0180)
		Age	0.0248	0.0163	(-0.0067, 0.0574)
	0.75	Treatment	0.0178	0.0046	(0.0087, 0.0266)
		Type of Diabetes	-0.0183	0.0069	(-0.0314, -0.0043)
		Age	0.0052	0.0046	(-0.0033, 0.0143)
1	Treatment	0.0047	0.0013	(0.0022, 0.0072)	
	Type of Diabetes	-0.0042	0.0015	(-0.0072, -0.0012)	
	Age	0.0009	0.0011	(-0.0008, 0.0034)	

We ran parallel chains with diverse starting values to evaluate the MCMC convergence properties. These chains converged to the same range of values for each parameter, and all the regression parameters appeared to mix satisfactorily. The 50% and 97.5% quantiles of the sampling distribution for the shrinking factor reported by the Gelman and Rubin (1992) diagnostics were very close to 1 for all the regression parameters.

To investigate the robustness of the proposed model and the influence of the prior distributions, we carried out several sensitivity analyses by varying the hyperparameters. Table 3 shows that the posterior estimates are very robust using noninformative prior distributions under a wide range of hyperparameters. We conducted another set of sensitivity analyses, summarized in Table 4, to examine the choice of the constrained regression parameter (β_1, β_2 , or β_3) for the truncated normal prior distribution. The results are very robust with respect to the constrained parameter, which demonstrates the feasibility of the proposed prior specification.

Table 3. Sensitivity analysis with different hyperparameters on the priors for the DRS data using $J = 4$ and $\gamma = 0.5$.

σ_β	ϕ	(ζ, d)	Covariate	Mean	Std. Dev.	95% HPD Interval
5	0.001	$(10^{-6}, 10^5)$	Treatment	0.0594	0.0155	(0.0289, 0.0895)
			Type of Diabetes	-0.0764	0.0288	(-0.1333, -0.0204)
			Age	0.0274	0.0166	(-0.0049, 0.0600)
10	0.001	$(10^{-6}, 10^5)$	Treatment	0.0593	0.0160	(0.0280, 0.0902)
			Type of Diabetes	-0.0701	0.0296	(-0.1277, -0.0125)
			Age	0.0237	0.0168	(-0.0096, 0.0564)
100	0.1	$(10^{-6}, 10^5)$	Treatment	0.0596	0.0158	(0.0287, 0.0907)
			Type of Diabetes	-0.0755	0.0317	(-0.1390, -0.0151)
			Age	0.0259	0.0175	(-0.0080, 0.0606)
100	0.01	$(10^{-6}, 10^5)$	Treatment	0.0586	0.0158	(0.0279, 0.0895)
			Type of Diabetes	-0.0740	0.0300	(-0.1325, -0.0157)
			Age	0.0253	0.0168	(-0.0075, 0.0582)
100	0.001	$(10^{-4}, 10^3)$	Treatment	0.0590	0.0159	(0.0275, 0.0904)
			Type of Diabetes	-0.0779	0.0291	(-0.1359, -0.0199)
			Age	0.0291	0.0164	(-0.0035, 0.0612)
100	0.001	$(10^{-5}, 10^4)$	Treatment	0.0596	0.0160	(0.0282, 0.0904)
			Type of Diabetes	-0.0700	0.0300	(-0.1270, -0.0105)
			Age	0.0227	0.0170	(-0.0116, 0.0553)

Table 4. Analysis of the DRS data with different regression parameters having truncated normal priors as in (3.2), using $J = 4$ and $\gamma = 0.5$.

Truncated	Covariate	Regression Coefficient	Mean	Std. Dev.	95% HPD Interval
<i>Type of Diabetes</i>	Treatment		0.0594	0.0156	(0.0291, 0.0902)
	Type of Diabetes		-0.0754	0.0290	(-0.1318, -0.0178)
	Age		0.0270	0.0166	(-0.0065, 0.0592)
<i>Age</i>	Treatment		0.0598	0.0156	(0.0290, 0.0908)
	Type of Diabetes		-0.0715	0.0294	(-0.1289, -0.0152)
	Age		0.0239	0.0167	(-0.0079, 0.0565)

Acknowledgements

We would like to thank the Editor, an associate editor and a referee for their constructive comments and suggestions which led to substantial improvement of the article. We thank Donald Berry and Peter Müller for helpful discussions. This research was partially supported by the National Cancer Institute Grant 74015 and National Institute of General Medical Sciences Grant GM 070335.

References

- Aranda-Ordaz, F. J. (1983). An extension of the proportional-hazards model for grouped data. *Biometrics* **39**, 109-117.
- Barlow, W. E. (1985). General relative risk models in stratified epidemiologic studies. *Appl. Statist.* **34**, 246-257.
- Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations (with discussion). *J. Roy. Statist. Soc. B* **26**, 211-252.
- Breslow, N. E. (1985). Cohort analysis in epidemiology. In *A Celebration of Statistics* (Edited by A. C. Atkinson and S. E. Fienberg), 109-143. Springer, New York.
- Breslow, N. E. and Storer, B. E. (1985). General relative risk functions for case-control studies. *Amer. J. Epidemiology* **122**, 149-162.
- Breslow, N. E. and Day, N. E. (1987). *Statistical Methods in Cancer Research, 2, The Design and Analysis of Case-Control Studies*. IARC, Lyon.
- Chen, M. and Shao, Q. (1998). Monte Carlo methods for Bayesian analysis of constrained parameter problems. *Biometrika* **85**, 73-87.
- Chen, M., Shao, Q. and Ibrahim, J. G. (2000). *Monte Carlo Methods in Bayesian Computation*. Springer, New York.
- Clayton, D. G. (1991). A Monte Carlo method for Bayesian inference in frailty models. *Biometrics* **47**, 476-485.
- Cox, D. R. (1972). Regression models and life tables (with discussion). *J. Roy. Statist. Soc. Ser. B* **34**, 187-220.
- Dey, D. K., Chen, M. and Chang, H. (1997). Bayesian approach for nonlinear random effects models. *Biometrics* **53**, 1239-1252.
- Diabetic Retinopathy Study Research Group (1985). Diabetic Retinopathy Study. *Invest. Ophthalmol. Visual Science* **21** (part 2), 149-226.
- Fleming, T. R. and Harrington, D. (1991). *Counting Processes and Survival Analysis*. John Wiley, New York.
- Geisser, S. (1993). *Predictive Inference: an Introduction*. Chapman and Hall, London.
- Gelfand, A. E., Dey, D. K. and Chang, H. (1992). Model determination using predictive distributions with implementation via sampling based methods (with discussion). In *Bayesian Statistics 4* (Edited by J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith), 147-167. Oxford University Press, Oxford.
- Gelfand, A. E., Smith, A. F. M. and Lee, T. (1992). Bayesian analysis of constrained parameter and truncated data problems using Gibbs sampling. *J. Amer. Statist. Assoc.* **87**, 523-532.
- Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statist. Science* **7**, 457-511.
- Gilks, W. R., Best, N. G. and Tan, K. K. C. (1995). Adaptive rejection Metropolis sampling within Gibbs sampling. *Appl. Statist.* **44**, 455-472.
- Gilks, W. R. and Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Appl. Statist.* **41**, 337-348.

- Gustafson, P. (1997). Large hierarchical Bayesian analysis of multivariate survival data. *Biometrics* **53**, 230-242.
- Hougaard, P. (2000). *Analysis of Multivariate Survival Data*. Springer, New York.
- Ibrahim, J. G., Chen, M. and Sinha, D. (2001). *Bayesian Survival Analysis*. Springer, New York.
- Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*, 2nd ed. John Wiley, New York.
- Lin, D. Y. and Ying, Z. (1994). Semiparametric analysis of the additive risk model. *Biometrika* **81**, 61-71.
- Murphy, S. A. (1994). Consistency in a proportional hazards model incorporating a random effect. *Ann. Statist.* **22**, 712-731.
- Murphy, S. A. (1995). Asymptotic theory for the frailty model. *Ann. Statist.* **23**, 182-198.
- Parner, E. (1998). Asymptotic theory for the correlated Gamma-frailty model. *Ann. Statist.* **26**, 183-214.
- Sinha, D. (1993). Semiparametric Bayesian analysis of multiple event time data. *J. Am. Statist. Assoc.* **88**, 979-983.

Department of Biostatistics and Applied Mathematics, M. D. Anderson Cancer Center, The University of Texas, Houston, TX, U.S.A.

E-mail: gyin@odin.mdacc.tmc.edu

Department of Biostatistics, The University of North Carolina at Chapel Hill, Chapel Hill, NC, U.S.A.

E-mail: ibrahim@bios.unc.edu

(Received January 2004; accepted July 2004)