Lifetime Data Analysis, 3, 123–137 (1997)

A Weibull Regression Model with Gamma Frailties for Multivariate Survival Data

SUJIT K. SAHU s.sahu@statslab.cam.ac.uk
Statistical Laboratory, University of Cambridge, UK

DIPAK K. DEY AND HELEN ASLANIDOU dey@stat.uconn.edu
Department of Statistics, University of Connecticut, Storrs, USA

DEBAJYOTI SINHA sinha@purabi.unh-edu
Department of Mathematics, University of New Hampshire, USA

Received March 5, 1996; Accepted February 10, 1997

Abstract. Frequently in the analysis of survival data, survival times within the same group are correlated due to unobserved co-variates. One way these co-variates can be included in the model is as frailties. These frailty random block effects generate dependency between the survival times of the individuals which are conditionally independent given the frailty. Using a conditional proportional hazards model, in conjunction with the frailty, a whole new family of models is introduced. By considering a gamma frailty model, often the issue is to find an appropriate model for the baseline hazard function. In this paper a flexible baseline hazard model based on a correlated prior process is proposed and is compared with a standard Weibull model. Several model diagnostics methods are developed and model comparison is made using recently developed Bayesian model selection criteria. The above methodologies are applied to the McGilchrist and Aisbett (1991) kidney infection data and the analysis is performed using Markov Chain Monte Carlo methods.

Keywords: Autocorrelated prior process, conditional predictive ordinate, frailty, Markov chain Monte Carlo methods, model determination, posterior predictive loss, proportional hazards model, Weibull model.

1. Introduction

Modeling dependence in multivariate survival data has received considerable attention in recent literature. Such data sets may come from subjects of the same group who are related to each other. For example, a typical group may consist of members from the same family, e.g., brothers and sisters. In a different context, the data may come from multiple recurrence times of a disease for the same patient. Straightforwardly, because of dependence in the data Cox proportional hazard models can not be used.

A key development in modeling such data is to consider the frailty models. The idea builds upon a familiar repeated measures trick. The event times are conditionally independent given the *frailty*, an individual random effect, see e.g., Clayton (1978), Oakes (1982) and Clayton and Cuzick (1985). These models formulate the variability of life times, coming from two distinct sources. The first source is natural variability and it is explained by the hazard function and the second is variability common to individuals of the same group or variability common to several events of an individual and it is explained by the frailty.

The frailty term in a model represents the common co-variates that are not observed or are neglected.

Many models compete for adoption as the baseline hazard function, see e.g., Sinha and Dey (1997) (and references therein) for a review. They range from simple constant hazard functions to Lévy processes, beta processes and so on. An attractive development here is to use a correlated prior process, (see e.g., Leonard, 1978; Gamerman, 1991 and Arjas and Gasbarra, 1994). Hazards in adjacent intervals are assumed to be correlated. A martingale prior process emerges from such considerations. Though, in general these processes do not produce monotonic hazard rate models, they impose smoothness on the hazard functions in adjacent intervals.

The issues of model adequacy and model choice has received far less attention in the survival analysis literature. In the frequentist analysis ad-hoc residual plots serve as model checking criterion since the residuals for the censored individuals do not have the desired distributions. Recent advances in Bayesian computations using Markov chain Monte Carlo (MCMC) methods (see e.g., Gilks *et al.*, 1996 for an excellent review) have enabled effective deliberations to be made on such questions using predictive distributions. MCMC methods enable the calculation of several features of the posterior distributions and help formulation of model choice criterion based on posterior predictive loss (Gelfand and Ghosh, 1997).

In this article, we consider conditional proportional hazard models given the frailty. We present many different parametric assumptions on the frailty models, considering both multiplicative and additive frailty distributions. In particular, we develop models based on multiplicative gamma frailty (Clayton and Cuzick, 1985 and also Sinha and Dey, 1997 for a review). We model the survival times using conditional piecewise exponential distribution (see, e.g., Breslow, 1974; Kalbfleisch and Prentice, 1973 and Gamerman, 1991). For the piecewise exponential model, we use correlated prior processes (Gamerman, 1991 and Arjas and Gasbarra, 1994) for the baseline hazard functions. We also consider the widely used Weibull hazard functions. We compare the piecewise exponential models with models built up using Weibull distributions. We illustrate our methodology for the kidney infection data set of McGilchrist and Aisbett (1991).

The remainder of the paper is organized as follows. Section 2 introduces the conditional proportional hazard model with frailty, the correlated prior process for the baseline hazard and the Weibull baseline hazard model. In Section 3 we discuss the issues of model adequacy and model choice using predictive distributions. In Section 4 we analyze a data set from McGilchrist and Aisbett (1991). Some important mathematical steps in our calculation are placed in the Appendix.

2. Frailty models

2.1. Hazard Functions Modeling

Clayton (1978) and Oakes (1982) considered first frailty models for multivariate survival data, using gamma distribution for the frailty. Hougaard (1986) used the positive stable model while Whitmore and Lee (1991) studied a model with inverse gamma frailties. We consider the gamma distribution which is most common to model the frailty. Assuming

that the survival time of the jth subject $(j = 1, \ldots, m)$ in the ith group $(i = 1, \ldots, n)$ is denoted by T_{ij} and given the unobserved frailty parameter denoted by w_i (for the ith group), the hazard function is as follows:

$$h(t_{ij}|\mathbf{z}_{ij}, w_i) \;=\; \lambda_o(t_{ij}) \exp(\boldsymbol{\beta}^T \mathbf{z}_{ij})\; w_i \tag{1}$$

where $\mathbf{z}_{ij}$ is the fixed co-variate vector, $\boldsymbol{\beta}$ is the regression parameter and $\lambda_o(\cdot)$ is the baseline hazard function. As is often done in practice, for identifiability purposes the linear model component in (1), $\log(\theta_{ij}) \equiv \boldsymbol{\beta}^T \mathbf{z}_{ij}$, does not include any intercept term. Components of $\boldsymbol{\beta}$ are assigned independent normal priors with large variances. The co-variate $\mathbf{z}_{ij}$ can be time dependent. The frailty parameters $w_i,\ i = 1, \ldots, n$ are assumed to be independent and identically distributed for every group. For identifiability reasons we need the w_i's to have mean one. In this paper we consider w_i to follow a gamma distribution, i.e.,

$$w_i \;\sim\; Gamma(\eta,\, \eta),\ i = 1, \ldots, n \tag{2}$$

where η^{-1} is the unknown variance of the w_i's. That is, we say $X \sim Gamma(a, b)$ if its density is $\propto x^{a-1} \exp(-bx)$. Small values of η signify closer positive relationship between the subjects of the same group and greater heterogeneity among the groups. An appropriate prior for η is a Gamma distribution with mean 1 and large variance, $Gamma(\phi, \phi)$ say with a small choice of ϕ.

2.2. Piecewise Exponential Models

Piecewise exponential models and prior processes on the components provide a very flexible framework for modeling univariate survival data. Modeling the baseline hazard using prior processes is very common, see Sinha and Dey (1997) for a review. Often in real life problems, not the actual baseline hazard, but the smoothness of it, is available as prior information (see e.g., Leonard, 1978 and Gamerman, 1991). We divide time into g pre-specified intervals $I_k = (t_{k-1}, t_k]$ for $k = 1, 2, \ldots, g$ where $0 = t_0 < t_1 < \ldots < t_g < \infty$, t_g being the last survival or censored time and assume the baseline hazard to be constant within intervals. That is,

$$\lambda_o(t_{ij}) = \lambda_k, \;\text{ for }\; t_{ij} \in I_k. \tag{3}$$

The model was first introduced by Breslow (1974) who used distinct failure times as end points of intervals. Kalbfleisch and Prentice (1973) suggested that the selection of the grid $\{t_1, t_2, \ldots, t_g\}$ should be made independent of the data. We will discuss the choice of g later in this section.

To correlate the λ_k's in adjacent intervals, a discrete-time martingale process is used, similar to that of Arjas and Gasbarra's (1994) for univariate survival model. Given $(\lambda_1, \ldots, \lambda_{k-1})$ we specify that

$$\lambda_k|\lambda_1, \ldots, \lambda_{k-1} \;\sim\; Gamma\left(\alpha_k, \frac{\alpha_k}{\lambda_{k-1}}\right), k = 1, \ldots, g \tag{4}$$

where $\lambda_0 = 1$. Hence, $E(\lambda_k|\lambda_1, \ldots, \lambda_{k-1}) = \lambda_{k-1}$. The parameter α_k in (4) controls the amount of smoothness available, i.e., small α_k indicates less information on the smoothing of λ_k's. If $\alpha_k = 0$ then λ_k and λ_{k-1} are independent. When $\alpha_k \to \infty$ then the baseline hazard is the same in the intervals I_k and I_{k-1} i.e., $\lambda_k = \lambda_{k-1}$. Though notice that if prior information is available, the shape and scale of the marginal prior of λ_k will change accordingly and the prior will be more informative. A version of the above process which can also be used, was given by e.g., Leonard (1978) and Gamerman (1991). They modeled $\log(\lambda_k) = c_k$ and

$$c_k|c_{k-1} \sim N(c_{k-1}, \tau^2),\ k = 1, \ldots, g \tag{5}$$

with $c_0 = 0$. Taking this further, we can assume a second difference prior process for c_k, i.e., $c_k|c_{k-2}, c_{k-1} \sim N(2c_{k-1} - c_{k-2},\ \tau^2), k = 3, \ldots, g$.

A few remarks are in order on the choice of g. It is clear that a very large choice of g will make the model non-parametric. However, in the parametric set-up of this paper, too large a g will produce unstable estimators of the λ's and too small a choice will lead to poor model fitting. Hence, a robust choice of g should be considered here. Note that, the maximum likelihood estimate of λ_k depend on the number of failures, d_k, in the kth interval I_k and is 0 if d_k is zero. One possible advantage of the Bayesian approach with the correlated process prior described here is to smooth out such jumps to zero. However, see the Appendix for further discussion on related computational issues. A random choice of g will make the posterior distribution in variable dimensions and sampling techniques other than the Gibbs sampler, e.g., reversible jump MCMC, Green (1995), can be used to compute the posterior distribution.

The above models can be easily altered to accommodate monotone baseline hazard functions. Suppose that one intends to model the λ's with the constraint $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_g$. Following Arjas and Gasbarra (1994) we can assume that

$$\lambda_k - \lambda_{k-1} \sim Gamma(\alpha_k, \alpha_k),\ k = 1, \ldots, g$$

instead of (4) or (5). However we do not consider these here and instead turn to the Weibull models which can accommodate monotonicity.

2.3. Weibull Models

To compare the performance of the above piecewise exponential models, we consider the usual Weibull distribution for modeling. The Weibull baseline hazard function given by

$$\lambda_o(t_{ij}) = \mu\alpha t_{ij}^{\alpha-1},\ \alpha,\ \mu > 0, j = 1, \ldots, m,\ \ i = 1, \ldots, n \tag{6}$$

where α and μ are unknown hyper-parameters, has been used extensively due to its simplicity and flexibility. We assume that a-priori $\mu \sim Gamma(\rho, \rho)$ and $\alpha \sim Gamma(\kappa_1, \kappa_2)$.

The above is a multiplicative frailty model. We can formulate an additive frailty model by assuming, instead of (1) and (6)

$$h(t_{ij}|\mathbf{z}_{ij}, b_i) = \xi_{ij}\alpha t_{ij}^{\alpha-1},\ \text{where}\ \log(\xi_{ij}) = \nu + \boldsymbol{\beta}^T\mathbf{z}_{ij} + b_i. \tag{7}$$

Now b_i's are assumed i.i.d $N\left(0, \eta^{-1}\right)$ and η is given a $Gamma(\phi, \phi)$ prior. Prior for α remains the same and ν is given a flat normal prior. This is the modeling strategy used in the BUGS, Spiegelhalter *et al.* (1995), software manual. However, it is expected that both the Weibull model formulations draw similar inference. We will return to this in Section 4.

2.4. Likelihood Specifications

Let δ_{ij} denote the indicator variable taking value 1 if the jth subject $(j = 1, \ldots, m)$ of the ith group $(i = 1, \ldots, n)$ fails and value 0 otherwise. Hence t_{ij} is a failure time if $\delta_{ij} = 1$ and a censoring time otherwise. Let $\mathbf{z}_{ij}$ be the co-variate for each subject. Hence, the triplet $(t_{ij}, \delta_{ij}, \mathbf{z}_{ij})$ is observed for all i and j. Let $(\mathbf{Y}, \mathbf{Z})$ denote the collection of all such triplets $(t_{ij}, \delta_{ij}, \mathbf{z}_{ij})$. The vector of unobserved w_i's, denoted by $\mathbf{W}$, is called the *augmented data* and the triplet $(\mathbf{W}, \mathbf{Y}, \mathbf{Z})$ is called the *complete data*. We only allow right-censored survival data and assume that the censoring is non-informative.

For the piecewise exponential model, the likelihood can be derived as follows. The jth subject of the ith group has a constant hazard of $h_{ij} = \lambda_k \theta_{ij} w_i$ in the kth interval $(k = 1, \ldots, g)$ given the unobserved frailty w_i. If the subject has survived beyond the kth interval, i.e., $t_{ij} > t_k$, the likelihood contribution is $\exp\{-\lambda_k \Delta_k \theta_{ij} w_i\}$ where $\Delta_k = t_k - t_{k-1}$. If the subject has failed or censored in the kth interval, i.e., $t_{k-1} < t_{ij} \leq t_k$ then the likelihood contribution is $\left(\lambda_k \theta_{ij} w_i\right)^{\delta_{ij}} \exp\left\{-\lambda_k (t_{ij} - t_{k-1}) \theta_{ij} w_i\right\}$. Hence, we arrive at the following complete data likelihood, $L(\boldsymbol{\beta}, \boldsymbol{\lambda}, |\mathbf{W}, \mathbf{Y}, \mathbf{Z})$ say,

$$\prod_{i=1}^{n} \prod_{j=1}^{m} \left\{ \prod_{k=1}^{g_{ij}} \exp(-\lambda_k \Delta_k \theta_{ij} w_i) \right\} \left(\lambda_{g_{ij}+1} \theta_{ij} w_i\right)^{\delta_{ij}} \exp\left\{-\lambda_{g_{ij}+1} (t_{ij} - t_{g_{ij}}) \theta_{ij} w_i\right\}, \tag{8}$$

where g_{ij} is such that $t_{ij} \in (t_{g_{ij}}, t_{g_{ij}+1}] = I_{g_{ij}+1}$. The data likelihood of $(\boldsymbol{\beta}, \boldsymbol{\lambda})$ based on the observed data $(\mathbf{Y}, \mathbf{Z})$ can be obtained by integrating out the w_i's from (8) with the density $\pi(w_i|\eta)$ as given in (2).

Using a Weibull hazard with parameters α and μ as given in (6), the jth subject of the ith group has a hazard $h_{ij} = \mu \alpha t_{ij}^{\alpha-1} \theta_{ij} w_i$. Given the unobserved frailty w_i, t_{ij}'s are independent. Hence, the complete data likelihood, $L(\boldsymbol{\beta}, \mu, \alpha|\mathbf{W}, \mathbf{Y}, \mathbf{Z})$ say, is given by

$$\prod_{i=1}^{n} \prod_{j=1}^{m} \left(\mu \alpha t_{ij}^{\alpha-1} \theta_{ij} w_i\right)^{\delta_{ij}} \exp\left\{-\mu t_{ij}^{\alpha} \theta_{ij} w_i\right\}. \tag{9}$$

Once again the data likelihood of $(\boldsymbol{\beta}, \mu, \alpha)$ based on observed data $(\mathbf{Y}, \mathbf{Z})$ can be obtained by integrating out the w_i's from (8) with the density $\pi(w_i|\eta)$ as given in (2).

The final forms of the data likelihoods after integration are too complicated to work with. Thus, it is not easy to evaluate the marginal posterior distributions of $(\boldsymbol{\beta}, \mu, \alpha)$ and $(\boldsymbol{\beta}, \boldsymbol{\lambda})$ analytically. To circumvent this problem, we use the Gibbs sampler, see e.g., Gelfand and Smith (1990) and Gilks *et al.* (1996) with the data augmentation method (Tanner and Wong, 1987) to generate samples from the appropriate marginal posterior distributions.

3. Model Determination

3.1. Model Adequacy

Model adequacy in survival analysis is an important issue. In the classical set up, there are two approaches. One is to assume a parametric function for the baseline hazard λ_o, assess an empirical estimate $\hat{\lambda_o}$ say, plot it against t or $\log(t)$ and compare the plot with the theoretical one. The other approach uses ad-hoc theoretical or empirical Q-Q plots of exponential residuals (Lawless, 1982 and Cox and Oakes, 1984 and the references therein). In the Bayesian framework very few papers address the issue of model adequacy with some notable exceptions, e.g., Ghosh (1996), Sinha and Dey (1997) and references therein.

Bayesian model examination for adequacy proceeds by calculating different predictive distributions. We outline the concepts and calculations below mainly following Gelfand *et al.* (1992); Gelfand and Dey (1994) and Gelfand (1996). In generic notation, let $\pi(\cdot)$ denotes the density of its argument and $\boldsymbol{\theta}$ denotes all the parameters under the assumed model. Let $\mathbf{y}_{obs}$ denotes the observed data. The posterior predictive density, $\pi(\mathbf{y}|\mathbf{y}_{obs})$, is the predictive density of a new independent set of observable under the model, given the actual set of observable, Gelfand (1996). By marginalizing $\pi(\mathbf{y}|\mathbf{y}_{obs})$ we obtain the posterior predictive density of one observation $y_r, r = 1, \ldots, N$ where N is the total number of observations, as follows,

$$\pi(y_r|\mathbf{y}_{obs}) = \int \pi(y_r|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathbf{y}_{obs})d\boldsymbol{\theta}. \tag{10}$$

Equation (10) may be used for model checking in the Bayesian framework. Suppose that we draw samples from the above density and form $100 \times q\%$ equal tailed credible intervals. Then under a given model we would expect at least $100 \times q\%$ of actual observations to lie in that interval. The model under consideration would be an adequate model for the data if it supports the above. This is the strategy we will use for model checking in our example in Section 4.

How do we sample from the predictive density in (10)? Suppose that $\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(B)}$ denotes B samples from $\pi(\boldsymbol{\theta}|\mathbf{y})$, possibly using one of the MCMC methods. Then, a random sample $y_r^{(j)}$ drawn from $\pi(y_r|\boldsymbol{\theta}^{(j)})$, is a sample from the above predictive density, see Gelfand (1996) for more details. Suppose that μ_r and σ_r^2 denote the posterior predictive mean and variance of y_r under the density (10). We can easily estimate μ_r and σ_r^2 by Monte Carlo integration using the samples $y_r^{(j)}, j = 1, \ldots, B$.

3.2. Model Selection

In this subsection we consider the problem of accounting for uncertainty about the model form. We are faced with many models that involve different assumptions, distributional forms, or sets of co-variates or other model parameters. Although we may wish to summarize our findings with a single model, there are usually many choices to be made. The pure Bayesian approach for model comparison, is to report the posterior probabilities of each

model, by comparing Bayes factors (see, e.g., Kass and Raftery, 1995). Though there has been recent advances in computing the Bayes factor, see e.g., Raftery (1996) for a review, there are problems in calculating the Bayes factor for high dimensional models that we work with. Also for improper priors the Bayes factor is not meaningful since it can not be calibrated. In this section we use two alternative approaches for model selection.

The first one is based on cross-validation predictive density. Suppose that $\mathbf{y}_{(r)}$ denotes the set of data points deleting the rth observation. Cross-validation predictive density is given by,

$$\pi(y_r|\mathbf{y}_{(r)}) = \int \pi(y_r|\boldsymbol{\theta}, \mathbf{y}_{(r)})\pi(\boldsymbol{\theta}|\mathbf{y}_{(r)})d\boldsymbol{\theta}. \tag{11}$$

This is popularly known as the conditional predictive ordinate (CPO). We can compare more than one models using the CPOs. The so called CPO plot, a plot which overlays the CPO's for different models in a single graph provides useful information for model examination. We prefer a model for which the CPO values are higher than those for the other models under consideration. Model comparison can also be done using summary measures. Suppose that we have two models M_1 and M_2 under consideration. The pseudo-Bayes factor, a surrogate for the Bayes factor, for comparing the two models is defined as

$$\text{PSBF}(M_1, M_2) = \frac{\prod_{r=1}^{N} \pi(y_r|\mathbf{y}_{(r)}|M_1)}{\prod_{r=1}^{N} \pi(y_r|\mathbf{y}_{(r)}|M_2)} \tag{12}$$

where $\pi(y_r|\mathbf{y}_{(r)}|M_i)$ is (11) under model $i = 1, 2$.

It is straightforward to estimate the CPO given by (11) when samples from the posterior distribution $\pi(\boldsymbol{\theta}|\mathbf{y})$ are available and given $\boldsymbol{\theta}$ the observations are conditionally independent, i.e., $\pi(y_r|\boldsymbol{\theta}, \mathbf{y}_{(r)}) = \pi(y_r|\boldsymbol{\theta})$. Note that given the frailty parameter w_i, t_{ij}'s are independent. This fact is crucial in the development below. Suppose that we have B samples $\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(B)}$ drawn approximately from $\pi(\boldsymbol{\theta}|\mathbf{y})$, possibly using one of the MCMC methods. A Monte Carlo estimate of $\pi(y_r|\mathbf{y}_{(r)})$ in (11) is then

$$\hat{\pi}(y_r|\mathbf{y}_{(r)}) = B\left[\sum_{j=1}^{B}\left\{\pi(y_r|\boldsymbol{\theta}^{(j)})\right\}^{-1}\right]^{-1}, \tag{13}$$

which is the harmonic mean of the conditional density of y_r evaluated at the posterior sample values.

Though, the above model selection methods based on cross-validation predictive densities have value, the summary measure (12) can not be calibrated because unlike the Bayes factor it does not have a probability interpretation. In simple terms, we would not know how large a pseudo-Bayes factor is large. Hence, we must look for alternative model selection criteria.

Recently Gelfand and Ghosh (1997) have proposed a model choice criterion by studying utility functions. They consider loss functions which reward an action for its closeness to the predictive value and penalizes the action if its too far from the observed value. The criterion is then obtained by minimizing this posterior predictive loss. As they claim, the

criterion emerges approximately as a form partitioned into a goodness-of-fit term and a penalty term for a wide range of models. With squared error loss the criterion is:

$$D'_\omega = \sum_{r=1}^{N} \sigma_r^2 + \frac{\omega}{\omega+1} \sum_{r=1}^{N} (\mu_r - y_{r,obs})^2 \tag{14}$$

where μ_r and σ_r^2 are as defined in Section 3.1 and $\omega > 0$ is some constant. The first term is a penalty term which penalizes both under-fitted and over-fitted models, since the predictive variances in such cases will tend to be larger. The second term without the factor involving ω is a goodness-of-fit measure. Model selection using D'_ω is usually not sensitive to ω and we will illustrate this in the example in Section 4. For censored data the criterion in (14) must be modified because $y_{r,obs}$ is not available for censored cases. The modified criterion (Gelfand and Ghosh, 1997) is

$$D_\omega = \sum_{r=1}^{N} \sigma_r^2 + \frac{\omega}{\omega+1} \sum_{r=1}^{N} (\mu_r - v_r)^2 \tag{15}$$

where $v_r = y_{r,obs}$ if the rth observation is a failure time and $v_r = \max(\mu_r, s_r)$ if the rth observation is censored at s_r. We illustrate the performance of different models under consideration using all of the above model choice criteria.

4. Kidney Infection Data Example

McGilchrist and Aisbett (1991) analyze time to first and second recurrence of infection in 38 kidney patients on dialysis using a Cox proportional hazard model with a multiplicative frailty parameter for each patient. Two primary co-variates are age of the patients at the time of each infection and sex of the patient. The data set has been reanalyzed by BUGS, Spiegelhalter *et al.* (1995), using an additive frailty model as described in Section 2.3. We consider the following four models for this data set.

1. Model I: Piecewise exponential model with gamma priors for the λ_k's as in (4).
2. Model II: Weibull model (6) with multiplicative gamma frailties.
3. Model III: Piecewise exponential model with normal priors for the $\log(\lambda_k)$'s as in (5).
4. Model IV: Weibull model (7) with additive frailties.

The proportional hazard component of each of the above models is

$$\theta_{ij} = \exp(\boldsymbol{\beta}^T \mathbf{z}_{ij}) = \exp(\beta_{sex} sex_i + \beta_{age} age_{ij})$$

where $sex_i = 1$ if the ith patient is a female and 0 otherwise, age_{ij} is the age at the jth infection of the ith patient.

Each of the above models was fitted using Gibbs sampling. In particular, we used the BUGS software (Spiegelhalter *et al.*, 1995) to fit Model IV. The adaptive rejection sampling

of Gilks and Wild (1992) was used whenever the complete conditional distribution was non-standard and log-concave, see Appendix for discussion on log-concavity. A Metropolis step was implemented if the distribution was not log-concave. Many convergence diagnostics measures were calculated, e.g., Geweke (1992) and Raftery and Lewis (1992) to monitor convergence. The first 1000 iterates in each case were discarded and the subsequent 10000 iterates were used to make inference.

The following values of hyper-parameters were used in the simulation. For the prior on inverse frailty variance, η, ϕ was taken to be 0.001. Each component of $\boldsymbol{\beta}$ was assumed a-priori normal with 0 mean and variance 10^3. The same prior was assumed for ν in Model IV. For Model I all the α_k's were assumed to be 0.01. For Model III τ^2 was taken as 10^4 to make it comparable with the corresponding prior precision for λ_k's in Model I. For models III and IV we took $\rho = 0.001$, $\kappa_1 = 1$ and $\kappa_2 = 0.001$.

We first investigated different choices of the grid size g for Models I and III. We experimented with three choices of $g = 5, 10$ and 20. The $g = 5$ case seemed to give worse model fitting than the $g = 10$ case and the last choice of g did not provide substantially better results than the $g = 10$ case. Hence we decided to use $g = 10$ throughout. Model fitting and/or model choice were not very much sensitive to small variations on the values of the other hyper-parameters as given above. Widely different α_k's in Model I did change the estimates a little bit. However, that did not alter the model choice ordering as reported below.

Table 1 shows the posterior mean, standard deviation and 95% credible intervals for β_{sex}, β_{age}, $\sigma^2_{frailty} = \eta^{-1}$. We show the estimates of and α and μ (for Model IV $\mu = \exp\{\nu\}$) in Table 2. Figure 1 shows the marginal posterior density estimates for β_{sex} under the four models. The estimates of β_{sex} show that the female patients have a slightly lower risk for infection. The estimates of $\sigma^2_{frailty}$ from different models show that there is a strong posterior evidence of high degree of heterogeneity in the population of patients. Some patients are expected to be very prone to infection compared to others with the same co-variate value. This is not very surprising, as in the data set there is a male patient with infection times 8 and 16, and there is also another male patient with infection times 152 and 562. The high posterior means of $\sigma^2_{frailty}$ also provide evidence of strong positive correlation between two infection times for the same patient.

Table 1. Parameter estimates from different Models. Posterior means are followed by (standard deviations) in the first row. 95% credible intervals are shown in the second row.

	Model I	Model II	Model III	Model IV
β_{sex}	–1.493 (0.468)	–1.888 (0.564)	–1.5 (0.480)	–1.69 (0.529)
	(–2.43, –0.6)	(–3.034, –0.846)	(–2.467, –0.624)	(–2.78, –0.699)
β_{age}	0.0061 (0.013)	0.0074 (0.013)	0.0065 (0.013)	0.006 (0.014)
	(–0.0184, 0.0321)	(–0.0178, 0.0322)	(–0.0179, 0.0356)	(–0.0189, 0.0356)
$\sigma^2_{frailty}$	0.499 (0.283)	0.585 (0.307)	0.523 (0.285)	0.816 (0.507)
	(0.061, 1.160)	(0.115, 1.317)	(0.089, 1.195)	(0.079, 2.05)

The above analysis suggests that Models I and III are very close to each other while Models II and IV are also somewhat similar. Hence we proceed with only Models I and II for further analysis based on predictive distributions. Under Model I, 44 and 75 out of 76

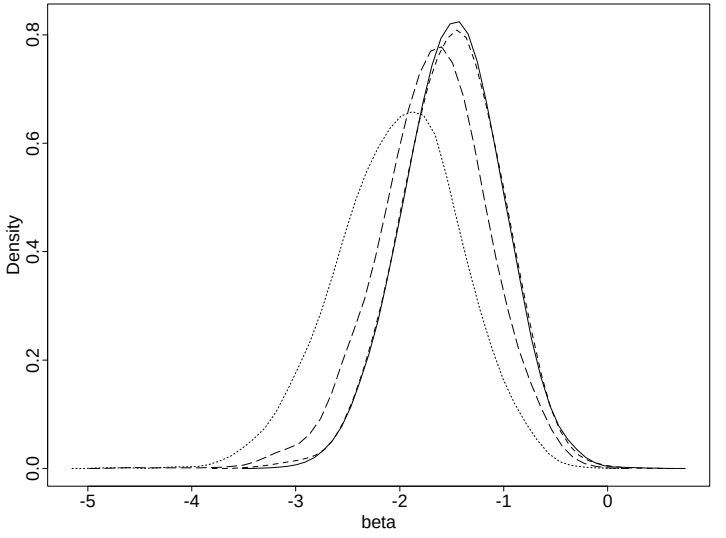

Figure 1. Marginal Posterior density estimates of β_{sex} under four models. Solid line for Model I; dotted line for Model II; dashed line for Model III and big dashed line for Model IV.

Table 2. Parameter estimates of α and μ for the Weibull Models II and IV. Posterior means are followed by (standard deviations) in the first row. 95% credible intervals are shown in the second row.

	Model II	Model IV
α	1.278 (0.190)	1.22 (0.16)
	(0.937, 1.692)	(0.916, 1.54)
μ	0.0157 (0.015)	0.0129 (0.014)
	(0.0012, 0.0575)	(0.0011, 0.0525)

observations were contained in the 50% and 95% predictive intervals respectively, while under Model II the number of observations were 39 and 74 respectively. This shows that Model I is doing better than Model II, though both may seem to be adequate.

Another diagnostic is based on CPO values obtained from Model I and II. Figure 2 plots the log(CPO, Model I) – log(CPO, Model II) against observation numbers. It shows approximately 50% of the observations are supporting Model I. The pseudo Bayes factor (12) for comparing models I and II is very close to 2.5, indicating that it could not discriminate between the two models effectively.

Finally, in Table 3 we show the values relating to the model choice criterion D_k as given in (15) for both the Models I and II. First two columns give the two parts of (15). As mentioned previously, the first part is the penalty term (P), the sum of the predictive variances and the second part is the goodness-of-fit term (G). Note that the Weibull model II received much

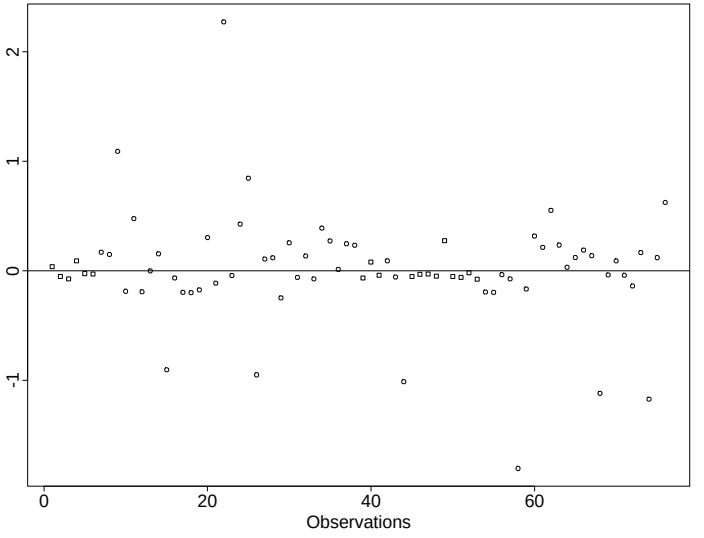

Figure 2. Log(CPO) difference between Model I and Model II. A point bigger than 0 supports Model I. □ represents a censored observation and ○ represents a failure time.

Table 3. Model selection criterion (15) ($\times 10^{-6}$) for Models I and II. P is the penalty (first) term in (15), G is the goodness-of-fit (second) term in (15) without the ω factor. Model I is favored by the criterion.

	P	G	D_1	D_5	D_{10}	D_∞
Model I	1.31	0.643	1.63	1.84	1.89	1.95
Model II	103.06	0.499	103.31	103.48	103.51	103.56

higher penalty than the piecewise exponential model I, but in terms of goodness-of-fit the Weibull model provided a slightly better fit. This is partially explained as follows. Note that the Weibull distribution with $\alpha < 1$ has larger variance than the exponential distribution ($\alpha = 1$). The posterior distribution of α in Model II has substantial mass below the point 1, see Table 2. Also, censoring effects the conclusions. The predictive variances for the censored observations are in general higher than those for the failed observations. If we recalculate the quantities in Table 3 ignoring the censored observations then the penalty for the Weibull Model II is only approximately twice. Hence, the Weibull model is much worse in predicting a censored observation than the piecewise exponential model in this situation. In conclusion, it is appropriate that Model I, i.e., a piecewise exponential model with multiplicative gamma frailties is the best choice for this data set among all the models we have considered here.

Acknowledgments

This paper has benefited from discussions with many people, particularly Alan Gelfand from the University of Connecticut. The authors acknowledge grants from the Engineering and Physical Sciences Research Council (UK), the National Science Foundation (USA, SCREMS DMS-09506557) and the National Cancer Institute (USA, R29-CA69222-02).

Appendix

Conditional distributions for the piecewise exponential model

Using the data likelihood $L(\boldsymbol{\beta}, \boldsymbol{\lambda}, |\mathbf{W}, \mathbf{Y}, \mathbf{Z})$ from (8) and the priors we compute all the complete conditional distributions (CCDs) needed for Gibbs sampling. Let $\pi(\cdot)$ denote the prior density of its argument and

$$S_i = \sum_{j=1}^{m} \theta_{ij} \left(\sum_{k=1}^{g_{ij}} \lambda_k \Delta_k + \lambda_{g_{ij}+1}(t_{ij} - t_{g_{ij}}) \right). \tag{A.1}$$

The CCD of each w_i is a gamma distribution as given below,

$$w_i \sim \ Gamma \left\{ \eta + \sum_{j=1}^{m} \delta_{ij},\ \eta + S_i \right\},\ i = 1, \ldots, n.$$

The CCD of η is,

$$\eta \propto \prod_{i=1}^{n} w_i^{\eta-1} \eta^{n\eta} \frac{\exp\{-\eta \sum_{i=1}^{n} w_i\}}{[\Gamma(\eta)]^n} \pi(\eta).$$

The CCD of $\boldsymbol{\beta}$ is,

$$\boldsymbol{\beta} \propto \ \exp\left\{ \boldsymbol{\beta}^T \sum_{i=1}^{n} \sum_{j=1}^{m} \delta_{ij} \mathbf{z}_{ij} - \sum_{i=1}^{n} w_i S_i \right\} \pi(\boldsymbol{\beta}).$$

Let

$$V_k = \sum_{(i,j) \in R_k} \Delta_k \theta_{ij} w_i + \sum_{(i,j) \in D_k} (t_{ij} - t_{k-1}) \theta_{ij} w_i,$$

where $R_k = \{(i,j); t_{ij} > t_k\}$ i.e., the risk set at t_k, and $D_k = R_{k-1} - R_k$. The CCD of $\lambda_k, k = 1, \ldots, g$ is,

$$\lambda_k \propto \ \lambda_k^{d_k} \exp\{-\lambda_k V_k\} \tau(\lambda_k) \tag{A.2}$$

where d_k is the number of failure times that occurred in the interval I_k and $\tau(\lambda_k)$, the conditional prior, given as follows

$$\tau(\lambda_1 | \lambda_2) = \lambda_1^{-\alpha_0} \exp\left(-\alpha_0 \frac{\lambda_2}{\lambda_1}\right) \pi(\lambda_1),$$

$$\tau(\lambda_k|\lambda_{k-1},\lambda_{k+1}) = \lambda_k^{-1}\exp\Big\{-\alpha_0(\frac{\lambda_k}{\lambda_{k-1}}+\frac{\lambda_{k+1}}{\lambda_k})\Big\},\ k=2,\ldots,g-1 \text{ and}$$

$$\tau(\lambda_g|\lambda_{g-1}) = \lambda_g^{\alpha_0-1}\exp\Big(-\alpha_0\frac{\lambda_g}{\lambda_{g-1}}\Big).$$

Conditional distributions for the Weibull model

Using the data likelihood $L(\boldsymbol{\beta},\mu,\alpha|\mathbf{W},\mathbf{Y},\mathbf{Z})$ from (9) and the priors the following are the CCDs for all the parameters needed for Gibbs sampling. Let $\pi(\cdot)$ denote the prior density of its argument as before and

$$S=\sum_{i=1}^{n}\sum_{j=1}^{m}t_{ij}^{\alpha}\theta_{ij}w_i. \tag{A.3}$$

The CCD of each w_i is a gamma distribution i.e.,

$$w_i \sim\ Gamma\left\{\eta+\sum_{j=1}^{m}\delta_{ij},\ \eta+\mu\sum_{j=1}^{m}t_{ij}^{\alpha}\theta_{ij}\right\},\ i=1,\ldots,n.$$

The CCD of η is

$$\eta \propto \prod_{i=1}^{n} w_i^{\eta-1}\eta^{n\eta}\frac{\exp\{-\eta\sum_{i=1}^{n}w_i\}}{[\Gamma(\eta)]^n}\pi(\eta).$$

The CCD of $\boldsymbol{\beta}$ is

$$\boldsymbol{\beta} \propto \exp\left\{\boldsymbol{\beta}^T\sum_{i=1}^{n}\sum_{j=1}^{m}\mathbf{z}_{ij}\delta_{ij}-\mu S\right\}\pi(\boldsymbol{\beta}).$$

The CCD of μ is a gamma distribution i.e.,

$$\mu \sim\ Gamma\left\{\rho+\sum_{i=1}^{n}\sum_{j=1}^{m}\delta_{ij},\ \rho+S\right\}.$$

Finally the CCD of α is

$$\alpha \propto \left(\prod_{i=1}^{n}\prod_{j=1}^{m}t_{ij}^{\delta_{ij}}\right)^{\alpha-1}\alpha^{\sum_{i=1}^{n}\sum_{j=1}^{m}\delta_{ij}}\exp\{-\mu S\}\,\pi(\alpha).$$

Remark on Log-Concavity: It is easy to see that, except for the density of λ_k in (A.2), each of the above conditional densities is *log-concave*, i.e., the second derivative of the

log-density is strictly decreasing. (Use a property of *digamma function*, $\Psi(x) = \frac{\Gamma'(x)}{\Gamma(x)}$, see e.g., Abramowitz and Stegun (1965, p260), viz.,

$$\Psi'(x) = \sum_{j=0}^{\infty} \frac{1}{(x+j)^2}$$

to prove log-concavity of the CCD of η.) A sufficient condition for the density of λ_k in (A.2) to be log-concave is $d_k \geq 1$, that is, there is at least one failure in the kth interval. Log-concavity of each λ_k, will be ensured if there is at least one failure in each of the intervals.

References

M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions*, Dover, 1965.

E. Arjas and D. Gasbarra, "Nonparametric Bayesian inference for right-censored survival data, using the Gibbs sampler," *Statistica Sinica*, vol.**4**, pp. 505–524, 1994.

N. E. Breslow, "Covariance analysis of censored survival data," *Biometrics*, vol.**30**, pp. 89–99, 1974.

D. Clayton, "A model for association in bivariate life tables and its application in epidemiological studies of familiar tendency in chronic disease incidence," *Biometrika*, vol.**65**, pp. 141–151, 1978.

D. Clayton and J. Cuzick, "Multivariate generalizations of the proportional hazards model (with discussion)," *J. Roy. Statist. Soc.*, A vol.**148**, pp. 82–117, 1985.

D. R. Cox and D. Oakes, *Analysis of Survival Data*, London: Chapman & Hall, 1984.

D. Gamerman, "Dynamic Bayesian models for survival data," *Appl. Statist.*, vol.**40**, pp. 63–79, 1991.

A. E. Gelfand, "Model determination using sampling based methods," in *Markov Chain Monte Carlo in Practice*, (Eds. W. R. Gilks, S. Richardson and D. J. Spiegelhalter), London: Chapman and Hall, 1996, pp. 145–161.

A. E. Gelfand and D. K. Dey, "Bayesian model choice: asymptotics and exact calculations," *J. Roy. Statist. Soc.*, B vol.**56**, pp. 501–514, 1994.

A. E. Gelfand, D. K. Dey and H. Chang, "Model determination using predictive distributions with implementation via sampling-based methods (with discussion)," in *Bayesian Statistics 4*, (Eds. J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith), Oxford: Oxford University Press, 1992, pp. 147–167.

A. E. Gelfand and S. K. Ghosh, "Model Choice: A Minimum Posterior Predictive Loss Approach," to appear in *Biometrika*, 1997.

A. E. Gelfand and A. F. M. Smith, "Sampling based approaches to calculating marginal densities," *J. Amer. Statist. Assoc.*, vol.**85**, pp. 398–409, 1990.

J. Geweke, "Evaluating the accuracy of the sampling-based approaches to the calculation of the posterior moments," in *Bayesian Statistics 4*, (Eds. J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith), Oxford: Oxford University Press, 1992, pp. 169–194.

S. K. Ghosh, "Modeling and Analysis of Multiple Event Survival Data." Ph. D. thesis, Department of Statistics, University of Connecticut, Storrs, Connecticut, 1996.

W. R. Gilks, S. Richardson and D. G. Spiegelhalter, *Markov Chain Monte Carlo In Practice*. London: Chapman and Hall 1996.

W. R. Gilks and P. Wild, "Adaptive rejection sampling for Gibbs sampling," *Appl. Statist.*, vol.**41**, pp. 337–348, 1992.

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

P. Hougaard, "A class of multivariate failure time distributions," *Biometrika*, vol.**73**, pp. 671–678, 1986.

J. D. Kalbfleisch and R. L. Prentice, "Marginal likelihoods based on Cox's regression and life model," *Biometrika*, vol.**60**, pp. 267–278, 1973.

R. E. Kass and A. E. Raftery, "Bayes factors," *J. Amer. Statist. Assoc.*, vol.**90**, pp. 773–795, 1995.

J. F. Lawless, *Statistical Models and Methods for Life Time Data*. New York: John Wiley, 1982.

T. Leonard, "Density estimation, stochastic processes and prior information," *J. Roy. Statist. Soc.*, B vol.**40**, pp. 113–146, 1978.

C. A. McGilchrist and C. W. Aisbett, "Regression with frailty in survival analysis," *Biometrics*, vol.**47**, pp. 461–466, 1991.

D. Oakes, "A model for association in bivariate survival data," *J. Roy. Statist. Soc.*, B, vol.**44**, pp. 414–422, 1982.

A. E. Raftery, "Hypothesis testing and model selection," in *Markov Chain Monte Carlo in Practice*, (Eds. W. R. Gilks, S. Richardson and D. J. Spiegelhalter), London: Chapman and Hall, 1996, pp. 163–187.

A. E. Raftery and S. Lewis, "How many iterations in the Gibbs sampler?," in *Bayesian Statistics 4*, (Eds. J. M. Bernado, J. O. Berger, A. P. Dawid and A. F. M. Smith), Oxford: Oxford University Press, 1992, pp. 765–776.

D. Sinha and D. K. Dey, "Semiparametric Bayesian Analysis of Survival Data," to appear in *J. Amer. Statist. Assoc.*, 1997.

D. J. Spiegelhalter, A. Thomas, N. G. Best and W. R. Gilks, "BUGS: Bayesian Inference Using Gibbs Sampling, Version 0.50." MRC Biostatistics Unit, Cambridge, England, 1995.

M. Tanner and W. Wong, "The Calculation of Posterior Distributions by Data Augmentation (with discussion)," *J. Amer. Statist. Assoc.*, vol.**82**, pp. 528–550, 1987.

G. A. Whitmore and M. L. T. Lee, "A multivariate survival distribution generated by an Inverse Gaussian mixture of exponentials," *Technometrics*, vol.**33**, pp. 39–50, 1991.